KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Rega Institute for Medical Research
Department of Microbiology and Immunology
Laboratory for Clinical and Evolutionary Virology

QUANTIFYING CHRONIC INFLAMMATORY BURDEN FROM TRANSCRIPTOMES IN VIRAL AND IMMUNE-MEDIATED PATHOLOGIES

Tim Dierckx

Jury members:

Promotor:	Prof. Dr. Anne-Mieke Vandamme
Co-promotor:	Dr. Johan Van Weyenbergh
Co-promotor:	Dr. Bram Vrancken
Chair:	Prof. Dr. Kathleen Freson
Secretary:	Prof. Dr. Johan Neyts
Jury members:	Prof. Dr. Johan Neyts
	Prof. Dr. Hendrik Blockeel
	Dr. Anne Van den Broeke
	Dr. Heli Salmela

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Biomedical Sciences

May 2019

Table of Contents

Table of Figures

Abbreviations

AA	Ascorbic Acid
AAT	alpha-1-antitrypsin
AC	Asymptomatic Carrier, Asymptomatic Control
ADM	Adalimumab
AGP	alpha-1-acid glycoprotein
AIDS	Acquired Immonodeficiency Syndrome
ALL	Acute Lymphoid Leukemia
ANN	Artificial Neural Networks
ANOVA	ANalysis Of VAriance
ATL	Adult T-cell leukemia/lymphoma
AUC	Area Under the Curve
AZT	zidovudine
CD	Crohn's Disease
CRP	C-reactive Protein
CVD	Cardiovascular Disease
DC	Dendritic Cell
DILGOM	DIetary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome
DNA	Deoxyribonucleic Acid
eGFR	estimated Glomerular Filtration Rate
ES	Enrichment Score
ESR	Erythrocyte Sedimentation Rate
fCal	fecal calprotectin
FCS	Fetal Calf Serum
FPR	False Positive Ratio
GEO	Gene Expression Omnibus
GlcNAc	N-Acetylglucosamine
GlycA	Glycoprotein Acetylation
GO	Gene Ontology
GSA	Gene Set Analysis
GSEA	Gene Set Enrichment Analysis
HAM/TSP	HTLV-1 Associated Myelopathy
HBI	Harvey-Bradshaw Index
HBZ	HTLV-1 bZIP factor
HC	Healthy Control
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
HP	Haptoglobin
hsCRP	high sensitivity CRP
HTLV	Human T-Cell Leukemia Virus

IBD	Inflammatory Bowel Disease
IDH	Infective Dermatitis associated with HTLV-1
IFN	Interferon
IFX	Infliximab
Ig	Immunoglobulin
IKKB	IkB kinase
IL	Interleukin
IQR	Inter-quartile Range
IU	International Unit
KD	Kawasaki Disease
kNN	k-Nearest Neighbors
LASSO	Least Aboslute Shrinkage and Selection Operator
LM	Linear Model
LN	Lupus Nephritis
mRMR	minimum Redundancy Maximum Relevance
MS	Multiple Sclerosis
NET	Neutrophil Extracellular Trap
NMR	Nulcear Magnetic Resonance spectroscopy
ORA	Overrepresentation Analysis
PBMC	Peripheral Blood Mononuclear Cell
PCA	Principal Component Analysis
PCNA	Proliferating Cell Nuclear Antigen
PD-L1	Programmed death Ligand 1
PI	para-inflammation
PRO2	Patient Reported Outcome
QD	Quantile Discretization
RA	Rheumatoid Arthritis
RAR	Retinoid Acid Receptor
rbf	radial basis function
RBV	ribavirin
RFE	Recursive Feature Elimination
RMSE	Root Mean Square Error
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
ROR	RAR-like Orphan Receptor
ROS	Reactive Oxygen Species
SD	Standard Deviation
SLE	Systemic Lupus Erythematosus
SLEDAI	SLE Disease Activity Index
STAT	Signal Transducer and Activator of Transcription
SVM	Support Vector Machine
T_c	cytotoxic T cells

T_h	T helper
TLR	Toll-like Receptor
TNF	Tumor Necrosis Factor
TPM	Transcripts per Million
TPR	True Positive Ratio
T_{reg}	Regulatory T cell
UC	Ulcerative Colitis
UPCR	Urinary Protein / Creatinine Ratio
UST	Ustekinumab
VDM	Vedolizumab
VSV	Vesicular Stomatitis Virus
WGCNA	Weighted Gene correlation Network Analysis
YFS	Young Finns Study

CHAPTER 1
Introduction

This PhD thesis is a result of research efforts grounded in a greater project which is based on the premise that the success of immunomodulatory treatment in viral infections depends on the state of the host immune system. More explicitly, does the immune system being 'primed', 'activated' or 'exhausted' play a role in determining the treatment outcome when treating a viral infection with a biological such as interferon (IFN). Characterizing the state of the immune system is not a straightforward task as it is the result of many different interlocking biological pathways. For the viral infections that are investigated by myself and my colleagues in my research group, namely Human T-cell Lymphotropic Virus 1 (HTLV-1), Human Immunodeficiency Virus (HIV), and Hepatitis C Virus (HCV), three aspects of the immune system are considered to be of particular interest: the antiviral effector pathways, the IFN related cellular machinery, and the inflammatory burden.

While the relevance of the host's antiviral effector molecules to antiviral treatment success is held as self-evident, the hypothesized importance of the IFN related cellular machinery stems from the so called 'IFN paradox' observed in HIV and HCV. While IFN is a major line of defense in the host's arsenal against viral infection and IFN-based therapies have been used in the treatment of HCV and HIV, it can also play a deleterious role, suppressing the immune system in ways that promote viral persistence [1]. A role for a third aspect, the inflammatory burden, was observed in the experiments comparing the effects of two types of IFN in *ex vivo* samples of HTLV-1 infected Adult T-cell Leukemia/Lymphoma (ATL) patients and in the analysis of publicly available gene expression data of ATL patients (reported in Chapters 2 and 3, respectively). Furthermore, the other HTLV-1 related pathology, HTLV 1 Associated Myelopathy or Tropical Spastic Paraparesis (HAM/TSP), is a chronic inflammatory disease of the central nervous system; providing further incentive to consider the inflammatory burden in this viral infection context. The characterization of the inflammatory burden is the central topic in this thesis. More specifically, as the enormous potential of accurately quantifying the chronic inflammatory burden from gene expression profiles burden became evident, it became my main research objective.

Considering the large range of pathologies explored in this thesis, and the myriad different techniques that has been used throughout it, this introduction must by necessity function as a primer on a multitude of different topics. As such, the rest of this introduction is laid out as follows: first, I provide an overview of what constitutes inflammation, how it can be quantified, and what methodology was used to characterize the chronic inflammatory burden

in Chapters 4 – 6. Second, I provide an introduction into the pathologies investigated explicitly in this thesis i.e. HTLV-1 and its related pathology Adult T-cell Leukemia/Lymphoma (ATL), followed by primers on two other inflammatory pathologies: inflammatory bowel disease (IBD) and Systemic Lupus Erythematosus (SLE). The final section of this introduction consists of a brief overview of the employed transcriptomic analysis strategies throughout these works (Chapters 2 – 4). The statistical and machine learning methods used to construct mathematical predictive models from these transcriptomic experiments is introduced and detailed in Chapter 4.

1.1 Inflammation

Inflammation is a complex biological process that is integral to our bodies' protective response to harmful stimuli. The term 'inflammation' is a blanket term that covers a variety of different responses that can result from a wide range of immune cells and molecular interactions. The cause, duration and intensity of the inflammation affect the underlying molecular interaction cascades, adding to the complexity of the process.

The typical physiological presentation of inflammation as heat, pain, redness and swelling is what is known as acute inflammation, which is a generic response to cellular damage, cellular detection of pathogens or irritants and is part of our innate immune system. This process is protective and critical to survival: its goal is to remove both the inflammatory trigger and damaged tissue, so that repair processes can commence. To prevent the harmful stimulus from compromising the survival of the host, the capacity of the acute inflammatory response needs to be high. On the other hand, continuous activation of the inflammatory processes is implicated in a number of diseases like hay fever [2], arthritis [3], cardiovascular disease (CVD) [4] or even cancer where it can initiate tumor formation or metastasis [5]. It is then unsurprising that the inflammatory response is extensively regulated in the body.

It is possible to classify inflammation by its severity and duration: where acute inflammation is the body's initial response to the detection of a harmful stimulus by increased recruitment of blood plasma and leukocytes (predominantly neutrophils) from the blood to the affected site, chronic inflammation is characterized by increased recruitment of mononuclear cells caused by prolonged acute inflammation. Recent research has recognized a chronic, subclinical inflammatory state which occurs with aging as a manifestation of immuno-senescence, termed 'inflammaging' [6], [7]. This inflammaging is believed to be caused by the accumulation of an

antigenic load of persistent infection throughout an individual's lifespan and the resulting chronic stimulation of macrophages [7]. Evidence exists that long-lived people, i.e. centenarians, better cope with the effects of inflammaging through an anti-inflammatory response termed "anti-inflammaging" [6], [8], [9]. These low-grade inflammatory processes are coming under greater scrutiny. Recently, an ongoing inflammatory response of a very low magnitude in the epithelium, barely distinct from tissue homeostasis, has been defined as 'para-inflammation' (PI) [5]. PI differs from canonical inflammatory processes by its lack of tumor necrosis factor (TNF) and NF-KB pathway activation in spite of interferon (IFN) activation, and has been observed in a multitude of different cancers [5].

Inflammation types can not only be identified by their duration, but also by their cause: sterile inflammation due to trauma or confirmed sterile systemic inflammatory response syndrome can be differentiated from systemic infectious inflammation (i.e. sepsis) using differences in gene transcription [10]. In what follows, I provide an overview of the principal processes and molecules in the inflammatory response.

1.1.1 Cell populations contributing to inflammation

Neutrophils

The recruitment of leukocytes from blood is considered the primary and integral hallmark of the acute-phase response. Even though this rapid response is facilitated through sentinel mast cells and macrophages stationed throughout all tissues [11], the attraction of leukocytes may still be considered as the initial effect of inflammation. Neutrophils are the most abundant leukocytes in human blood. These short-lived polymorphonuclear cells are the first leukocytes to be recruited to a site of inflammation, where they perform their function by engulfing microbes (phagocytosis) and secreting anti-microbials (reactive oxygen species, ROS, and granules of microbicidal molecules). However, these are not the only means by which neutrophils attack bacteria.

First reported in 2004, the process of neutrophil extracellular trap (NET) formation showed neutrophils using chromatin fiber, enzymes and granule-derived antimicrobial peptides to ensnare and immobilize microbes [12]. This prevents the physical spread of microbes and creates an environment thought to limit host damaging molecules secreted as part of the inflammatory process from escaping to, and unnecessarily damaging, nearby healthy tissue

[13]. However, exaggerated NET formation and inefficient NET clearing has also been linked to autoreactivity to the NET components [14], which causes inflammation.

Monocytes and Macrophages

Monocytes and macrophages are a diverse and plastic set of leukocytes which can be activated in different ways. Broadly, activated macrophages present either a M1 or M2 phenotype, although these are better thought of extremes in a spectrum rather than discrete states [15]. M1 macrophages are highly pro-inflammatory and secrete IL-1β, IL-6 and reactive oxygen and nitrogen intermediates, whereas M2 macrophages have a more anti-inflammatory role and perform regulatory functions, tissue repair and immunosuppression [16]. The induction of macrophages to the M1 state is achieved by IFN-γ, pathogen-derived molecules through Toll-like receptors and granulocyte-macrophage colony-stimulating factor, whereas M2 differentiation is promoted by IL-4, IL-13, IL-10 and steroids [15], [16].

T cells

T cells circulate through the secondary lymphoid organs and are important components of the adaptive immune response, with a small fraction of the total T cells taking part in the innate T cell response. They have pro-inflammatory properties, activate macrophages and other T cells and perform a broad range of regulatory activities. Their properties and function depend on their subtype: T cells can be classified as helpers, cytotoxic, regulatory or memory T cells, with smaller subsets of natural killer T cells, γδ T cells and mucosal associated invariant T cells. A variety of T helper (T_H) subtypes exist, but their overarching function is the maturation of other immune cells (e.g. B cells) and the activation of cytotoxic T cells and macrophages. Once activated, cytotoxic T cells (T_C) destroy virus-infected and tumor cells. Memory T cells are a long-lived subtype which occurs when T cells are presented with an appropriate antigen by a professional antigen presenting cell. Like T_H cells, different memory T cell subtypes can be identified: of particular note is the T memory stem cell [17]. These cells are at the hierarchical apex of the memory T lymphocytes and recent reports suggest that they are crucial to the maintenance of immune homeostasis [18]. At current, little is known about their precise relation to auto-immune disease, but early reports suggest that long-lasting autoreactive or abnormally activated T_{SCM} trigger an inflammatory response that is self-renewing, thereby contributing to the persistence of these diseases [19]. Their importance in retroviral infections targeting $CD4^+$ T cells (i.e. HIV-1 and HTLV-1) has already been investigated: in HIV, T_{SCM} cells can support both productive and transcriptionally silent

infection [20] and due to their long lifespan (a half-life estimated at 227 months in antiretroviral-treated HIV patients [21]) they represent an extremely durable and self-renewing viral reservoir [22]. In HTLV-1 infection, the T_{SCM} cells have been proven to be the progenitors of dominant circulating ATL clones [23].

1.1.2 Inflammation-related Molecules

As the term inflammation covers a broad variety of cellular responses, mediated by a host of different cell types, regulated by dozens of cytokines and chemokines, it is impossible to provide a comprehensive overview of all molecules involved in these processes. Below, I highlight several key proteins relevant to the research topic.

Interleukin 1β

IL-1β is a fever-inducing, pro-inflammatory cytokine, mainly produced by monocytes and macrophages as a result of Toll-like receptor (TLR) activation. IL-1β acts as a chemokine, recruiting inflammatory cells and induces gene expression of several enzymes leading to production of inflammatory mediators prostaglandin E2 and nitric oxide. It also has several systemic effects, increasing platelet and neutrophil and the production of other acute-phase proteins. It induces T helper (Th) 17 cell differentiation and IL-17 production.

Interleukin 17

The IL-17 cytokine is secreted by Th17 cells, CD8 T cells and NK cells and upregulates IL-1β, IL-6 and TNF-α production, in addition to the NF-KB pathway [24]. The cytokine induced the activation and mobilization of neutrophils to sites of inflammation and it's a key mediator of protection against extracellular microbes, with an outspoken protective role against the *Candida albicans* commensal fungus [25], though its effects are not exclusively protective: it drives pathology in various autoimmune diseases [26], one of the functional differences between MS patients and healthy subjects is increased T cell IL-17 production [27], and in CVD with high plasma CRP concentrations, IL-17 has proven to be an actionable clinical target [28].

Interleukin 6

The IL-6 cytokine activates and promotes the expansion of T cells and the differentiation of B cells, as part of its pleiotropic effects. Not only does it regulate the acute-phase response, but it exhibits hormone-like attributes that affect the vascular system, lipid metabolism and

insulin resistance, the neuroendocrine system and neuropsychological behavior [29]. It is produced by stromal cells, as well as the cells of the immune system in response to IL-1β or TNF-α stimulation. IL-6 has both pro- and anti-inflammatory properties depending on the immunological context [30] and is investigated as a target for clinical intervention. While IL-6 plays a protective role in many infections, it also contributes to the maintenance of chronic inflammation in several models of auto-inflammation [31], [32]. The importance of this cytokine is further underscored by the observation that several microorganisms have evolved ways to disrupt or even mimic this immunological pathway: human cytomegalovirus can antagonize IL-6 expression [33], and human herpesvirus 8 expresses a form of IL-6 which can block the recruitment of neutrophils and inhibition of type I interferons owing to a 60% similarity with the human cytokine [34].

Interleukin 10

IL-10 is likely the most well-known anti-inflammatory cytokine. Its anti-inflammatory activities include the downregulation of antigen presentation and T cell activation and the inhibition of pro-inflammatory cytokine production. The mode of action for these anti-inflammatory effects is an intricate cascade which involves its receptor IL-10R and the activation of members of the Janus kinase family which, in turn, activate the transcription factor signal inducer and activator of transcription (STAT) 3. STAT3's binding of a set of target genes across all cell-types (e.g. the *Bcl3* gene which impairs NF-KB DNA binding and suppresses TNF-α production), in addition to a wide range of cell-type-specific targets, facilitates IL-10's anti-inflammatory response [35]. IL-10 is produced by a wide variety of immune cells, including T cells, monocytes, macrophages, neutrophils and dendritic cells, though macrophages are considered to be its main target cells [36], [37].

Tumor Necrosis Factor a

TNF-α is an acute-phase, pro-inflammatory cytokine which is produced by monocytes, macrophages and activated T cells. The mode of action of TNF-α consists of the release of the inhibitory silencer of death domain protein from its receptor to enable the binding of the tumor necrosis factor receptor type 1-associated DEATH domain adaptor protein, which results in the activation of three pathways: activation of NF-KB, activation of mitogen-activated protein kinase pathways and apoptotic pathways, though this latter function is usually counteracted by strong anti-apoptotic effects of NF-KB [38], [39]. Targeting TNF-α

has shown to be a viable strategy in the treatment of a variety of auto-immune diseases [40], [41], and can even complement highly active anti-retroviral treatment of HIV [42].

Interferon

Identified in 1957 as a substance which 'interfered' with, and thus protected cells from, viral infection [43], interferon (IFN) is now known to constitute an integral part of the host immune response to viral infections. The IFN cytokines consist of three families (Type I, II and III), totaling nine different classes [44], [45]. Though they are central to our anti-viral defense, IFN molecules have pleiotropic effects [46] on many different immune cells, including but not limited to: antiproliferative effects, activation of pro-apoptotic genes and proteins, differentiation modulation, antiangiogenic activity (reviewed in [44]). It is therefore unsurprising that IFNs were, and still are, considered a promising option for the treatment of a variety of diseases.

IFN-γ, as the only type II IFN, stands apart from type I IFN-α and -β because its immunomodulatory activity is much greater relative to its antiviral activities [47]. While it is considered a pro-inflammatory cytokine owing to its activation of macrophages and innate immune pathways, it also exerts regulatory activities like the modulation of regulatory and helper T cell differentiation and the limitation of inflammation-associated tissue damage [48], [49].

IFN is used as a biological in several pathologies: arguably the best-known application is the use of IFN-β in multiple sclerosis (MS) treatment. But a number of other important applications can be noted: IFN-α is used in the treatment of leukemia and, in combination therapy with ribavirin (RBV), of hepatitis C virus (HCV) infection. IFN-γ is occasionally employed in the treatment of rare immunodeficiencies [50] and IFN-λ has only recently undergone its first phase 1 clinical trials [51]. The downside of their potent and pleiotropic effects is the significant side-effects of a treatment course [52], [53].

1.1.3 Biomarkers for inflammation

1.1.3.1 C-Reactive Protein

Clinical assessment of the severity of inflammation is routinely performed through quantification of CRP circulating in the blood. The protein takes its name from its ability to

precipitate *Streptococcus pneumoniae* C-polysaccharide and is found in blood plasma [54], [55]. It was the first so-called 'acute-phase' protein to be described, with 'acute phase' referring to the rapid increase from a baseline concentration between 0 and 3 mg/L in healthy individuals following IL-6 secretion of macrophages and T-cells. As such, it can be a highly sensitive marker of inflammation and tissue damage [54], [55]. CRP is produced in the liver, following signals produced by macrophages and adipocytes.

Multiple tests can be used to quantify CRP and overall, they are highly sensitive, cheap, and easy to perform. However, the term 'inflammation' covers a wide range of heterogeneous physiological and biochemical conditions; not all of which are associated with CRP concentration. A clear example of this is found in two chronic inflammatory bowel diseases: in over 25% of Crohn's disease patients, no significant elevation of CRP is observed and in ulcerative colitis patients the CRP response is generally modest or even absent [56]. In addition, CRP is a highly variable biomarker: the determination of the true homeostatic setpoint of a patient can require up to 33 concurrent measurements [57]. Furthermore, the biochemical properties of CRP limit its usefulness in the characterization of an anti-inflammatory state as CRP is supposed to be absent in healthy individuals.

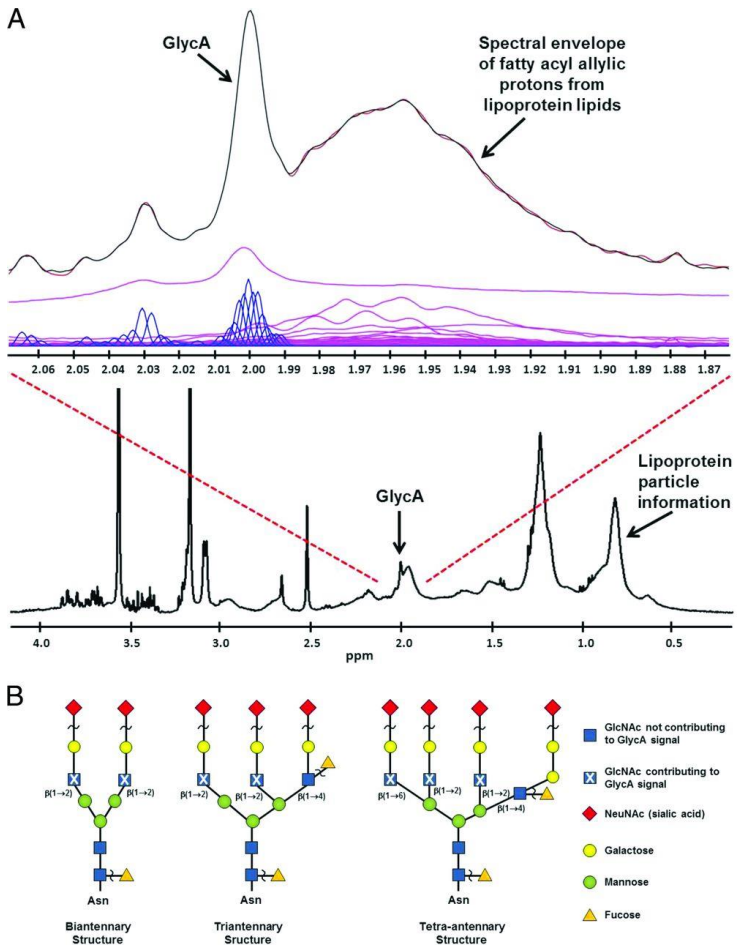


Figure 1: GlycA NMR signal and its glycoprotein glycan origins. (A) A representative plasma NMR spectrum showing the GlycA signal and deconvolution model used by Labcorp (NC, USA) for its quantification. The red line is the measured plasma signal envelope, and the virtually superimposed black line is the calculated sum of the deconvolution-derived amplitudes of the lipoprotein and protein (purple) and N-acetyl methyl (blue) reference signals. GlycA measurements by Nightingale Health (Helsinki, Finland) do not make use of this deconvolution. (B) Locations of the N-acetylglucosamine moieties in major forms of the branched glycans of acute-phase glycoproteins that contribute to the GlycA signal. Figure from [58], reused with permission.

1.1.3.2 Glycoprotein Acetylation: description

The quantification of Glycoprotein Acetylation (GlycA) through NMR spectroscopy was first achieved in 1987 [59]. The amount of GlycA summarizes the NMR signal located between 2.04 and 2.08 ppm on serum NMR spectrum (Figure *1*). It is a non-specific, composite signal which measures the concentration of mobile N-acetyl methyl groups on N-Acetylglucosamine residues of antennary branches of glycoproteins in circulation, i.e. proteins with carbohydrate

sugar groups covalently linked to their amino-acid protein structure in blood plasma [59]. The first reports showed that the signal arose from four main glycoproteins: alpha-1-acid glycoprotein (AGP), alpha-1-antitrypsin (AAT), haptoglobin (HP) and transferrin [59]. Recent, more comprehensive reports also showed Alpha 1-antichymotrypsin as a possible source of the signal [58]. These glycoproteins are acute-phase reactants (like CRP), meaning their glycan structures and their concentrations change during the acute-phase response of inflammation [60]. Though the magnitude of the change in concentration, as well as its kinetics, are different for each of the contributing proteins [61], their combined NMR signal, in the form of GlycA, not only exhibits acute-phase reactant characteristics but has also shown to more robustly associate to inflammatory conditions than traditional inflammatory biomarkers such as CRP concentration. Though GlycA was first described in 1987, the advances in technology that have made it possible to quantify the biomarker in a high throughput fashion are quite recent [58], [62].

Investigations into the molecular underpinnings of GlycA's association to long-term risk of disease and mortality have tried to determine which of GlycA's constitutive glycoproteins best explain the observed associations, and what their the relative contributions to the signal are [63]. These efforts have made use of large scale imputation of the glycoprotein levels using machine learning methods on a set of 626 paired NMR and immunoassay experiments and suggest that while alpha-1-acid glycoprotein correlates best with GlycA levels, alpha-1-antitrypsin (AAT) is more predictive of morbidity and mortality in these population based cohorts [63]. As the literature describes AAT as an inflammation suppressant [64], these results are significant in that they suggest that increased AAT levels should not be interpreted as evidence of anti-inflammatory conditions within the host, but rather that they are actually compensating for the presence of low-grade chronic infection. However, these conclusions are based on imputed concentrations of the constitutive proteins, and studies which simultaneously quantify GlycA and the concentrations of the constitutive proteins are rare. Similarly, though HP is an acute-phase protein, it has also been described as having anti-inflammatory effects (e.g. the reduction of IFN-γ) [65] making its apparent importance to the GlycA signal surprising. However, reports in rheumatoid arthritis (RA) [66], and ovarian [67] and colon [68] cancer have shown that the glycosylation profiles of HP are affected in these pathologies, which can explain the importance of its contribution to the GlycA signal.

1.1.3.3 Glycoprotein Acetylation: published associations

Increases in the GlycA biomarker have been linked to a great number of adverse outcomes. For example, serum NMR profiling of population biobanks have shown associations with fatal and non-fatal cardiovascular disease (CVD) events [69]–[71], type II diabetes [72], mortality from chronic inflammation-related causes [71], non-alcoholic fatty liver disease [73] and all-cause mortality [70], [71]. In datasets generated from cohorts with pre-existing CVD, increased GlycA concentrations associated robustly with greater long-term risk of fatal CVD events [74] and all-cause mortality [75]. Throughout these studies, GlycA is shown to be a better indicator than, and independent of, standard risk factors associated with the considered outcomes like sex, BMI, smoking and activity levels.

Of the associations noted in population-based research, the two most well-researched associations are those with CVD and obesity. In CVD, the GlycA association to CVD and all-cause mortality persists after adjusting for high sensitivity CRP (hsCRP), which can be interpreted as GlycA better capturing the heterogeneous nature of inflammation [69], [76], [77]. In fact, in patients undergoing coronary angiography, GlycA and hsCRP are independent and additive markers for CV event risk [74], [77].

GlycA also seems to be an ideal biomarker for inflammation in the obesity context, where it outperforms other acute-phase proteins hsCRP and IL-1RA as a predictor for type 2 diabetes and CVD event incidence [78]. Curiously, while GlycA has been repeatedly shown to be associated to BMI [79], [80] and bariatric surgery normalizes the elevated GlycA in obese subjects, this normalization occurs even in those patients which remain overweight and obese post-surgery [81]. Multivariate analysis showed that the changes in GlycA following bariatric surgery could be explained by an increase in HDL particle size [81]. The impact of regular exercise on GlycA has been examined in several different cohorts [82]. Not only does regular exercise significantly lower GlycA, this reduction remains significant following adjustment for age, sex, race, baseline BMI and baseline GlycA [82].

While the majority of GlycA studies have been demographic, population-based analyses, the GlycA biomarker has also been investigated in several disease contexts. Rheumatoid Arthritis (RA), as an inflammatory pathology, was an attractive context for initial GlycA quantification studies. Consistently, these studies have found GlycA to be elevated in RA when compared to healthy controls and to be associated to disease activity [83]–[85]. The

association to CVD biomarkers like coronary artery atherosclerosis is evident in RA patients as it is in healthy controls [84].

In Systemic Lupus Erythematosus (SLE), another major inflammatory auto-immune disease, early reports about GlycA appeared to contradict each other: Chung et al. reported increased GlycA levels in SLE patients and noted an absence of correlation with disease activity scores and coronary artery calcification [86], while Durcan et al. observed similar increased GlycA levels which, contrary to the findings of Chung et al, did correlate with disease activity and atherosclerosis markers [87]. I provide a hypothesis to explain this apparent discrepancy in Chapter 6: it is likely that Chung et al.'s choice of SLE patients, i.e. a cohort with low disease activity scores and no nephritic involvement [86], was too homogeneous to identify an association with disease activity and coronary artery calcification.

This view is supported by reports on GlycA in relation to kidney function: in the Brazilian Longitudinal Study of Adult Health, GlycA was found to be related to albuminuria and inversely related to estimated glomerular filtration rate (eGFR) [88]. These observations are echoed in the Dutch PREVEND study cohort [89]. It is worth noting that despite GlycA's association with these renal health biomarkers, the association of GlycA with incident CVD is not attenuated after adjustment for eGFR and albuminuria [89].

Similar observations were made in the chronic inflammatory skin disease psoriasis, where GlycA, in a cohort of 122 patients, shows an increase over healthy controls, and associates with disease severity [90]. Owing to the increased vascular inflammation of this disease, psoriasis patients show increased risk of CV events and mortality and biomarkers for this increased risk are highly valuable. GlycA's association with CVD in healthy controls persists in the psoriasis patients, in contrast to hsCRP and traditional CVD risk factors [90]. GlycA's relevance in this context goes beyond tracking CVD risk, as the authors show that GlycA can be used to accurately track anti-TNF treatment response in psoriasis.

GlycA's ability to characterize CVD incidence and risk is further illustrated in Kawasaki disease (KD). This acute auto-immune disease predominantly affecting young children is characterized by vascular inflammation and presents with fever and mucocutaneous changes [91]. Though diagnosis is very difficult, it is exceedingly important to diagnose the syndrome in a timely fashion because, even though a very effective therapy exists, there is a high chance of serious CV damage if the treatment is not administered early in the disease [92]. This makes the lack of a diagnostic laboratory test a serious issue and explains the importance of GlycA's

ability to differentiate between KD and febrile infections. Again, GlycA outperforms traditional inflammatory biomarkers CRP and erythrocyte sedimentation rate (ESR) (especially in combination with the other lipoprotein biomarkers measured simultaneously in GlycA's NMR spectrum) [57].

There are also settings in which GlycA is not a relevant biomarker: experiments in Sickle Cell Disease (SCD) have shown that the biomarker is not a good fit for that particular pathology, despite SCD being a pathology that is driven by recurrent vascular inflammation [93]. Considering GlycA's robust and extensively proven association to CVD, this is somewhat surprising. The GlycA decrease in SCD patients when compared to healthy controls, in addition to the observed (statistically non-significant) decrease of GlycA in acute SCD cases compared to steady state observations in the same patient, suggest that HP (an important contributor to the GlycA signal) depletion due to hemolysis prevents GlycA from acting as a suitable biomarker for inflammation in this disease context [93].

Finally, a single study has been reported in which GlycA has been investigated in the context of a viral infection. In a cross-sectional study of 935 subjects, of which 63% were HIV-infected, GlycA levels were significantly higher in the HIV-infected individuals [94]. Controlling for HIV serostatus, demographic and CVD risk factors in this cohort preserved GlycA's association to the presence of coronary plaque [94].

While a mechanistic explanation for GlycA's associations has not been determined yet, correlation network analysis in transcriptomic experiments has indicated a neutrophil transcriptional signature that is associated with GlycA levels [76].

1.2 Examined pathologies

1.2.1 Human T-Cell Leukemia Virus Type 1

HTLV-1 is the first discovered human pathogenic retrovirus, and was first isolated from a patient with T-cell malignancy in 1980 [95]. A second strain of the virus was identified in 1982 and dubbed HTLV-2 [96]. The addition of the viral causative agent of acquired immunodeficiency syndrome (AIDS) as HTLV-3 to the same group warranted a name change from Leukemia to Lymphotrophic virus. However, the official nomenclature of the AIDS virus changed from HTLV-3 to human immunodeficiency virus (HIV) the 'lymphotropic' term in the name was deemed inappropriate for the remaining viruses in the HTLV group,

and its nomenclature has been restored to its original identity of the original 'leukemia' virus, emphasizing its extremely pronounced oncogenic properties [97].

Recent estimates about the global burden of HTLV-1 range between 5-10 million infected individuals [98], mostly limited to certain endemic regions. However, as systematic population screening is lacking, true infectious burden could be manifolds greater [99]. High prevalences have been identified in Southwest Japan, sub-Saharan Africa [100], Iran [101], Central Australia [102], the Melanesian islands and South America [103]. Endemic hotspots like the city of Salvador, Brazil and Tsushima Island, Japan have a prevalence of 1.8% and 36.4% respectively [99], [104] and are visualized in Figure *2*. Note that for China and India, which together make up about a third of the world's population (United Nations Department of Economic and Social Affairs, 2018), no information is available.

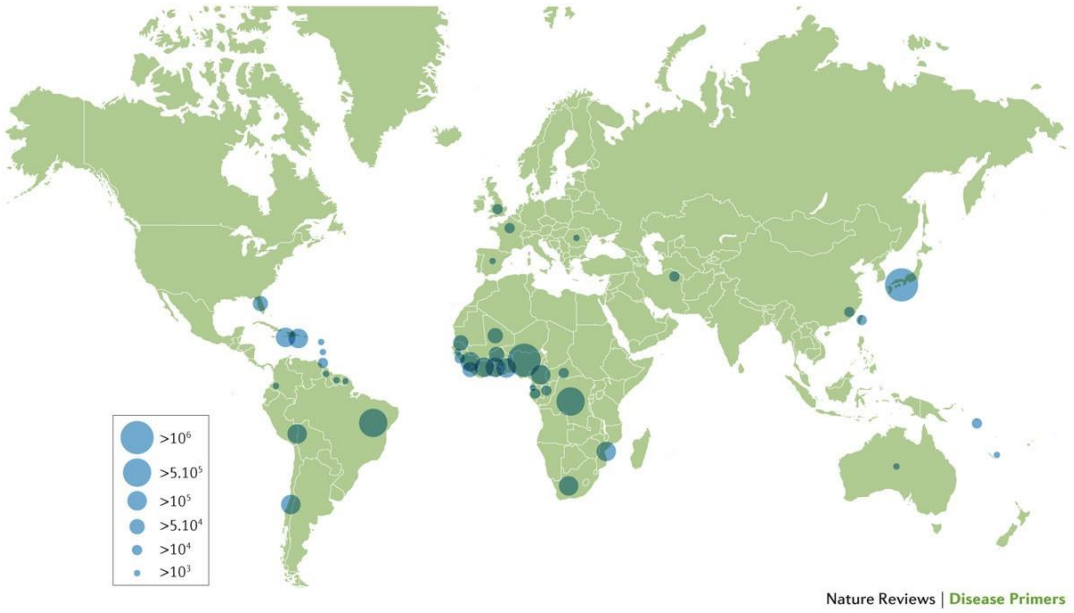


Figure 2: World map with indicated prevalence of HTLV-1 infection. Depicted are estimates of the number of HTLV-1 infected carriers, based on approximately 1.5 billion individuals from endemic regions. (Image from [105], reused with permission.)

Transmission of HTLV-1 primarily occurs via sexual contact [106], mother-to-child transmission through breastfeeding [107], [108] and blood-to-blood contact e.g. by using contaminated needles [109], transfusion of cellular blood products [110] and organ transplantation [111]. When receiving HTLV-1 infected blood by transfusion, the risk of seroconversion is between 40% and 60% with the risk of infection being higher when

transfusing fresh whole blood or platelets (plasma is not infectious) [112]. Some endemic countries screen blood and tissue donors, as well as pregnant women, for HTLV-1 infection but globally, HTLV-1 screening is largely absent [110].

In total, seven genetic subtypes (named 'a' through 'g') are recognized, with different geographic spreads. The cosmopolitan subtype is HTLV-1a, whereas HTLV-1b, -d, -e, -f and -g are African subtypes which have been identified predominantly in Gabon, Congo, and Cameroon, while subtype -c is the Australo-Melanesia subtype [113]–[116]. The within-subtype diversity of HTLV-1 is low when compared to what is observed in HIV-1 and HTLV-1 transmitted within families is highly conserved [105], [117], [118].

HTLV-1 is classified as a group VI virus, meaning its genome is a single + strand of RNA. All retroviral genomes consist of at least 4 genes: *gag*, *pro*, *pol* and *env*. These genes encode proteins related to the assembly of viral-like particles, the maturation of the viral particles, proteins related to viral DNA synthesis and integration in host DNA, and the structural and surface proteins of the viral envelope, respectively. Aside from these typical retroviral genes, the pX region of the HTLV-1 genome encodes regulatory proteins Tax, Rex, p12, p13 and p30 plus the HTLV-1 basic leucine zipper factor (HBZ) as a minus strand protein. Of these, Tax and HBZ have particularly important, but opposite, roles in the pathogenesis and persistence of the virus [105], [119]. Tax is capable of activating transcription of the provirus and many host genes, chiefly CD25, IFN-γ and intercellular adhesion molecule 1. Together, these functions allow the virus to influence cellular activation, proliferation and cell cycle checkpoints, while also inhibiting DNA repair [120]–[122]. Furthermore, Tax persistently activates nuclear factor kappa B (NF-KB) canonical and non-canonical pathways which results in deregulated expression of a large array of genes [123]. In contrast, HBZ limits the effects of the Tax protein, but still drives proliferation of HTLV-1 infected cells [119], [124].

The *in vivo* reservoir for HTLV-1 is CD4+ T-cells, harboring ~95% of the pro-viral load, though the virus is capable of infecting most nucleated cell types (e.g. B cells, NK cells, dendritic cells, monocytes/macrophages, endothelial cells and hematopoietic stem cells [125], [126]). It is then unsurprising that HTLV-1 is the causative agent for a wide range of pathologies, though the reason why most HTLV-1 infected individuals remain asymptomatic throughout their life is poorly understood. Those HTLV-1 infected individuals that progress to a severe disease state mainly develop the blood cancer adult T-cell leukemia/lymphoma (ATL) and the neuroinflammatory disease HTLV-1-associated myelopathy/tropical spastic

paraparesis (HAM/TSP). Other related pathologies include an intraocular inflammatory disorder termed HTLV-1 uveitis and the severe, exudative, infective dermatitis associated with HTLV-1 (IDH) [118]. However, a myriad of other inflammatory conditions that have been linked to HTLV-1 including pulmonary disease, polymyositis, arthritis, inflammatory myopathy, polyneuropathy, motor neuron disease, dysautonomia and non-neurological manifestations Sjögren's syndrome, pulmonary alveolitis, conjunctivitis, sicca syndrome and interstitial keratitis, Hashimoto's thyroiditis, and Graves' disease [118], [127], [128]. Thus far, clinical manifestations of disease have not been successfully linked to particular genotypes of the virus [129], [130]. This suggests that the varied disease manifestations could be attributed to variations in host genetics [105] or to variations in host immunological condition [131]–[135].

1.2.1.1 Adult T-cell leukemia/lymphoma

ATL is an aggressive T-cell leukemia, predominantly of CD4+CD25+ T cells. The disease develops after a long latency period post HTLV-1 infection, with typical diagnoses occurring between the age of 40 to 60 years. ATL can be classified into four subtypes based on their symptoms and prognosis: acute, lymphoma, chronic and smoldering subtypes making up 60%, 20%, 15% and 5% of ATL cases, respectively. The two most aggressive subtypes, i.e. acute and lymphoma, are characterized by their heavy tumor burden with lymphadenopathy, hepatosplenomegaly, as well as skin and visceral lesions in addition to opportunistic infections due to immunodeficiency. The difference between acute and lymphoma subtypes is found in the degree of infiltration of leukemic cells into peripheral blood and hypercalcemia: lymphoma patients have less than 1% of leukemic infiltrates in the blood and are usually less burdened with hypercalcemia. Chronic and smoldering subtypes are typified by 1-5% of peripheral leukemic blood lymphocytes but with an absence of both visceral involvement and hypercalcemia [136], [137].

While confirmation of a diagnosis has been simplified since the discovery of CADM1/TSLC1 as a bona fide cell surface marker for ATL cells [138], [139], prognosis remains poor: the cancer is incurable, with acute ATL patients having a mean survival time of <6 months, and is refractory to current combination chemotherapy. Reported viable treatment options are subtype-specific and include allogeneic hematopoietic stem cell transplantation and monoclonal antibodies to CC chemokine receptor 4 [140]. However, aggressive leukemic ATL

is, according to international consensus following meta-analysis, best treated with IFN-α and zidovudine [140]–[142]. This combination of an immunomodulatory biologic and antiretroviral medication has shown some effectiveness in multiple uncontrolled studies, extending patients' lifespans by 3 to 11 months [143]. In addition to the cancer, patients often suffer from and die of opportunistic infections due to a severe immuno-depressed state.

1.2.1.2 HTLV-1 Associated Myelopathy

HTLV-1 Associated Myelopathy / Tropical Spastic Paraparesis was first described in the nineteenth century [105], though its etiology was unknown until the 1985 when HTLV-1 infection was identified as the cause [144], [145]. In an early inflammatory phase of the disease, $CD4^+$ and $CD8^+$ T cells (some carrying the HTLV-1 infection) invade the spinal cord and produce inflammatory factors and cytokines. The resulting chronic inflammation leads to atrophy of the spinal cord, though precise mechanisms have not yet been elucidated. The neuroinflammatory disease leads to urinary symptoms, back pain and paralysis of the legs. No curative treatment is available for HAM/TSP, and clinical care is focused on symptomatic treatment. Several drugs, like corticosteroids, IFN and antiretrovirals have been tested and have shown some beneficial symptomatic effects and decreased viral loads (reviewed in [105]). Immunomodulatory drugs in particular seem effective in early disease, but their usefulness tapers off in later stages when the neurodegeneration predominates.

1.2.2 Inflammatory Bowel Disease

Inflammatory Bowel Disease (IBD) is a term used to describe chronic inflammatory disorders of the gastrointestinal tract. The two primary types of IBD are Crohn's Disease (CD) and Ulcerative Colitis (UC). Though many clinical symptoms of IBD are common to both diseases (e.g. diarrhea, cramps, abdominal pain, rectal bleeding and weight loss), CD and UC are distinct pathologies. CD affects both small and large intestine, as well as other parts of the gastrointestinal tract like the esophagus and stomach, whereas UC primarily affects the colon and rectum. The intestinal mucosa are continuously subjected to external stresses originating from microbial antigens and dietary metabolites. As such, defects in this innate immune system can lead to persistent immune activation, a key characteristic of IBD [146]. These diseases are characterized by periods of disease activity and remission with pathogenesis ascribed to a combination of host genetics, immune dysregulation and gut microbiota [147].

Biomarkers allowing for differential diagnosis, assessment of disease activity and severity and tracking of therapy outcome are an active research field, and currently no single gold standard exists [56], [148]. It is worth noting that C-reactive protein (CRP), the most commonly used biomarker for acute inflammation, is poorly suited to these tasks in IBD: the variability of CRP in CD patients is considerable and UC patients are often noted to have only a modest or even absent CRP response [56], [149]. This heterogeneity is peculiar considering concentrations of e.g. interleukin (IL)-6, IL-1β and TNF-α (inflammatory proteins discussed in Section 1.1.2) are detected in UC even when CRP is not. Genetic studies on CRP polymorphisms have thus far failed to explain this heterogeneity [150].

Treatment options include dietary interventions [151], corticosteroid usage and the sectioning of the GI tract, though TNF-α blockade by neutralizing antibodies, like the biological infliximab, has shown to be an effective strategy in some subgroups of patients, paving the path for trials with many other biologicals targeting different cytokines, with varying degrees of success [152].

1.2.3 Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a relatively rare chronic autoimmune disease observed mostly in young women. The disease is characterized by repeated disease activity flares of varying severity: 80% of patients with active disease show either persistent disease activation or frequent disease flares [153]. These disease flare-ups can cause severe visceral damage. For example: while renal impairment is observed in approximately 15% of patients at SLE diagnosis in Europe, over 40% of SLE patients experience kidney damage over the course of their disease and the risk of end-stage renal failure after 10 years in patients with severe renal involvement is estimated at 10-30% [154]. Moreover, the 10-year mortality attributed directly to SLE is estimated at less than 10% [155]. Nonetheless, the treatment of SLE is based on the use of immunosuppressive therapies with significant long-term side effects (infections, cardiovascular pathologies, cancers, osteoporosis) [156] that remain a relevant concern even for novel biologicals as recently shown by my collaborators [157].

The clinical presentation of SLE is polymorphic and includes heterogeneous manifestations ranging from renal and neurological damage to anemia, polyarthritis and thrombosis. Biologically, manifestations of SLE are characterized by the presence of hypocomplementemia, anti-nuclear factors, anti-double-stranded DNA antibodies and/or

anti-nucleosome antibodies. In the lapses between active disease, i.e. when the disease is quiescent, biological surveillance involves regular screening for urinary abnormalities which may indicate a recurrence of kidney disease and repeated quantification of the complement and anti-DNA antibodies. However, to date, there are no biomarkers to assess/predict lupus disease activity and thus personalize the level and duration of immunosuppression. This lack of biomarkers means treatment is often continued in clinically quiescent patients for fear of a new outbreak of the disease for extended or even indefinite duration. The resulting iatrogenic effects are justified by the fact that once treatment is stopped or reduced, the diagnosis of a relapse of SLE often comes too late; when the organs have already sustained significant damage. Indeed, the biomarkers in current clinical use, i.e. clinical-biological activity scores, determination of complement fractions, native anti-DNA antibody titer and renal puncture biopsy, have essentially been validated only for the diagnosis of confirmed disease flare and to determine their severity. The need for predictive biomarkers of SLE flares is dire and has been internationally highlighted [158].

1.3 Transcriptomic analysis

Transcriptomics is the term describing the examination of the expression levels of a comprehensive set of genes in a sample. This data is typically generated by fluorescent microarrays or by RNA-sequencing (RNA-seq). While extensive repositories for microarray studies exist (like the Gene Expression Omnibus, GEO [159]), RNA-seq is now commonplace and is typically performed for new differential expression studies [160].

Microarrays were developed in the mid-1990s and were the first high-throughput technology that enabled quantification of gene expression levels [161]. Microarray chips contain oligonucleotide probes which are specifically designed to target known gene sequences. These probes are then laid out in specific grid-positions on a chip so that the fluorescent light resulting from the hybridization between the sample's complementary RNA (created by transcription from complementary DNA which was generated in turn by reverse transcribing the sample RNA) and the probes can be attributed to the correct gene of origin.

RNA-seq is replacing traditional microarray technologies, owing to technological advances and concurrent decreases in the costs of sequencing technologies. RNA-seq has many advantages: the technology measures all RNA in a sample, not just those belonging to known gene sequences, and alternative splicing events (where exons of a gene are combined to

generate alternative versions of a protein) can also be identified [162]. The core goal of expression quantification also benefits from RNA-seq's greater dynamic range: the technology better captures changes in transcripts with very low expression without giving up accuracy in highly expressed genes [162]. Analysis of RNA-seq data is not without its challenges as many different analysis pipelines for the raw data can be employed, each with its own advantages and disadvantages.

Traditional analysis of the sequence reads generated in a transcriptomic experiment generally starts with a pre-processing where low quality reads and adapters are removed from the data [163], [164]. The high-quality reads are then aligned to a reference genome by a splice-aware aligner before the reads covering each gene are quantified [165]. The latest generation of RNA-seq analysis software employs a different strategy for quantification: pseudoalignment. Traditional alignment strategies try to find the position where a given read originated from, pseudoalignment instead generates a list of all possible transcripts which could have generated the read without caring exactly where within the transcript the read originated [166]. These techniques are orders of magnitude faster than traditional alignment based methods without sacrificing quantification accuracy [166].

1.3.1 Modular Analysis of Transcriptomic Data

Transcriptomic analysis frequently takes the form of associative studies between traits of interest and the expression of specific genes. However, correlation networks are increasingly used to describe the complex correlational patterns observed across the expression of genes measured in transcriptomic experiments. Weighted Gene Correlation Network Analysis (WGCNA) is a widely used R software package which identifies clusters or 'modules' of highly correlated genes and summarizes the expression of such modules using their eigengene or intramodular hub genes [167].

Briefly, WGCNA starts by generating an adjacency matrix, using robust biweight-midcorrelation to calculate correlation between expression of all genes which are then subjected to a soft-thresholding procedure so that the resulting correlational network approximates the scale-free network criterion. This criterion states that the majority of genes within a network correlate to very few other genes, while a small subset is highly interconnected. From this adjacency matrix, the topological overlap measure for each gene pair is calculated, which quantifies how many gene-correlates are common to both members

of each gene pair. Advanced dynamic tree cut algorithms are then used to identify coherent clusters or 'modules' of genes. Each resulting module can then be summarized by the expression value of its virtual 'eigengene', i.e. its first principal component, which can be interpreted as a weighted average expression value, so named because of an algebraic eigenvector's ability to essentially summarize a matrix. Alternatively, if virtual gene expression is undesirable, a module's hub genes can be identified: these are genes whose intramodular connectivity is highest (and are therefore highly correlated to the module's eigengene expression) [167]. The gene modules identified by WGCNA can then be annotated manually or by using standard Gene Ontology (GO) approaches using either Fischer overrepresentation analysis (ORA), or gene set enrichment analysis (GSEA) methodologies [168].

1.4 Predictive Models

The centerpiece of this thesis is the construction of a model which estimates chronic inflammatory burden from gene expression. The methodology used in this process is introduced and detailed in Chapter 4. However, these modelling approaches can also be used to assess the value of new biomarkers, as I did for the GlycA biomarker in the IBD and SLE disease contexts (Chapters 5 and 6).

Assessing the performance of a biomarker in a binary classification model is routinely done using receiver operating characteristic (ROC) curves which shows the model's relation between its sensitivity or true positive ratio (TPR) and its false positive ratio (FPR) or 1 minus its specificity. Though the area under the ROC curve (AUC or c-statistic) is often used as a performance metric in the literature, this measure has important limitations: models with high baseline accuracy are hard to improve upon and meaningful improvements to model accuracy may have comparatively small impacts on the c-statistic [169], [170]. Furthermore, in real applications a classification model with high false positive rates might be unacceptable, but the c-statistic summarizes performance across all false positive rates. One alternative to the c-statistic is to report a model's performance as its leave-one-out cross-validation accuracy and Cohen's Kappa value for relative improvement between models (either over an existing model or one based on random chance). This method is to be preferred, provided the considered sample accurately reflects the population it's drawn from [171], [172], although it is not without its own limitations and other alternatives have been suggested [172], [173].

CHAPTER 2

IFN-β induces greater antiproliferative and proapoptotic effects and increased p53 signaling compared with IFN-α in PBMCs of Adult T-cell Leukemia/Lymphoma patients.

Letter to the editor

Published in *Blood Cancer Journal,* 2017, 7, e519,
Dierckx T., Khouri R., Menezes S., Decanine D., Farre L., Bittencourt A., Vandamme A-M., Van Weyenbergh J. (2017). **IFN-β induces greater antiproliferative and proapoptotic effects and increased p53 signaling compared with IFN-α in PBMCs of Adult T-cell Leukemia/Lymphoma patients.**

doi: 10.1038/bcj.2016.126.
PMID: 28128792

Abstract

Current first-line treatment for Adult T-cell leukemia (ATL) includes combination therapy with interferon alpha (IFN-α) and zidovudine (AZT). The use of IFN-α in this treatment is mostly empirical in origin, whereas the therapeutic potential of interferon beta (IFN-β), has not yet been thoroughly explored in this context. Here we compare the effects of IFN-α and IFN-β treatment in short term *ex vivo* peripheral blood mononuclear cell (PBMC) cultures from 22 ATL patients. Using proliferation, apoptosis and antiviral bioassays, complemented with microarray and gene set analysis, we demonstrate that in this setting IFN-β has superior antiproliferative and pro apoptotic effects than IFN-α. Increased p53 signaling is observed under the IFN-β treatment, while the antiviral effects are equivalent to those of IFN-α treatment. Notably, the genes in a published in vivo AZT/IFN-α response profile are affected more strongly by IFN-β than by IFN-α stimulus in these *ex vivo* PBMCs. In conclusion, this first comprehensive analysis comparing the effects of IFN-α and IFN-β on *ex vivo* primary cells from ATL patients demonstrates that IFN-β has a greater impact than IFN-α on biological processes which have been shown to be crucial in the treatment of ATL, making IFN-β an intriguing candidate for further in vivo testing.

Adult T-cell leukemia/lymphoma (ATLL) is an aggressive T-cell malignancy with a poorly understood pathology that manifests from human T-lymphotropic virus type I (HTLV-1) infected T-cells, typically after long latency periods (>30 years). Current treatment regimens for ATLL, reviewed in [174], include zidovudine (AZT) and interferon alpha (IFN-α) combination therapy. The established usage of IFN-α in the treatment of ATLL is largely empirical in origin. The effects of the other widely used IFN subtype, IFN-β, have not been thoroughly examined in this setting, even though IFN-β's more potent induction of antiproliferative and apoptotic pathways has been described in solid cancers [175]. Few studies have compared the effects of IFN-α versus those of IFN-β in ATLL and these were generally performed in cell lines, rather than primary patient cells [176], [177]. *In vitro* experiments with IFN-α using cell lines report minimal effects of IFN-α on the viability of HTLV-1 infected T-cells and viral replication, in contrast to the high *in vivo* response rates obtained by this therapy [178], [179]. Reports regarding *in vitro* and *in vivo* activity of IFN-β in this context also seem contradictory: *in vitro* experiments showed no antiproliferative action of IFN-β [180] while a single early *in vivo* trial using IFN-β reported promising results, with 50% of 6 treated patients achieving partial response to treatment using IFN-β monotherapy [181]. The reported success rate of this IFN-β monotherapy was comparable to early AZT/IFN-α combination therapy trials where 67% of 24 treated patients achieved partial response [182].

We performed the first direct comparison between the response to IFN-α and IFN-β in *ex vivo* ATLL patient PBMCs. We analyzed samples obtained between 2001 and 2007 from 9 men and 13 women aged 21-78 years (median 47.5), diagnosed as HIV negative and definite ATLL with serology, inverted PCR and/or flow cytometry, at the "Hospital Universitário Professor Edgar Santos" (HUPES) in Salvador, Bahia, Brazil. Seven of these patients were classified as acute, ten as smoldering, three as chronic and two as lymphoma according to Shimoyama criteria [136]. This study was approved by the Ethics Review Board of HUPES (number 32050106). Data handling and processing was additionally approved by the Medical Ethics Commission of the UZ Leuven hospital, Belgium (number s57931).

Proliferation, antiviral activity and apoptosis were all measured in three distinct treatment conditions: cultures were either left untreated or stimulated with either IFN-α (1000 U/ml) or IFN-β (1000 U/ml) at the start of the experiment. Bioactivity of IFN-α and IFN-β was determined using a VSV/Wish bioassay in order to preclude any bias owing to different antiviral effects of the interferon subtypes. Neither IL-2 nor PHA were added to the *ex vivo*

cultures in order to approximate *in vivo* conditions as closely as possible. Proliferation was measured by [^{3}H] thymidine incorporation assay in the cultures of *ex vivo* PBMCs of 19 patients. Active caspase-3 was then measured by flow cytometry (FACSort, BD Biosciences, Franklin Lakes, NJ) using a CBA apoptosis kit (BD Biosciences). HTLV p19 protein levels in PBMC 48-hour culture supernatants were measured using the HTLV-I/II p19 antigen ELISA (ZeptoMetrix, Buffalo, NY), according to the manufacturer's instructions. Detailed methods are provided as supplementary materials.

Unless otherwise noted, Bonferroni corrected, nonparametric Friedman rank sum tests were used to test for statistically significant differences between the three experimental conditions. The results of these tests are summarized in Figure 3. IFN-α caused a small but significant 24±36% decrease in proliferation in the nineteen examined samples, whereas IFN-β treatment decreased proliferation significantly by 47±58%. Direct comparison of IFN-α vs IFN-β treatment conditions shows that IFN-β exerted superior antiproliferative activity. Caspase 3 activation, measured in six samples, showed an increase in apoptosis for both IFN subtypes, but IFN-β showed a significantly higher increase in apoptosis than IFN-α (12.8±7.2 and 4.9±7.1 pg/ml, mean±sd, respectively). Fourteen out of 16 tested samples had detectable virus production in the supernatants of 48-hour cultures. Viral p19 levels varied strongly between patient samples, ranging from 4.8 to 10792.7 pg/ml (mean±sd, 2131.3±3796.9) in the control condition. Both IFN-α and IFN-β treatments resulted in comparable reductions in viral p19 levels when contrasted with the untreated control condition (a mean±sd decrease of 49±32% versus 69±70%), suggesting that the observed differential effects of the two IFN types on proliferation and apoptosis do not stem from a differential impact on viral replication.

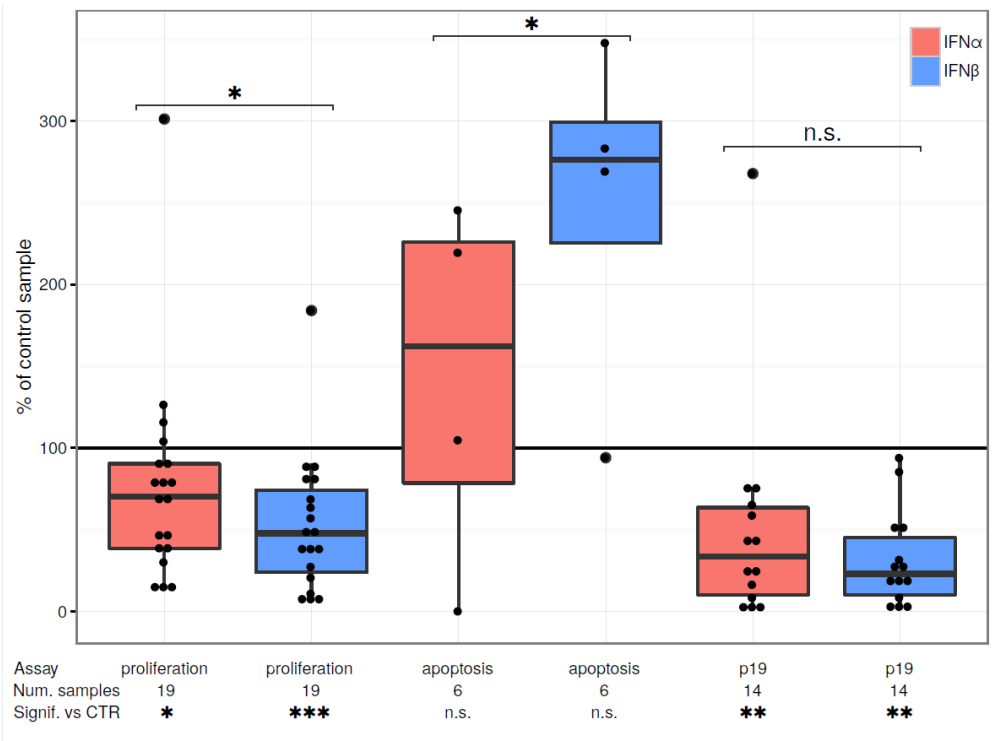


Figure 3: Boxplots of the effects of IFN-α and IFN-β on measured proliferation, apoptosis and viral protein p19 production in ex vivo PBMCs of ATL patients. Viral protein was quantified using ELISA, proliferation by [3H] incorporation and apoptosis through Flow cytometry of active caspase-3. Each sample was treated in parallel in three different conditions: either left untreated, stimulated with 1000 IU of IFN-α or IFN-β (red and blue, respectively). Data is depicted here as the percentage of the value measured in the corresponding untreated control condition of the sample. Statistical significance of the Friedman rank sum test (* = p<0.05, **= p<0.01, ***= p<0.001) and number of samples used in the comparison of each condition versus its control (CTR) is indicated underneath the graph. Statistical significance of the difference between IFN-α and IFN-β is depicted above the boxplots. All datapoints are shown in the graph, excepting two samples that had no measurable apoptosis in the untreated control condition. These could not be included in this graph, but have been used in the statistical comparison of significance versus control.

Sample quantities did not allow for the purification of leukemic cells, so their relative contribution to the composition of the examined PBMCs could not be determined. Additionally, at the time of patient recruitment and sample processing for *ex vivo* experiments, TSLC1/CADM1 had not yet been described as a reliable flow cytometry marker for HTLV-1 infected and ATLL leukemic cells [138], [183]. As such, we could not determine whether the observed differential effects of IFN-α and IFN-β take place in leukemic, HTLV-1 infected non-leukemic or HTLV-1 negative cells. However, in a re-analysis of the data comparing the effects of IFN-α and IFN-β on proliferation in ATLL subtypes with high percentages of circulating ATLL cells (i.e. acute and chronic subtypes, n=8), the differences remained significant (p=0.004). In contrast, patients with low percentages of circulating ATLL cells (i.e. lymphoma and smouldering subtypes, n=11) did not show these significant differences

(p=0.36), suggesting that the differential effects occur in the leukemic cells. Viral p19 protein measurements show no significant differences between IFN-α and IFN-β treatments in these subgroups (p=0.32 and p=0.71 for high and low ATLL cell percentage subgroups, respectively).

These findings were complemented with microarray analysis. Both preprocessed and raw data from the microarray experiments are available at the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE85487. Paired differential expression analysis comparing control, IFN-α and IFN-β treated samples of six patients shows that the IFN-β response can be regarded as both broader and stronger than the IFN-α response: all but two of the significantly IFN-α regulated genes were regulated more strongly by IFN-β (one-sided, paired t-test, $p<0.005$).

Although the mechanism behind the AZT/IFN-α treatment in ATLL has not been fully elucidated [179], [184], recent research pointed to the protein kinase R (PKR or EIF2AK2) gene as a critical gene in the antiviral response and the activation of the p53 pathway and the induction of apoptosis as key components in its mechanism [185]. We report a superior effect of IFN-β on all of these processes. First, our microarray results show that IFN-β affects PKR gene transcription more strongly than IFN-α (21% vs 15% increase, p=0.04). Second, Gene Set Analysis (GSA) shows that the enrichment of the canonical apoptosis pathway that is prominent in the IFN-β condition is absent in the IFN-α condition (Figure 4). Third, while p53 gene expression (TP53) is not significantly affected by either IFN-α or IFN-β treatment, in agreement with the post-transcriptional stabilization of p53 by AZT [186], GSA shows that the p53 pathway is more strongly activated by IFN-β than by IFN-α (Figure 4). Finally, in agreement with the results of Kinpara et al. [185], GSA of the IFN-α treatment microarray results revealed a modest downregulation of the NF-κB pathway, which has been identified as integral to the ATLL transcriptome in a recent integrated omics analysis in a large number of patients [187]. In contrast to the IFN-α effects, IFN-β treatment upregulated NF-κB pathway activation. Combined, these results suggest that IFN-β has a greater impact on the crucial elements of the AZT/IFN-α treatment mechanism than IFN-α, already making a strong case for possible clinical trials using IFN-β or AZT/IFN-β. But perhaps the clearest argument in favor of clinical trials using IFN-β in ATLL is that most of the genes in the *in vivo* AZT/IFN-α response gene set reported by Alizadeh et al. [184] respond more strongly to IFN-β mono-stimulus than to IFN-α mono-stimulus in these *ex vivo* PBMCs.

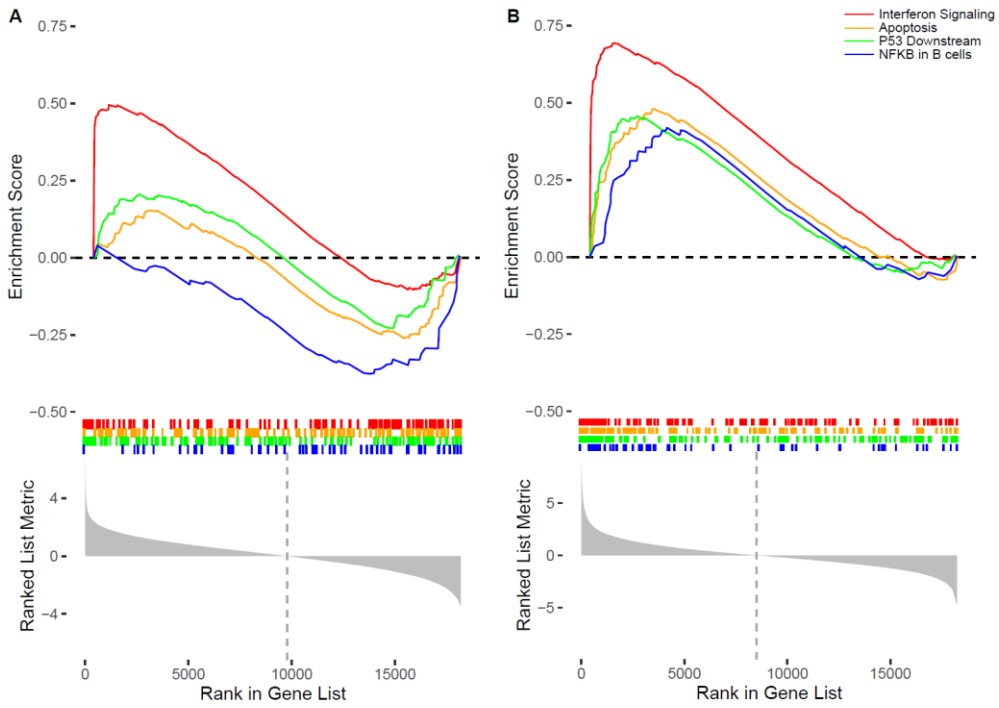


Figure 4: Selected Gene Set Enrichment Analysis results. The impact of IFN-α treatment (A) and IFN-β treatment (B) on four selected gene sets relevant to ATL associated with interferon signaling (red), apoptosis (yellow), p53 signaling (green) and NF-κB signaling (blue). The top graph depicts the running Enrichment Score (ES) graphed versus the rank of the gene when ordered by the t-statistic of their differential expression (Untreated vs IFN treatment) result, which is depicted at the bottom.

Other preclinical cancer models have previously shown superior antiproliferative and pro-apoptotic effects of IFN-β when compared to IFN-α [175], [188] but to our knowledge this is the first time that these differential effects have been reported in *ex vivo* PBMCs of leukemia/lymphoma patients. Due to three strong similarities with observations made by Sancéau et al. in a preclinical Ewing Sarcoma model [188], we hypothesize that other leukemias with a functional or wild-type p53 could also be more sensitive to the pro-apoptotic effects of IFN-β as compared to IFN-α. First, the four cell lines examined in their study proved to be more susceptible to the antiproliferative effects of IFN-β than to those of IFN-α, similar to the PBMCs of ATLL patients used in the present report. Second, IFN-β, but not IFN-α, induced apoptosis in the two cell lines with wild-type p53 but not in those with mutant p53 [188] which is in line with p53's mutant status as a predictor of poor response to AZT/IFN-α treatment [186] and the p53-dependence AZT/IFN-α-induced apoptosis [185]. Third, Sancéau et al. showed that IFN-β-induced apoptosis was mediated by IRF1 [188], which is also associated with AZT/IFN-α treatment response in ATLL [184].

In conclusion, we performed the first comprehensive analysis comparing the effects of IFN-α and IFN-β on *ex vivo* PBMCs of ATLL patients. Our observations suggest that IFN-β is a worthwhile candidate for clinical trials in ATLL and is also worth investigating in other leukemic contexts where IFN-α has shown modest success.

Supplementary Materials and Methods

Ethics

This study was approved by the Ethics Review Board of "Hospital Universitário Professor Edgar Santos" (registration number 32050106), according to the principles of the Declaration of Helsinki, and all individuals included in this report signed an informed consent form before enrolment. Data handling and processing was additionally approved by the Medical Ethics Commission of the UZ Leuven hospital, Belgium, under registration number s57931.

Patient recruitment, diagnosis and treatment

Between 2001 and 2007, a total of 30 leukemia patients were recruited from the "Hospital Universitário Professor Edgar Santos" (HUPES, Federal University of Bahia), inclusion and exclusion criteria have been previously described [189]. All cases were confirmed as HIV negative and 26 were diagnosed as clinically definite ATLL according to [136], with serology, inverted PCR and/or flow cytometry carried out as previously described [190]. Blood samples were obtained from 22 individuals. Of the examined 22 patients, 7 were classified as acute, 10 as smoldering, 3 as chronic and 2 as lymphoma. The male:female ratio was 1:1.44. The median age was 47,5 years, with a range of 21 to 78 years. Samples were obtained before treatment, survival at five year follow-up data was available for 21 patients, with one lost to follow-up. Patient treatment was in agreement with published international consensus [137] with "watchful waiting" for smoldering forms followed by treatment upon disease progression, IFN+AZT combination therapy for chronic/acute forms and chemotherapy for lymphoma patients. Demographical and clinical details for all patients as well as full experimental assay details can be found in Supplementary Table S1. The clinical details of one patient in this data set have previously been described in a case report [191].

IFN-a and -β bioactivity determination

A single batch of clinical grade IFN-α2A ($3x10^6$ IU/ml, a gift from Blausiegel Farmacêutica, São Paulo, Brazil) and clinical grade IFN-β1a ($1x10^6$ IU/ml, Biogen, Cambridge, Massachusetts, U.S., a gift from Dr. D. Brassat, Toulouse, France) was used throughout the study to eliminate any possible variation in bioactivity during the study period. Stock solutions were prepared in sterile saline and working solutions in RPMI 1640 medium, supplemented with 10% heat inactivated fetal calf serum, 20 µg/ml gentamicin (GIBCO® Invitrogen, Belgium). Bioactivity of IFN-α and IFN-β was determined according to WHO guidelines in

order to preclude any potential bias owing to the different antiviral effects of the two interferon types. The antiviral activity of both IFNs was measured against Vesicular Stomatitis Virus (VSV) in Wish cells and showed no statistically significant differences, in agreement with the reports of Sancéau et al. [188].

Treatment conditions

Antiviral activity, proliferation and apoptosis were measured as described below in three distinct treatment conditions: either left untreated or stimulated at the start of the experiment with either IFN-α (1000 U/ml) or IFN-β (1000 U/ml), as in [192]–[194]. Neither IL-2 nor PHA was added to the *ex vivo* cultures so that *in vivo* conditions are approximated as closely as possible.

Proliferation assay

Peripheral blood mononuclear cells (PBMCs, 1 x 10^6 cells/ml) were plated in 96-well U bottom plates in RPMI + 10% fetal calf serum (FCS), 200 µl/well in the three conditions and left for 5 days at 37°C and 5% CO2 as in [195]. Lymphoproliferation was quantified by [^{3}H] thymidine incorporation after a 12-16h pulse with [^{3}H] thymidine (1 µCi/well). Incorporation of radioactive label was measured by gas phase scintillation (Direct Beta Counter Matrix 9600, PerkinElmer Life Sciences, MA). Results are expressed as the mean counts per minute in triplicate cultures.

Apoptosis assays

PBMCs (1-2 × 10^6 cells/ml) were plated in 24-well plates in RPMI + 10% FCS, 1 ml/well in the three treatment conditions, for 48h at 37°C and 5% CO2 as in [192], [196]. Apoptosis resistance of the samples was tested through etoposide and serum starvation (1% FCS) treatment. Active caspase-3 was measured by flow cytometry (FACSort, BD Biosciences, Franklin Lakes, NJ) using a CBA apoptosis kit (BD Biosciences).

Quantification of HTLV-1 p19 expression

HTLV p19 protein levels in PBMC 48h culture supernatants were measured in the three treatment conditions, using the HTLV-I/II p19 antigen ELISA (ZeptoMetrix, Buffalo, NY), according to the manufacturer's instructions.

Statistical analysis

Comparison of proliferation, apoptosis and viral protein expression assay results between the different treatment conditions was performed using the nonparametric Friedman rank sum test. Unless otherwise noted, reported p-values for these tests were corrected for multiple testing using the Bonferroni method.

Microarray analysis

Total RNA was extracted according to manufacturer's protocol (QIAgen, Benelux B.V. Venlo, Netherlands) from a total of 20 samples: six patients' parallel 48h cell cultures in the three treatment conditions plus two additional patients' untreated control samples. Whole genome microarray was performed at the VIB Nucleomics Facility (Leuven, Belgium) using the GeneChip Human Gene 1.0 ST Array with the WT PLUS reagent kit (Affymetrix, Santa Clara, CA), according to manufacturer's instructions. Data were RMA preprocessed in R using the Bioconductor oligo package [197]. Differential expression analysis was then performed using the Bioconductor limma package [198]. A moderated, paired t-test was used to determine differential transcript expression between the three conditions. Genes were accepted as differentially expressed in a condition if their Benjamini-Hochberg corrected p-value was lower than 0.05. Both the preprocessed and raw data from the microarray experiments is available at the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE85487.

Gene Set Analysis (GSA)

GSA was performed in R using the platform for integrative analysis of omics data (piano) package [199]. This package implements a range of GSA methods, including Gene Set Enrichment Analysis (GSEA) [168], and provides a consensus score of the GSA results. The C2 canonical pathways and the C5 Gene Ontology Biological Processes gene sets were collected from the Molecular Signatures Database (MSigDB) and tested for enrichment.

CHAPTER 3

Decreased RORC expression and downstream signaling in HTLV-1-associated adult T-cell lymphoma/leukemia uncovers an antiproliferative IL17 link: A potential target for immunotherapy?

Research Article

Published in *International Journal of Cancer,* 2019, 144(7):1664-1675.
Subramanian K.*, Dierckx T.*, Khouri R., Menezes SM., Kagdi H., Taylor GP., Farre L., Bittencourt A., Kataoka K., Ogawa S., Van Weyenbergh J. **Decreased RORC expression and downstream signaling in HTLV-1-associated adult T-cell lymphoma/leukemia uncovers an antiproliferative IL17 link: A potential target for immunotherapy?**

doi: 10.1002/ijc.31922.
PMID: 30303535

* Both authors contributed equally to this work

Abstract

Retinoic acid-related drugs have shown promising pre-clinical activity in Adult T-cell Leukemia/Lymphoma, but RORC signaling has not been explored. Therefore, we investigated transcriptome-wide interactions of the RORC pathway in HTLV-1 and ATL, using our own and publicly available gene expression data for ATL and other leukemias. Gene expression data from ATL patients were analyzed using WGCNA to determine gene modules and their correlation to clinical and molecular data. Both PBMCs and $CD4^+$ T-cells showed decreased RORC expression in four different ATL cohorts. A small subset of $RORC^{hi}$ ATL patients was identified with significantly lower pathognomonic CADM1 and HBZ levels but similar levels of other ATL markers (CD4/CD25/CCR4), hinting at a less aggressive ATL subtype. An age-dependent decrease in RORC expression was found in HTLV-1-infected individuals, but not in healthy controls, suggesting an early molecular event predisposing to leukemogenesis. Genes upstream of RORC signaling were members of a proliferative gene module (containing proliferation markers PCNA/Ki67), whereas downstream members clustered in an anti-proliferative gene module. IL17C transcripts showed the strongest negative correlation to PCNA in both ATL cohorts, which was replicated in two large cohorts of T- and B-cell acute lymphoid leukemia (ALL). Finally, *IL17C* expression in purified $CD4^+CCR4^+CD26^-CD7^-$ 'ATL-like' cells from HTLV-1-infected individuals and ATL patients was negatively correlated with clonality, underscoring a possible antileukemic/antiproliferative role. In conclusion, decreased RORC expression and downstream signaling might represent an early event in ATL pathogenesis. An antiproliferative IL17C/PCNA link is shared between ATL, T-ALL and B-ALL, suggesting (immuno)therapeutic benefit of boosting RORC/IL17 signaling.

Introduction

Human T-Lymphotropic Virus -1 (HTLV-1) is a retrovirus with an estimated prevalence of 10-20 million worldwide [109]. A recent return to the original name of Human T-cell Leukemia Virus-1 [97] is in agreement with its exceptional oncogenicity [200]. Although most HTLV-1 infections are asymptomatic, 2-6% of HTLV-1 infected individuals develop a CD4+CD25+ chemotherapy-resistant and aggressive leukemia known as Adult T-cell Lymphoma/Leukemia (ATL) [183], [201], [202]. ATL presents after a long latency period of the virus, commonly several decades [203]. Patients therefore tend to be older individuals with an average age at diagnosis of 40 years in Central and South America and 60 years in Japan. Depending on the subtype (acute, lymphomatous, chronic, and smoldering), survival ranges from 4 months to over 5 years [189].

HTLV-1 has two viral oncoproteins: Tax and HBZ. Tax benefits cell survival in HTLV-1 infected T-cells by interacting with NFκB [204], a key player in immune regulation. However, Tax levels are undetectable in most ATL patients, either due to gene deletion or altered DNA methylation levels, whereas HBZ is expressed consistently in ATL [204]. HBZ modulates Tax expression and induces CD4+ T-cell proliferation [183]. CADM1/TSLC1 is also consistently expressed in ATL cells, such that CADM1 staining overlaps with the CD4+CD25+ T-cells in ATL and proviral sequences from these leukemic CD4+CADM1+ cells were consistently positive for the HBZ region [183]. Thus, CADM1 is a sensitive biomarker for ATL and might be used to determine treatment efficacy [183], [204].

ATL patients display an increased incidence of opportunistic infections [205], which could be attributed to a deregulation of the Th17 axis, as an intact Th17 response is necessary for the clearance of opportunistic infections [206]–[209]. IL-17 and its upstream regulator IL-6 were increased in long-term cultured Tax+CD4+ T-cell supernatant [210]. IL-17 mRNA was also found to be highly expressed in HTLV-1 infected T-cells and Tax-expressing Jurkat cells [211]. Therefore, we hypothesize Tax-negative ATL cells are unlikely to express IL-17. Induction of the Th17 axis via retinoic acid receptors (RARs) and RAR-like orphan receptors (RORs) could potentially alleviate the increased opportunistic infection frequency caused by Th17 deregulation.

Retinoic acid blocks Th17 differentiation and stimulates regulatory T-cell (Treg) production [212]. Although HTLV-1 proviral integration in the host genome showed greater

enrichment of promotor sequence motifs binding p53 and STAT1 instead of the RORC locus [213], downstream effects of p53 and STAT1 downregulate RORC expression by suppressing the transcription factor STAT3 [214]. The relevance of RORC in leukemogenesis is further supported by the observed increased proliferation and apoptosis rates in mice deficient in the protein product of the RORC gene RORγ, leading to the development of T-cell lymphoma [215] and lymphoblastic lymphoma [216].

Taken together, deregulation of the RORC/Th17 axis can provide an explanation to both the oncogenic persistence of ATL and to patient susceptibility to opportunistic infections. In this study, we generate a representative consensus gene set for the RORC pathway of the Th17 axis and proceed to a multi-cohort analysis of novel and existing data to test the biological significance of this pathway in ATL.

Results

Transcriptomic analysis of four independent cohorts reveals a RORC^{lo} ex vivo phenotype in ATL

Gene expression profiling of *ex vivo* primary cells from ATL patients showed decreased RORC normalized expression in all four independent cohorts, revealing a common RORC^{lo} phenotype (Figure 5A-B-C-D). Japanese Cohort #3 (n=73) and Caribbean Cohort (n=38) had significant decreases in RORC expression of ATL patients ($p<0.0001$ and $p=0.016$ respectively). Japanese Cohorts #1 (n=18) and #2 (n=50) had borderline significant decreases in RORC expression of ATL patients ($p=0.083$ and $p=0.10$). HAM patients in the Caribbean Cohort did not have a significant change in RORC expression ($p=0.54$), however asymptomatic HTLV-1 infected individuals (AC) did display a significant decrease in RORC expression ($p=0.016$) when compared to healthy controls. ACs in other cohorts were not found to have a significant change in RORC expression, relative to healthy controls. Thus, RORC expression, measured as normalized expression (Figure 5) and percentile rank (Supplementary Figure S1), is consistently lower in ATL than in HC, but varies among cohorts for AC.

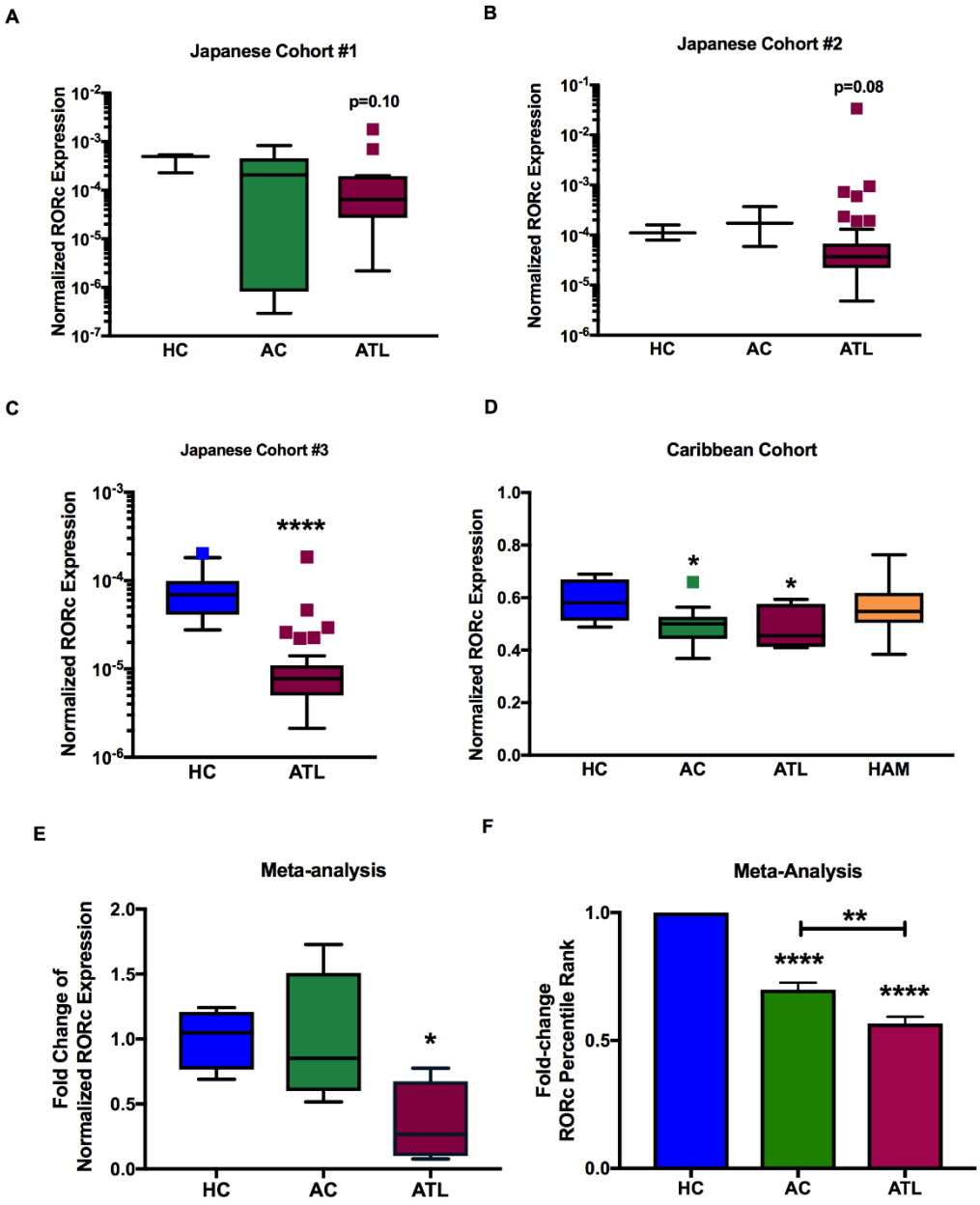


Figure 5: Normalized RORC expression levels for four independent cohorts consisting of ATL patients and healthy uninfected (HC) and/or HTLV-1 infected healthy controls (AC) showed a consistent decreased expression in ATL (A-D). (E) Meta-analysis of RORC normalized expression fold-change by disease status shows a significant decrease in ATL patients but not ACs. (F) Meta-analysis of RORC percentile rank fold-change by disease status shows a significant two-step decrease for ACs and ATL. HCs = Healthy Controls, ACs = Asymptomatic Controls, ATL = Adult T-cell Lymphoma/Leukemia Patients, HAM = HTLV-1-Associated Myelopathy patients. *p<0.05, **p<0.01, ****p<0.0001.

Since RORC gene expression had been previously shown to decrease in AC [217], we performed a meta-analysis of the fold changes in RORC expression in all four cohorts, using

both normalized expression and percentile ranks. Normalized gene expression allows for a comparison of the fold change in absolute RORC levels across cohorts but does not consider the profound perturbation of the cellular transcriptome between healthy vs. leukemic $CD4^+$ cells in AC and ATL patients, respectively. In contrast, percentile ranks are a measure of RORC expression relative to the overall transcriptome for each individual, which is more suitable for a comparison between divergent disease states. This was confirmed by two-way ANOVA, analyzing cohorts and disease status (HC-AC-ATL) as separate variables. For fold-change RORC normalized expression, cohort differences accounted for 21.2% of variation ($p=0.34$) and disease status for 47.2% of variation ($p=0.06$). For fold-change RORC percentile ranks, only 1.3% of variation ($p=0.48$) was explained by cohorts and 95.8% of variation ($p<0.0001$) was explained by disease status. As shown in Figure 5E, RORC normalized expression was significantly ($p<0.01$) decreased in ATL patients, but not AC. Figure 5F displays the median RORC percentile rank fold-change, showing a significant 30% decrease in asymptomatic HTLV-1-infected individuals ($p<0.0001$, vs. HC) and an even further (43%) decrease in ATL patients ($p<0.0001$, vs. HC; $p=0.006$ vs. AC). This two-step decrease in RORC gene expression in ATL pathogenesis, first upon HTLV-1 infection and next upon progression to malignant disease, prompted us to investigate the possible influence of age upon RORC expression.

RORC Expression is not influenced by Age in Healthy Controls but decreases with Age in HTLV-1 Infected individuals

Since ATL usually occurs after several decades of HTLV-1 infection5 and RORγt Tregs were shown to increase with age in mice [218], we investigated the effect of age upon RORC gene expression in healthy controls and HTLV-1-infected individuals from several cohorts. We found that RORC expression significantly decreased with age in HTLV-1 infected individuals without ATL, either AC and HAM/TSP patients ($r=-0.57$, $p=0.0002$, $n=30$ from UK Cohort, Figure 6A). We observed a similar tendency of decreased RORC expression with age in our Brazilian cohort ($r=-0.62$), but this observation did not reach statistical significance levels ($p=0.10$), most probably due to the small size of this ATL cohort ($n=8$) (Figure 6B). Unfortunately, the age of ATL patients was not available for the larger Japanese cohort. Next, we examined paired $CD4^+$ T-cells ($n=293$), $CD8^+$ T-cells ($n=283$), and PBMC ($n=77$) microarray results from a cohort of healthy controls with sufficient power to study the effects of age (Healthy Estonian Cohort, Table 1 in supplementary methods). We found that RORC

expression levels did not significantly change with age in CD4+ T-cells (r= 0.002, p= 0.45, Figure 6C), CD8+ T-cells (r=0.0001, p= 0.86, Figure 6D) or PBMC (r=0.0001, p=0.93), nor with gender (data not shown).

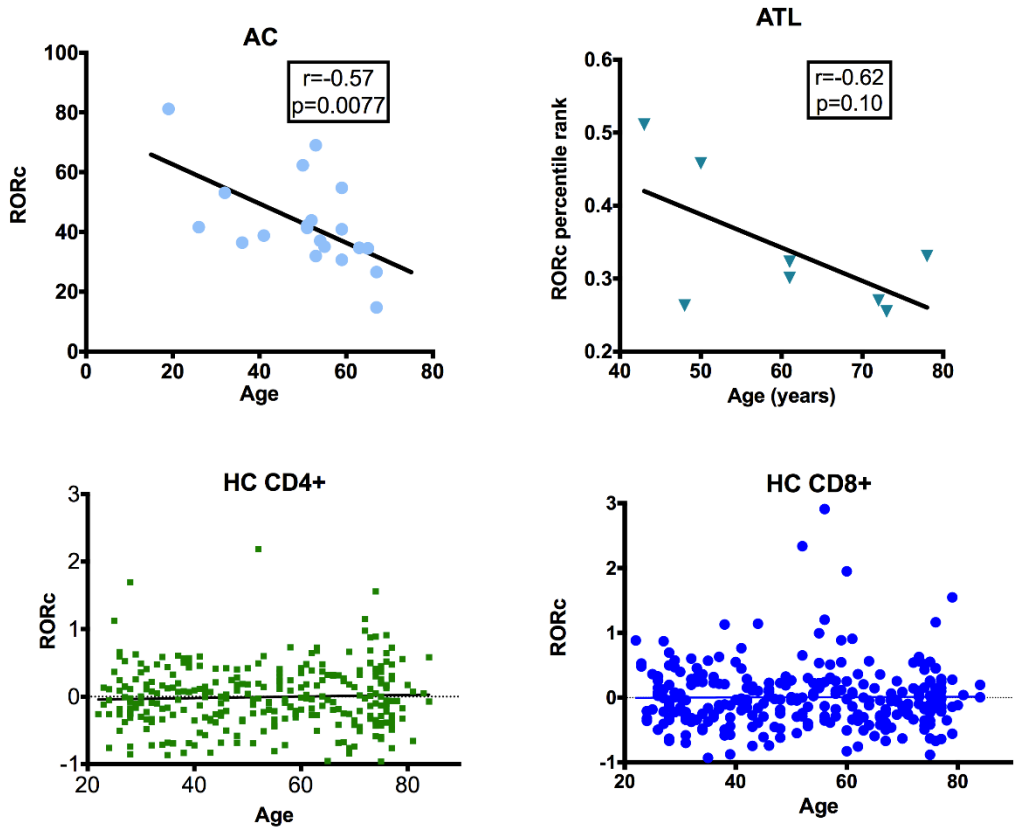


Figure 6: HTLV-1-infected Individuals from UK Cohort (A, top left) and ATL patients from Brazilian Cohort (B, top right) showed a decrease in RORC expression as age increased. This decrease was absent in healthy controls, in either CD4+ cells (C, bottom left), CD8+ cells (D, bottom right) or PBMCs (not shown).

A minor RORC^{hi} subgroup of ATL patients displays a unique CADM1^{lo}HBZ^{lo} phenotype

RORC^{hi} outliers (Rout Method [219], Q=0.1%) were observed in the three Japanese cohorts, accounting for a total of 13 out of 108 ATL patients (12.04%), Therefore, we examined this phenotype more closely in the largest examined cohort (Japanese Cohort #2), where 7 outliers with a higher normalized RORC expression were identified (Figure 5B). The patients from this cohort were then split into two groups, according to their RORC levels, as shown in Figure 7. Interestingly, we noted that RORC expression was inversely associated with expression levels of pathognomonic, or unique disease identifying, ATL biomarkers *CADM1* and HBZ. Thus, RORC^{hi} patients displayed significantly lower HBZ (p=0.0061) and *CADM1*

(p=0.045) levels, but similar expression levels of other ATL surface marker genes (*CD4*, *CD25/IL2RA*, *CCR4*) suggesting the RORC^{hi} subgroup might represent a distinct, possibly clinically relevant, molecular subgroup of ATL. The lower *CADM1* and *HBZ* expression levels in RORC^{hi} patients may represent the decreased proliferation rate of chronic or less aggressive ATL subtypes. As shown in Supplementary Figure S2, RORC^{hi} patients showed similar expression levels of other ATL driver genes (*STAT3*, *PLCG1*, *NFκB1*, *RELA*, *FAS*) [187], [190], highlighting the specificity of the RORC^{hi}CADM1^{lo}HBZ^{lo} phenotype. Positive expression of *IRF4* and *c-REL* has been associated with resistance to IFN-α+AZT therapy in ATL patients [190], [220]. Interestingly, IRF4 and c-REL expression did not differ between RORC^{hi} and RORC^{lo} patients (Supplementary Figure S2). This finding suggests RORC expression is independent of IFN-α+AZT therapeutic resistance and offers an additional molecular target for patients failing this therapy.

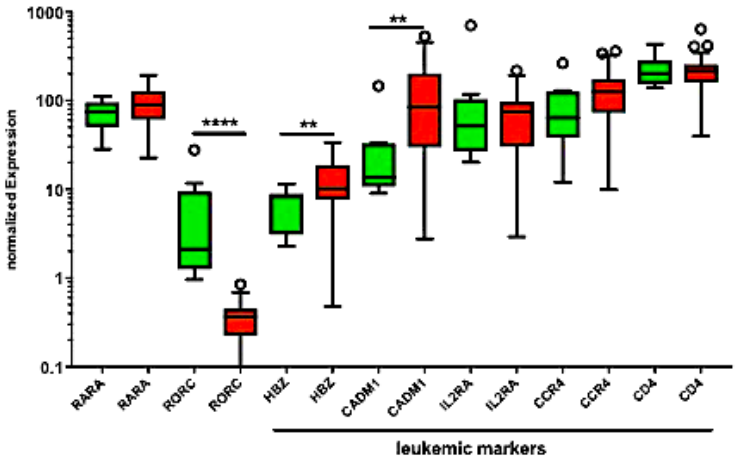


Figure 7: RORC expression levels of ATL patients from Japanese Cohort #2 separated into two groups: RORC^{hi} (Green) and RORC^{lo} (Red) show RORC^{hi} levels were associated with lower *HBZ* and *CADM1* expression levels. RORC^{hi} (7 outliers) and RORC^{lo} groups were compared with expression levels for ATL driver/mutated genes. *p<0.05 **p<0.01 ****p<0.0001

Definition of a consensus RORC pathway and gene set and its relevance to ATL oncogenesis

To facilitate the molecular exploration of the RORC^{hi} phenotype, a RORC gene set was determined based on published literature findings, integrating the intrinsic oncogenic pathway for *STAT3* activation, as defined by Yu et al. [221], and *RARA/RORC* signaling summarized by Muranski and Restifo (2013) [222]. The consensus RORC pathway included *IL6*, *IL23*, *IL21*, *IRF4*, *BATF*, *STAT1*, *STAT5*, *RARα*, *TGFβ*, *NFκB*, *SLC2A1* (*GLUT1*), *BCL6*, *STAT3*, *FOXP3*, *SOCS1*, *RORC*, and *IL17A/F*. Figure 8A illustrates the interplay between

these genes, as detailed in the legend. *RNX1*, *T-bet*, *RORA*, and *TGFB1R* were not measured by the microarray used for the initial WGCNA analysis on the Brazilian cohort (pilot cohort) and therefore excluded from the gene set. To validate the biological significance of this manually compiled pathway, we applied STRING protein-protein interaction enrichment analysis, which confirmed highly significant interaction for the RORC consensus pathway (expected number of edges: 11, observed number of edges: 83, enrichment $p<10^{-16}$). As displayed in Figure 8B, our compiled RORC pathway was significantly enriched for "Positive regulation of cytokine production" (p=7.9x10-12), "Regulation of T-helper cell differentiation" ($p=2.2x10^{-10}$), "Th17 immune response" ($p=1.9x10^{-9}$), "Jak-STAT signaling pathway" ($p=5.1x10^{-9}$), "Pathways in cancer" ($p=1.2x10^{-8}$), "HTLV-1 infection" ($4.4x10^{-5}$) and "Viral carcinogenesis" (p=0.0047), thus validating our approach.

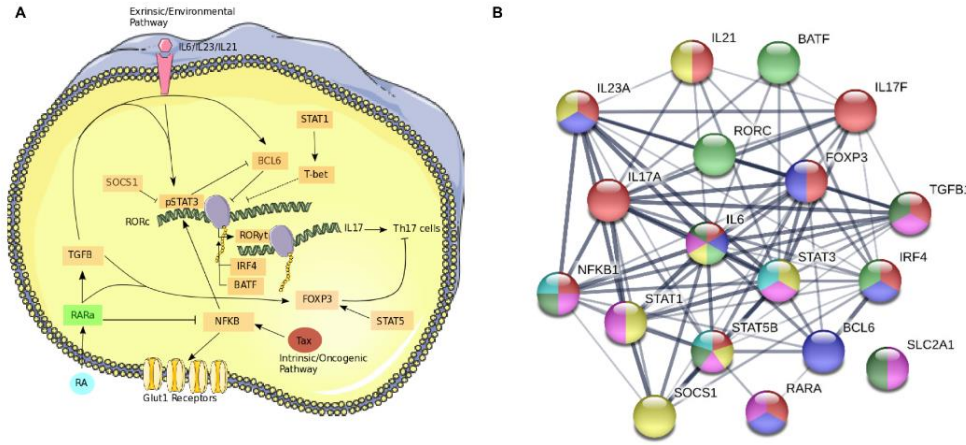


Figure 8: A) Simplified figure depicting the roles of RORC consensus pathway members, adapted from Yu et al. (2009) [221] and Muranski and Restifo (2013) [222]. The figure was produced using Servier Medical Art (http://www.servier.com) and edited using Inkscape software. B) The RORC consensus pathway was validated using STRING protein-protein interaction, GO biological process and KEGG pathway enrichment analysis. Significant enrichment (genome-wide FDR<0.05) is shown for "Positive regulation of cytokine production" (red), "Regulation of T-helper cell differentiation" (purple), "Th17 immune response" (green), "Jak-STAT signaling pathway" (yellow), "Pathways in cancer" (magenta), "HTLV-1 infection" (dark green) and "Viral carcinogenesis" (turquoise).

A modular approach reveals a link between the RORC consensus pathway, proliferation and leukemogenesis

Transcriptomic expression levels of RORC pathway members extracted from a UK HTLV-1-infected asymptomatic control dataset (UK Cohort; GSE29312) and ATL cohort (Japanese Cohort #2; EGAD1001411) showed that the majority were expressed at highly

variable levels (Supplementary Figure S3). First, prominent STAT1 expression is in line with published findings in AC [217], [222] and ATL [190], [223], [224]. On the other hand, downstream members of the RORC pathway and particularly, IL17 family genes were either undetectable or poorly expressed.

WGCNA analysis of PBMCs from our pilot ATL cohort (n=8, Brazil, Figure 9A) showed overlap of the RORC pathway with a gene module correlated to proliferation, containing bona fide proliferation markers PCNA (Proliferating Cell Nuclear Antigen) and *MKI67* (the gene coding for *Ki67* antigen, routinely used in flow cytometric quantification of proliferation). As shown in Figure 9B, downstream pathway members *RORC* and the *IL17* family, under-expressed in ATL, were negatively correlated with the proliferative module and positively correlated with the anti-proliferative module. Likewise, upstream and overexpressed gene members of the RORC pathway displayed the reverse trend. This resulted in a significant bifurcation in the RORC pathway, as shown by linear regression of correlation coefficients of member genes with proliferative and antiproliferative modules, respectively (Supplementary Figure S4, $r=-0.97$, $p<0.0001$). Overall, WGCNA analysis suggested that inducing RORC and its downstream signaling, as well as blocking upstream pathway members may decrease the cell proliferation rate in ATL.

To confirm and extend these findings on proliferation, we repeated the WGCNA in the larger cohort of ATL patients (n=44, Japanese cohort #2). We additionally obtained in silico estimates of the relative size of 22 immune cell type populations using the CIBERSORT software [225]. As shown in Figure 9B, RORC was the only pathway member which was significantly and positively correlated ($r=0.42$) with the presence of resting memory $CD4^+$ T-cells ($p=0.0041$). Downstream pathway members *IL17B* ($r=0.62$, $p= 0.0000054$) and *IL17C* ($r=0.42$, $p=0.04$) were positively correlated with the presence of naïve $CD4^+$ T-cells. STAT3 inducer NFκB subunits 1 and 2 were negatively correlated with naïve $CD4^+$ T-cells ($p=0.02$ and $p=0.052$ respectively) and resting memory $CD4^+$ T-cells (p-0.000073 and $p=0.00084$ respectively). Similar to the observations in the WGCNA of the pilot cohort, a reverse trend was also seen in the CIBERSORT analysis between upstream and downstream members of the RORC pathway and their correlation with naïve and activated memory $CD4^+$ T-cell fractions (Figure 9B). Together, the two WGCNA analyses, combined with CIBERSORT $CD4^+$ subtype quantification, suggest a distinct change in proliferative pathways between upstream and downstream members of the RORC/IL17 pathway with opposite effects in

activated memory vs. naïve and resting memory CD4+ T-cells. Among downstream pathway members, *IL17C* showed the strongest antiproliferative gene module membership in both cohorts and was also more frequently detected than other *IL17* family members (*IL17A/B/D/F*). Therefore, we classified ATL patients from the largest cohort (Japanese cohort #2) into *IL17C* expressing, (*IL17C*^{pos}, n=17) and *IL17C* negative (*IL17C*^{neg}, n=28). As shown in Figure 9C, *IL17C*-positive patients had significantly lower gene expression levels of proliferative marker *PCNA* (Mann-Whitney p=0.022) and in those patients, *IL17C* was positively correlated to *RORC* gene expression (r=0.54, p=0.026), confirming the findings of our modular analysis.

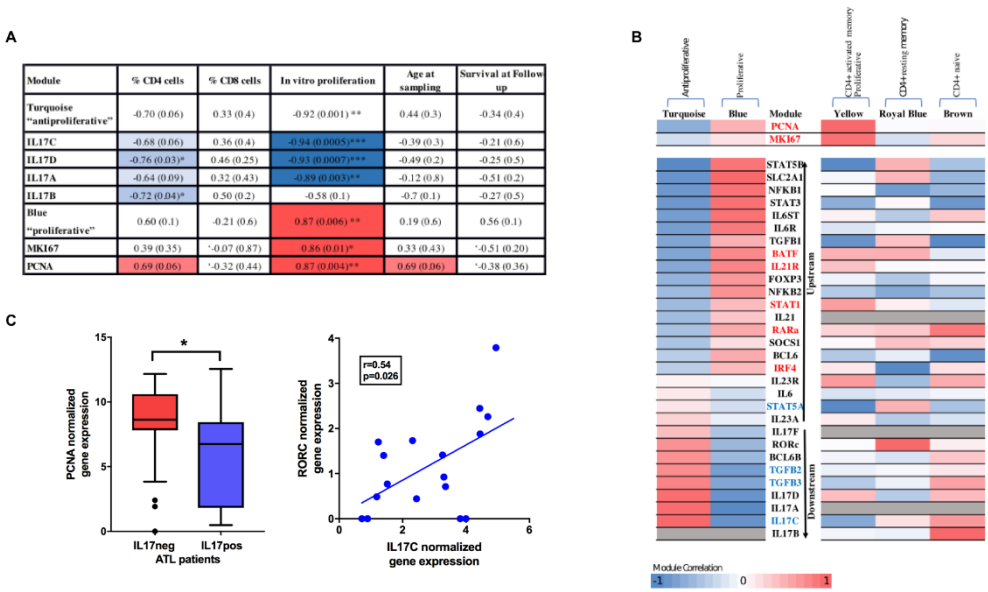


Figure 9: Modular transcriptomic analysis of primary ATL cells reveals a strong association of the RORC/IL17 pathway with proliferation. A) WGCNA findings for the Turquoise and Blue modules are shown for selected molecular and clinical correlations. R-value and p-values (in parenthesis) are shown *p<0.05 **p<0.01. B) Most RORC pathway members were positively correlated with the turquoise ('antiproliferative') gene module in primary ATL cells and negatively correlated with the blue ('proliferative') modules, such that most downstream members were found to be associated with the antiproliferative module (Brazilian cohort, n=8, left panel). WGCNA of Japanese Cohort #2 (right panel, n=44) shows *RORC* has a positive correlation with CD4+ resting memory T-cells (Royal Blue module, r=0.47, p=0.001). Yellow module is positively correlated with CD4+ memory active T-cells (r=0.39, p=0.007), as well proliferation markers *PCNA* (r=0.92, p=10-15) and *MKI67* (r=0.87, p=10-12). Brown module is positively correlated with CD4+ naïve T-cells (r=0.49, p=0.0006). Genes which were validated in both ATL cohorts for proliferative modules are colored according to their R-values. C) ATL patients expressing *IL17C* (IL17^{pos}, n=17, Japanese Cohort #2) showed decreased PCNA expression as compared to patients with undetectable *IL17C* (IL17^{neg}, n=28), and *IL17C* levels were positively correlated with *RORC* levels.

Validation of IL17C as a potential 'antileukemic' target in multiple ATL, T-ALL and B-ALL cohorts

First, we analyzed *IL17C* expression in an independent UK cohort for which clonality analysis as well as clinical data (including therapeutic response) were available. RNAseq analysis of purified 'ATL-like' cells with a $CD4^{+}CCR4^{+}CD26^{-}CD7^{-}$ immunophenotype demonstrated that *IL17C* transcripts were detectable in all ATL patients, but at significantly lower levels, as compared to AC. As shown in Figure 10A, *IL17C* transcripts were significantly decreased in both indolent and aggressive ATL (One-way ANOVA, Bonferroni's post-test $p<0.05$). No difference in *IL17C* levels was observed between ATL clinical forms or with regard to therapeutic response (chemotherapy and IFN+AZT resistance, not shown). However, *IL17C* expression was negatively correlated to clonality ($r=-0.72$, $p=0.0086$, $n=12$) in AC (fraction of largest clone 0.02-0.34) and patients with ATL (fraction of largest clone 0.68-0.99) (Figure 10B), in support of our hypothesized antiproliferative/antileukemic role for *IL17C*. Of note, *IL17A* and *IL17F* transcripts were not expressed (data not shown) in 'ATL-like' cells, in agreement with Kagdi et al. (2018) [226], who demonstrated compartmentalized expression of most cytokines in non-leukemic cells.

Second, to explore if the antiproliferative *IL17C*/*PCNA* link might be specific to ATL or shared with other leukemias, we analyzed two large cohorts of acute T- and B-cell leukemia (T-ALL, n=138; B-ALL, n=300). Similar to ATL, we found a significant negative correlation between *IL17C* and *PCNA* expression levels in both T-ALL ($r=-0.24$, $p=0.007$) and B-ALL ($r=-0.28$, $p<0.0001$), as shown in Figure 10C and Figure 10D. Unfortunately, no clinical follow-up data (survival or therapeutic response) are available for the T-ALL and B-ALL cohorts.

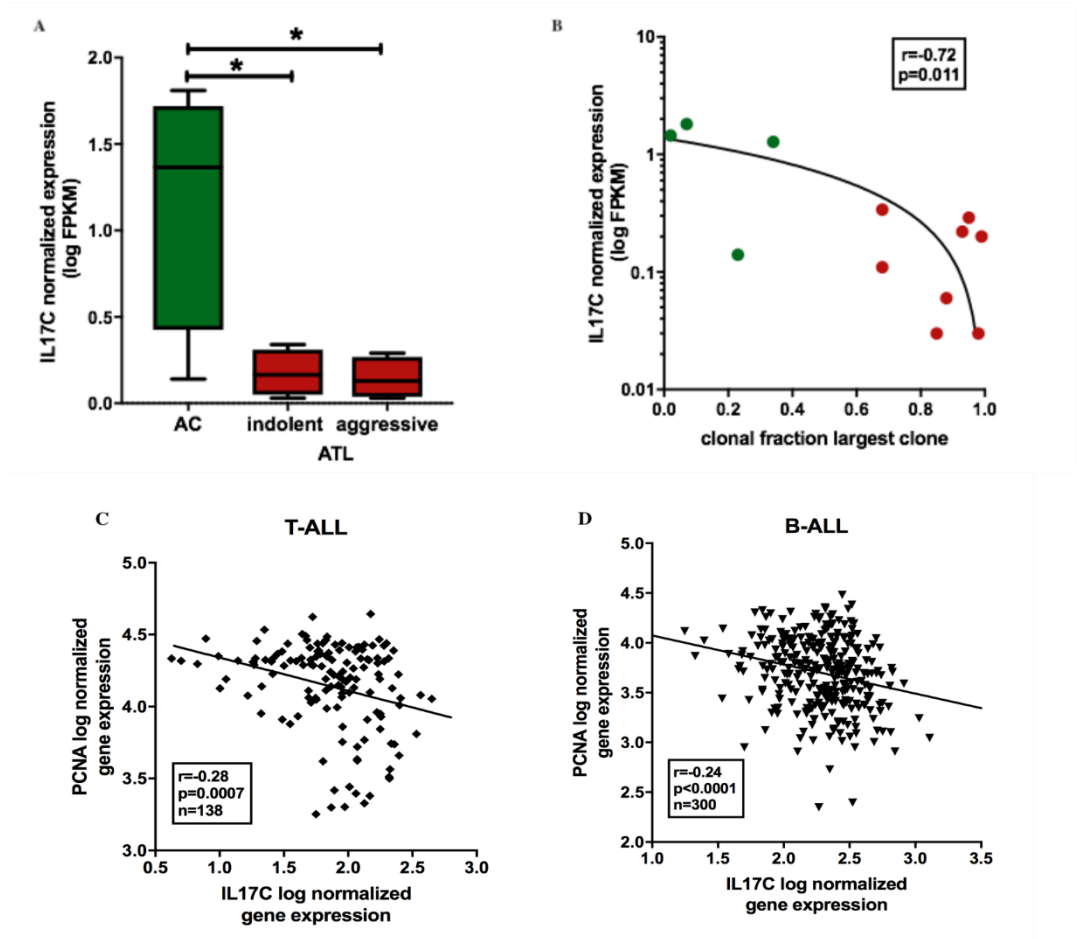


Figure 10: Validation of IL17C as a potential 'antileukemic' target in independent ATL, T-ALL and B-ALL cohorts. A) In an independent UK cohort of HTLV-1-infected individuals, *IL17C* transcripts were significantly decreased in purified $CD4^+CCR4^+CD26^-CD7^-$ cells from both indolent and aggressive ATL patients, as compared to AC (One-way ANOVA, Bonferroni's post-test $p<0.05$), with no difference in *IL17C* levels between ATL clinical forms. B) IL17C expression was negatively correlated to clonality in AC and ATL patients (fraction of largest clone 0.02-0.34 and 0.68-0.99, respectively). A negative correlation between *IL17C* and *PCNA* transcript levels was replicated in C) T-cell Acute Lymphoid Leukemia (ALL) (n=138) and B-cell ALL cohorts (n=300).

IFN-α, IFN-β and Ascorbic Acid in vitro treatment differentially regulates RORC pathway members in primary ATL cells and HTLV-1 transformed cell lines.

We previously tested the effects of IFN-α and Ascorbic Acid (AA) on HTLV-1-infected transformed cell lines (MT2, MT4, C8166) [192], [227], [228]. Although both drugs have shown moderate success in decreasing HTLV-1-induced proliferation [178], [192], [227], [228], only the high-dose AA affected the retinoic acid pathway, specifically the shared RORC/Th17 pathway. Reanalysis of our transcriptomic data showed that neither IFN-α nor high-dose AA altered *RORC* expression levels (log fold-change=0.042, p=0.59 and log fold-

change=0.069, p=0.39, respectively). AA stimulated an increase in expression of a key gene in Th17 differentiation, *IL23R* (log fold-change=0.81, p=0.000024), in support of its possible use in (combination) therapy for ATL. *RARα* expression was unchanged by IFN-α (log fold change = 0.003, p=0.98) but decreased by AA (log fold change=-0.28, p=0.064). Interestingly, our *in vitro* data (Brazilian cohort) demonstrated that *RARα* levels are upregulated upon *in vitro* treatment of ATL PBMCs with IFN-β (log fold-change=0.31, p=0.017), but not IFN-α. IFN-β also significantly modulated the expression levels of *STAT1*, *IRF4*, *TGFB1R*, *IL23R*, *FOXP3*, and *IL6*, while IFN-α significantly altered *BCL6* only, as shown in Supplementary Figure S5. This is in accord with our recently demonstrated differential anti-proliferative and pro-apoptotic effect of both IFN subtypes [229].

Discussion

Upon transcriptomic meta-analysis of four different cohorts, we found a specific and consistent RORC^{lo} phenotype in primary ATL cells and to a lesser extent in HTLV-1-infected individuals, in contrast to healthy controls. In addition, HTLV-1-infected individuals displayed an age-dependent decrease in RORC expression. The observed two-step decrease of RORC in ACs and ATL patients might thus represent an early event in HTLV-1-driven leukemogenesis. We also identified a small subset (12.0%) of RORC^{hi} ATL patients with significantly lower pathognomonic *CADM1* and *HBZ* levels but similar levels of other ATL markers (CD4, CD25 and CCR4), hinting at a less aggressive ATL subtype.

ATL pathogenesis develops over decades as is seen by patients presenting at least 20 years after HTLV-1 infection; yet not all infected patients develop ATL. Observational studies suggest that ATL, at least in the Caribbean and Brazil, can be triggered by the pediatric cutaneous manifestation known as Infectious Dermatitis [230]–[233]. ID is a chronic, eczematous condition with scaly, crusted lesions often superimposed by Staphylococcus aureus or Streptococcus pyogenes infections [230], [231]. Defects in the Th17 axis increase vulnerability to *S. aureus* and *C. albicans* infections, whereas *in vivo S. aureus* primed memory Th17 cells inhibited *IL-17* production and increased *IL-10* production [232], [234]. Of interest, two recent papers have demonstrated a role for IL-10 as an unexpected proliferative trigger of infected CD4+ T-cell clones and, possibly, leukemogenesis [235], [236]. Corroborating these findings, *IL-10* was found to be a significant (r=0.36, p= 0.013) member of the proliferative gene module, together with *PCNA* and *MKI67*, in our WGCNA analysis. In addition,

RNAseq analysis of purified cells with a $CD4^+CCR4^+CD26^-CD7^-$ leukemic phenotype from an independent UK cohort revealed *IL17C* is expressed in 'ATL-like' cells. In agreement with our proposed protective role, *IL17C* was significantly and negatively correlated with clonality (Figure 10B). Therefore, our findings underscore an IL-10 vs. RORC/IL-17 antagonism in HTLV-1-associated pathologies and provide a possible molecular basis for the epidemiological link between ID and ATL, alteration of the RORC/Th17 axis, and subsequent progression to leukemogenesis.

Modular transcriptomic analysis in ATL shows a strong correlation of the RORC pathway with cell proliferation and possibly oncogenesis, which supports its therapeutic potential. WGCNA analysis combined with CIBERSORT suggested the involvement of RORC pathway members in the homeostasis of resting memory and naïve $CD4^+$ T-cells. Combining the $RORC^{lo}$ observation in ATL cohorts with our WGCNA analysis, we find that decreased RORC expression is correlated with proliferation and ATL driver genes (*STAT3*, *NF-κB*). Thus, inducing RORC and switching to a $RORC^{hi}$ phenotype may convert ATL cells to a less aggressive subtype, suggested by the lower *CADM1* and *HBZ* levels seen in the $RORC^{hi}$ subset (Figure 7). However, no overlap was found between the module memberships of *RORC*, *HBZ*, and *Tax* in the WGCNA of Japanese Cohort #2 (data not shown). In addition, no RORC gene module members were significantly correlated to *HBZ* or *Tax* transcript levels, suggesting decreased RORC levels and signaling in ATL are not a direct consequence of retroviral transcription. Therefore, we hypothesized the RORC/IL17 axis might be linked to proliferation in other (lymphoid) leukemias. Indeed, the strongest negative correlation observed in both ATL cohorts, between *IL17C* and proliferation marker *PCNA*, was replicated in two large cohorts of other acute lymphoid leukemias, namely T-ALL and B-ALL (Figure 10C-D).

Thus, our data reveal a widely prevalent antagonistic regulation between Th17 cells, usually considered as pro-inflammatory, and leukemic cell proliferation. Regarding the clinical translation of these results, antitumor immunotherapy using Th17 cells has recently shown promising results in animal models. Adoptive cell therapy using *ex vivo* Th17 cell selection enhanced antitumor activity [235], [237], to a greater extent than Th1 cells and other $CD4^+$ T-cells [235], [237]. In addition, inducing *IL17* expression via RORC stimulation would also subsequently alter the host immune response to reduce the risk of opportunistic infections by increasing Th17 cell count [205]–[209].

Although most often believed to antagonize IL17 production, IFN-β can trigger and even exacerbate IL17 production, especially in Th17-mediated inflammatory diseases [236], [238]. This becomes problematic in cases of multiple sclerosis, where 30-50% of patients are resistant to IFN-β therapy [236]. However, this same exacerbation could be useful in ATL as a means of increasing Th17 cell production and decreasing proliferation of leukemic clones. IFN-β significantly alters the expression of more RORC pathway members than IFN-α (Supplementary Figure S5), a common therapeutic adjuvant to zidovudine in ATL treatment. This finding, along with the observation that IFN-β has superior anti-proliferative and pro-apoptotic properties compared to IFN-α [229], makes IFN-β a novel, valuable option for combination therapy in ATL.

Recently, immune checkpoint inhibitors have come to the forefront of anticancer immunotherapy [237], [238]. Immunotherapeutic targeting of Programmed death ligand -1 (PD-L1) can increase Th17 cell count, restoring IL-17A protein levels in naïve T-cells of patients with a loss-of-function STAT3 mutation [237]–[239]. Conversely, inducing Th17 cell differentiation by RORγ agonist LYC-54143 simultaneously reduced PD-1+ cell numbers and PD-1 expression *in vitro*, and resulted in tumor growth inhibition *in vivo* in two murine models [240]. In ATL, PDL1 gene amplifications have been associated with worse prognosis, especially in aggressive subtypes [241]. For *PDL1* transcript levels, we observed a trend for positive correlation to $CD4^+$ cells ($r=0.66$, $p=0.091$) as well as proliferation ($r=0.66$, $p=0.075$) in our Brazilian ATL cohort (Subramanian et al. unpublished), in agreement with a deleterious role for PD-L1. Again, combination immunotherapy by RORγ agonists and PD-L1 blockade might be a more effective option in ATL, similar to the superior response rate to dual therapy with PD-1 and CTLA-4 blocking antibodies in advanced melanoma, as compared to monotherapy [242].

In conclusion, we describe a predominant $RORC^{lo}$ phenotype observed in four cohorts of ATL patients and a minor $RORC^{hi}$ molecular subgroup with significantly lower mRNA levels of pathognomonic ATL biomarkers *CADM1* and *HBZ* mRNA levels. An age-dependent decline in *RORC* level indicates a possible early event in HTLV-1-driven leukemogenesis, supported by modular transcriptomic analysis of ATL patients, revealing a strong negative correlation of the RORC/IL17 pathway with proliferation, which was shared with T-ALL and B-ALL patients. Thus, inducing *RORC* levels and/or downstream signaling might represent

(immuno)therapeutic benefit in ATL and possibly other acute lymphoid leukemias, which awaits further testing in clinical settings.

Supplementary Material

Methods

In silico analysis

RORC expression levels were examined in publicly available transcriptomic data sets from patients with ATL, HTLV-1 infected asymptomatic controls, and healthy controls. A total of 135 untreated ATL patients, 12 HAM patients, 40 asymptomatic controls (AC), and 242 healthy controls (HC) from the Gene Expression Omnibus datasets GSE55851, GSE33615, GSE19080, GSE85487, and the European Genome-phenome archive EGAD1001411 dataset were used in this study (Table 1). EGAD1001411 initially contained 45 ATL patients, but one outlier with an overall strongly divergent transcriptome was removed. The effect of age on RORC expression was investigated in the Healthy Estonian Cohort for healthy controls (n=293) and the UK Cohort for HTLV-1 infected individuals (n=30).

The Japanese Cohort #2 (EGAD1001411) RNA-Seq data was quality- and adapter-cleaned using trimmomatic [163] and cutadapt [164] and quantified with kallisto [166] using an index built on the transcriptome obtained from the Genome Reference Consortium GRCh38, rel79. CIBERSORT was used to generate an *in silico* approximation of the relative composition of 22 immune cell types in the samples [225].

To facilitate consistent analysis of both the microarray and RNA-Seq data, the ensemble and/or Agilent IDs of the datasets were matched with corresponding Entrez IDs using the biomaRt package [243], [244] in R. The Entrez IDs were verified with the associated GPL files on GEO where available. Considering transcriptomic analysis of the Caribbean Cohort was performed on a limited (non-genome-wide) microarray platform, 2134 Entrez IDs were common to all examined microarrays and comprised the list of genes examined in this study. To address the bias in the measurements inherent to each platform, we adapted the quantile discretization method proposed by Warnat et al. [245] and transformed gene expression levels into percentile ranks among the surveyed 2134 genes for the meta-analysis. To further exclude the possibility of biasing our results, we refrain from making direct statistical comparisons of gene expression levels between datasets.

Published literature on RORC and RORγ, as cited and detailed in the results section, was used to develop a consensus molecular pathway, which was validated using STRING (version 10.5) protein-protein interaction enrichment analysis (www.string-db.org) using the whole genome as background.

Weighted Gene Correlation Network Analysis (WGCNA) [167] clusters genes into modules according to their topological overlap measure which quantifies how many gene-correlates were common to both members of each gene pair. To determine coherent gene modules and their correlation to clinical and molecular data, we performed WGCNA on each of the transcriptomic datasets from two independent ATL cohorts recently published by our group: in vitro gene expression data from short-term cultured ATL patient PBMCs (n=8, Brazilian Cohort) performed in parallel with lymphoproliferation, and *ex vivo* expression data from ATL patient PBMCs (n=44) of Japanese Cohort #225. Module membership of the RORC gene set and the ATL signature genes were determined and correlated to demographic, clinical, and in vitro data.

In vitro analysis

Spontaneous lymphoproliferation of primary cells (PBMC) from ATL patients (n=8, Brazilian Cohort) was measured by [3H]-thymidine incorporation, as described previously [229].

Statistical analysis

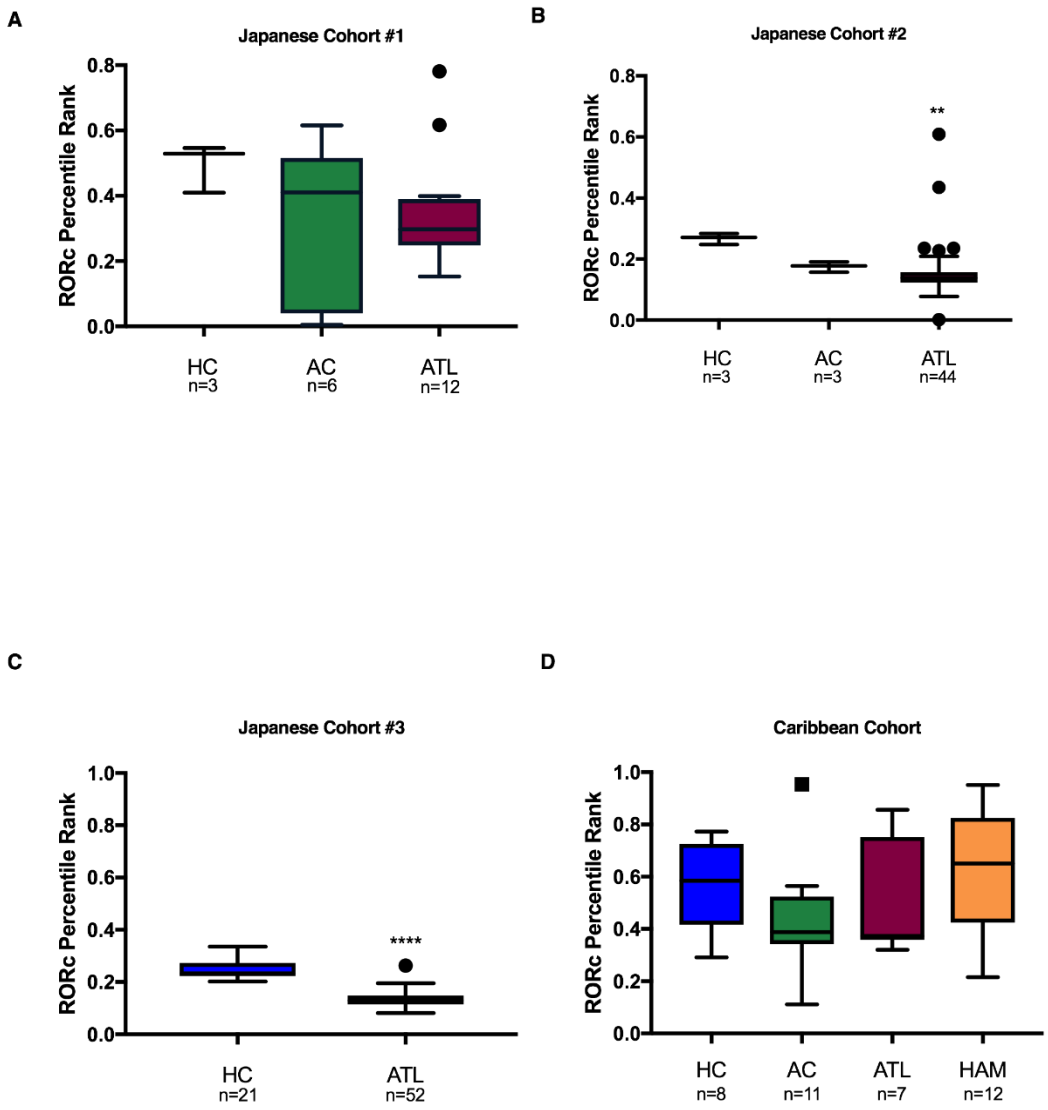
Statistical analysis was performed using GraphPad Prism 7.0. Differences in RORC gene expression were analyzed by Kruskall-Wallis test for Japanese Cohorts #1 and #2, and the Caribbean Cohort. For Japanese Cohort #3, where ACs were not included, Mann-Whitney was used to compare HC and ATL patients. The false discovery rate two-stage method of Benjamini, Krieger, and Yekutieli was used to correct for multiple comparisons. Spearman's Rho was used to correlate gene expression (either per gene or per WGCNA module using their eigengene expression) to demographic (age), clinical data (patient survival) and in vitro data (proliferation and apoptosis).

Table 1. Transcriptomic (microarray and RNAseq) datasets used in RORC expression analyses

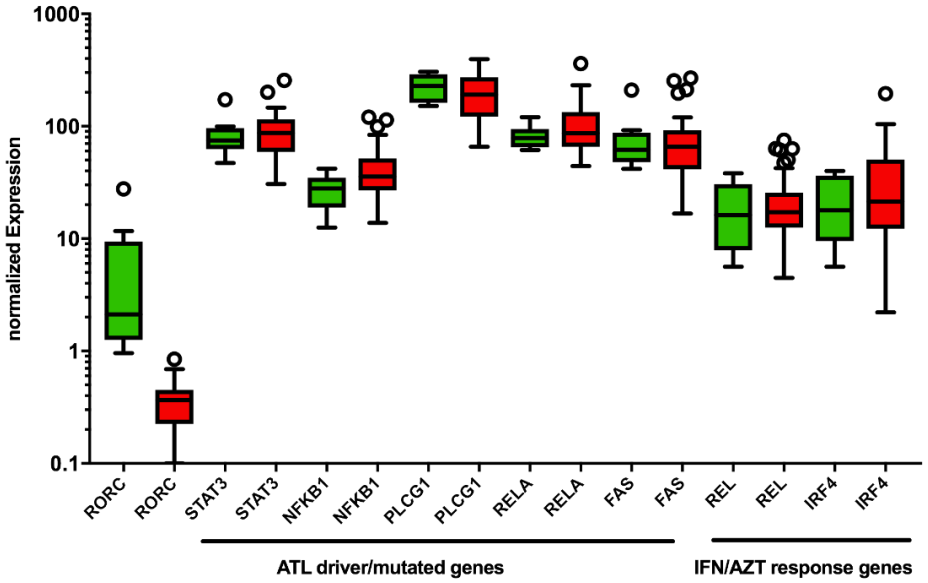
Data set	Source	Cell Type	Disease Status	Sample Size
GSE55851 Japanese Cohort #1	Kobayashi et al. (2014)	CD4+ T-cells	HC	3
			AC	6
			ATL	12
EGAD1001411 Japanese Cohort #2	Kataoka et al. (2015)	CD4+ T-cells	HC	3
		PBMCs	AC	3
			ATL	44
GSE33615 Japanese Cohort #3	Yamagishi et al. (2012)	CD4+ T-cells	HC	21
		PBMCs	ATL	52
GSE19080 Caribbean Cohort	Oliere et al. (2010)	CD4+ T-cells (Immunoarray)	HC	8
			AC	11
			ATL	7
			HAM	12
GSE85487 Brazilian Cohort	Dierckx et al. (2017)	PBMCs*	HC	5
			ATL – Untreated	8
			ATL - IFN-α	6
			ATL - IFN-β	6
GSE29312 UK Cohort	Tattermusch et al. (2012)	Whole Blood	HC	9
			AC	20
			HAM	10
GSE78840 Healthy Estonian Cohort	Kasela et al. (2017)	CD4+ T-cells	HC	293
		CD8+ T-cells	HC	283
		PBMCs	HC	77
ImmuCo	Wang et al. (2015)	CD4+ T-cells	HC	551
		CD8+ T-cells	HC	149
		Bone marrow Mononuclear Cells	AML	814
		Acute T-cell Leukemia	T-ALL	138
		Acute B-cell Leukemia	B-ALL	300

HC = Healthy Control. AC = Asymptomatic HTLV-1 Infected Control. ATL = Adult T-cell Lymphoma/Leukemia patients. HAM = HTLV-1 Associated Myopathy. AML = Acute Myeloid Leukemia. ALL = Acute Lymphoblastic Leukemia.

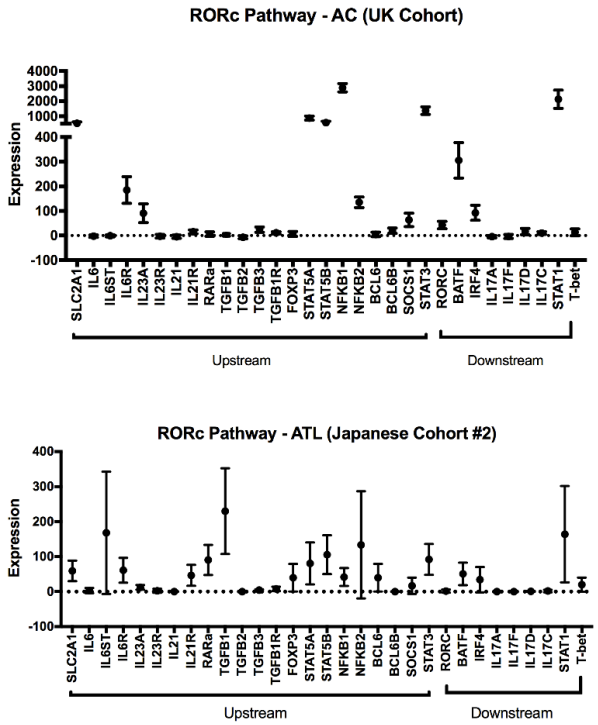
Supplementary Figures



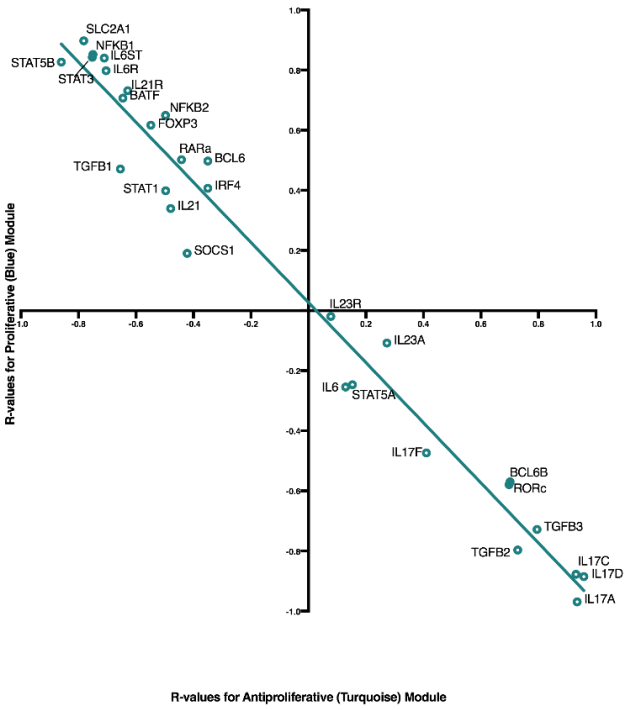
Supplementary Figure S1: RORC percentile ranks for four independent cohorts consisting of ATL patients and healthy uninfected (HC) and/or HTLV-1 infected healthy controls (AC) showed a consistent decrease in expression in ATL (A-D). HCs = Healthy Controls, ACs = Asymptomatic Controls, ATL = Adult T-cell Lymphoma/Leukemia Patients, HAM = HTLV-1-Associated Myelopathy patients. *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$.



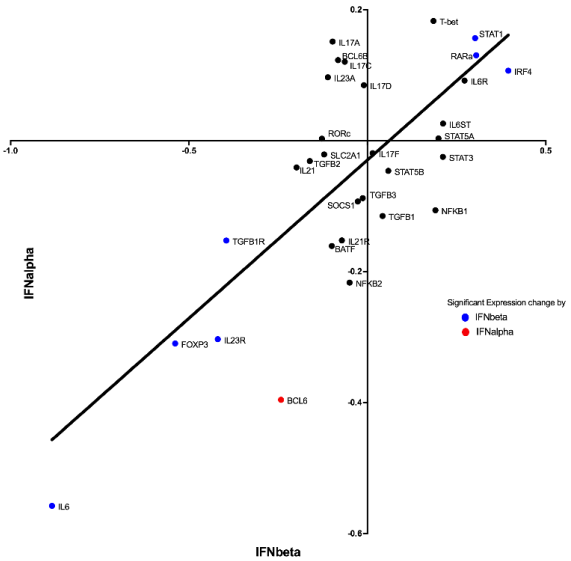
Supplementary Figure S2: Differential RORC expression of ATL patients from Japanese Cohort #2 is independent of ATL driver genes and genes involved in IFN/AZT response.



Supplementary Figure S3: RORC and downstream IL17 family genes are poorly expressed in both asymptomatic HTLV-1 infected individuals (top) and ATL patients (bottom).



Supplementary Figure S4: RORC and downstream IL17 family genes cluster in an antiproliferative gene module. Using WGCNA analysis, a significant bifurcation in the RORC pathway is observed when correlation coefficients of member genes from the antiproliferative module (Turquoise) were regressed with the correlation coefficients from the proliferative (Blue) module ($r=-0.97$, $p<0.0001$) in primary cells from ATL patients (Brazilian cohort).



Supplementary Figure S5: Differential effects of IFN-β vs. IFN-α upon RORC pathway members in primary ATL cells. Blue represents genes which had significantly ($p<0.05$) altered expression levels with IFN-β treatment. BCL6 (Red) was the only gene member which was significantly altered by IFN-α treatment.

CHAPTER 4

Quantifying chronic inflammatory burden as Glycoprotein Acetylation from gene expression data.

Progress report

Ongoing research, manuscript in preparation.

Introduction

Whenever a new technology proves to be capable of measuring a promising biomarker, finding out in which other contexts the biomarker is relevant becomes an important consideration. Generally, answering this question requires the initiation of a new project, with sample collection and the empirical testing of the new biomarker to test its viability in each setting of interest. If the biomarker measurements could be estimated (or imputed) from other available data then this process becomes far cheaper, as costs inherent to the measurements and the sample collection can be avoided. To facilitate this imputation, sufficient measurements of associated variables must be available. Publicly available transcriptomic data represents an attractive source of data for this purpose: generation of gene expression profiles is steadily becoming more affordable, and the amount of available transcriptomic data is ever-increasing. If a predictive model for some novel biomarker could be generated from a dataset of transcriptomic experiments for which paired biomarker data are available, then this model can be used to predict biomarker measurements in all other publicly available transcriptomic datasets. Thus, the model leverages existing gene expression profile data to quickly screen potential contexts in which the new biomarker could be relevant, and to estimate an effect size that can be used in power analysis to calculate minimum sample sizes required to show the existence of the effect with statistical significance.

A host of limitations applies to this relatively straightforward concept of data re-use, which does limit its applicability. First, a sufficiently large dataset of experiments is required where paired measurements of both the novel biomarker of interest and transcriptomic experiments are available. Second, transcriptomic experiments are sensitive to bias introduced throughout the transcriptomic pipeline: from sample selection, to sample handling and the transcriptomic platform used to generate the expression profile, to the bio-informatic data processing methods used to quantify gene expression. Each of these aspects can have a substantial effect on the 'virtual' biomarker measurement and needs to be accounted for. Third, the resulting model is only applicable to transcriptomic experiments performed on similar tissues used to generate the transcriptomic training data.

Applying machine learning methods to gene expression data is not a novel concept. However, the bulk of the literature available in this context does not aim to quantify a metric, but rather aims to classify samples into several groups, e.g.: bacterial *versus* viral infection [246], [247], presence *versus* absence of specific pathogens [248], presence *versus* absence of growth hormone deficiency [249], response *versus* non-response to treatment [250], evidence of sterile inflammation *versus* inflammation of infectious origin [10]. Furthermore, most research in this context limits itself to datasets generated explicitly for their own project. This is unsurprising, as the potential sources of bias in transcriptome quantification are legion and the state of the art technology, RNA-Seq, is based on vastly different principles than the ubiquitous microarray transcriptomic experiment data, to the point that they essentially measure very different things [251]. An entire field of research exists aimed eliminating platform bias from microarray measurements to make sample comparison across different microarray platforms feasible [252]–[256], with some brave souls even attempting to make microarray and RNA-Seq data comparable [257]–[259].

The situation described above is what we set out to achieve for the Glycoprotein Acetylation (GlycA) biomarker. This novel biomarker is measured using Nuclear Magnetic Resonance (NMR) spectroscopy in blood serum or plasma. First discovered in 1987 [59], high throughput measurement techniques have only recently been developed [58], [260], [261], and while the biomarker has been robustly examined in large demographic studies where it was found to be associated to chronic inflammatory burden [58], [69], [76], [77], the number of reports about disease contexts where the biomarker has been explored stands in stark contrast to the range of morbidity and mortality associations observed in cross-sectional studies [262]. Here, we aim to contribute to the discovery of useful contexts for this biomarker by levering the existence of three datasets for which both crucial components are present. In the DIetary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) and Cardiovascular Risk in Young Finns Study (YFS) datasets, whole blood transcriptomic experiments are available in conjunction with serum GlycA measurements [263]–[265]. This data can be used to generate a predictive model for GlycA, with the ambitious aim to then apply it to all publicly available transcriptomic

datasets generated from blood in order to identify disease contexts where it is worthwhile to investigate the usefulness of the GlycA biomarker.

Methods

Ethics

Secondary use of samples for the quantification of inflammatory burden through NMR and the secondary data analysis performed in this study was approved by the ethics commission of University Hospitals Leuven (MEC UZ Leuven, #s57931). Validation samples submitted for metabolic NMR analysis originate from multiple studies conducted in accordance with the principles of the declaration of Helsinki. Patients supplied written consent before sampling procedures were performed.

Model Generation Data Overview

For model construction and testing, we rely on the 2007 and 2014 DILGOM transcriptomic datasets (D07 and D14, respectively), in addition to the YFS dataset. D07 and D14 are subsets of the FINRISK study and are cross-sectional sample collections from FINRISK participants aged 25-74 years and 32-81 years, respectively. YFS is an ongoing population-based cohort study, designed to study the emergence and progression of cardiovascular risk factors from childhood onwards. The baseline study started in 1980 and recruited participants in six age groups (ages 3, 6, 9, 12, 15 and 18), in 2011, when participants were between ages 34 and 49, transcriptomic experiments were performed. The D07 (n=518), D14 (n=331) and YFS (n=1651) gene expression profiles are generated using the Illumina Human HT12 microarray platform (with D07 generated using v3, whereas D14 and YFS were generated using D14). Distinct data processing has been used to quantify expression for each of these datasets: for the YFS study, the raw data obtained from Illumina Beadstudio underwent nonparametric background correction, followed by quantile normalization with control and expression probes, by the neqc function of the limma R package [198], prior to log2 normalization. D07 data processing has been described in [76] and differs from YFS data processing in its use of the normalise.quantiles function of the preprocessCore R package, instead of the neqc function, after log2 transformation and

in its removal of probes mapping to erythrocyte globin components (i.e. HBA1, HBA2 and HBB). D14 data processing is similar to D07 with one key difference: the exclusion of 507 probes which hybridize to multiple genomic positions more than 10000 bases apart, prior to normalization. Expression of genes for which multiple probes are measured, is calculated using the collapseRows function of the WGCNA R package with default settings [167], [266]. This results in 28142 quantified genes for the YFS dataset and 17012 and 18612 genes for the D07 and D14 datasets, respectively.

Comparability

To achieve the ambitious goal of making the predictive models applicable to microarray and RNA-Seq datasets, regardless of the platform used to generate them, the data needs additional pre-processing to ensure that at least the range of the inputs is the same across datasets. Microarray probe binding intensity is usually reported as a unitless variable in a range between 3-12, representing gene expression abundance in terms of the observed intensity in comparison to reference probes, whereas RNA-Seq data is commonly presented as RPKM or FPKM, that is, Reads Per Kilobase of Transcript per Million mapped reads, though recently Transcripts Per Million (TPM) has become the preferred unit. While the body of literature about microarray comparability is extensive, approaches aimed at making microarray and RNA-Seq data comparable are much less common. In addition, many published methods apply a normalization or standardization across genes in the combined dataset. This would mean that a new normalization is required across the full range of available samples every time new samples are added, or an additional dataset is considered. Furthermore, we do not always have access to the raw data of published datasets, which is a requirement for many of the more sophisticated published solutions to the comparability problem. Here, we implement a method which accommodates both these issues: we adapt the Quantile Discretization (QD) method originally proposed in [245] so that each gene within a sample is discretized into a predetermined number of bins (here, b = 10), with a value of 10 and 1 indicating the gene ranks in the top 10% and the bottom 10% of expressed genes within the sample, respectively.

Dimensionality Reduction

While the availability of 2500 paired GlycA measurements and gene expression profiles represents the largest dataset to date, the number of available variables (i.e. gene expression levels) for each experiment is considerably larger than the available number of samples, with up to 28142 genes quantified in the YFS dataset, meaning our datasets operate under the curse of dimensionality [267]. Therefore, prior to model generation a dimensionality reduction step is required, to ensure meaningful results from the machine learning algorithms. Two types of strategies can be employed to reduce the dimensionality of a dataset: feature extraction and feature selection [268]. The key difference between these two strategies lies in how the original data is preserved. Feature extraction reduces the dimensionality by creating new variables as combinations of the existing variables, in a linear, e.g. the well-known principal component analysis (PCA), or non-linear fashion. On the other hand, feature selection identifies a subset of the original data which is deemed to carry the most relevant or least redundant information. In this work, we opt to use a feature selection method for two reasons: first, the predictive models created on these features are more easily interpreted, as each single variable represents the expression of a gene instead of some combination of genes. Second, predictive models should be applicable to publicly available transcriptomic data which frequently lack quantification data on specific genes. If one of the features used in the predictive models is absent in the data, a different model needs to be constructed which accommodates the absence of this variable. When extracted features are used, not only is the likelihood that this happens much greater than would be the case when using selected features, but the feature extraction would need to be performed again prior to each new predictive model generation.

Feature selection can be performed in a variety of ways, here we make use of the Random Forest based method Boruta [269]. Briefly, the Boruta algorithm is a Random Forest based feature selection algorithm that determines the importance of variables by quantifying the loss of predictive accuracy of Random Forest models when they are denied access to the variable. To enable an absolute decision about a variable's relevance, the first step in the algorithm is the duplication of each variable in the

dataset, followed by the random permutation of these duplicated entries. These so-called 'shadow attributes' introduce known random variables in the model, which can be used as a cut-off in the subsequent determination of each variable's relevance to the response variable, illustrated in Figure 11. The importance of each variable in each of the RF runs is compared in a two-sided test for equality to the highest importance seen for any shadow variable, thereby enabling the calculation of a test statistic and corresponding p-value for true association to the response variable.

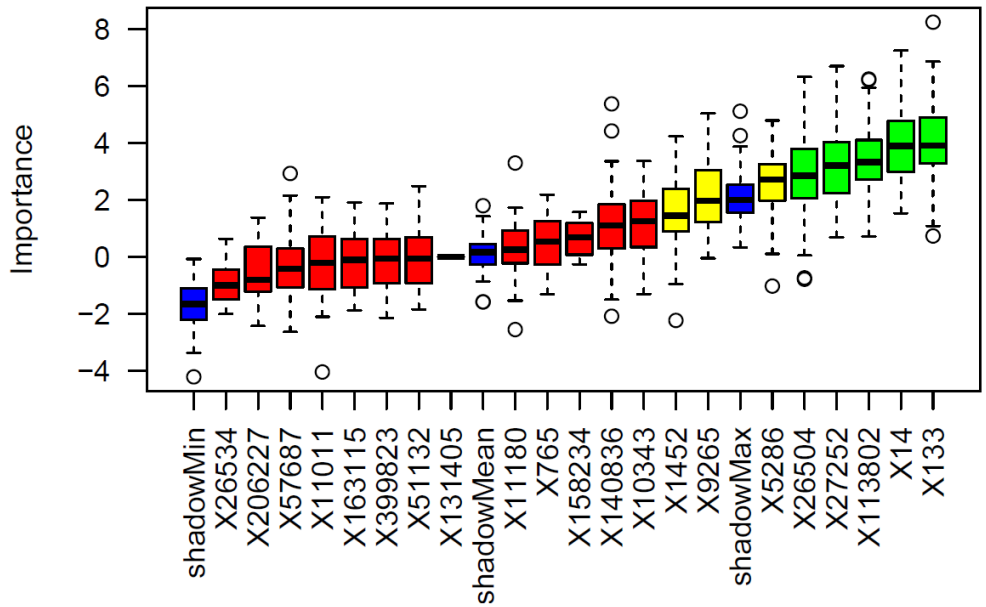


Figure 11: Illustrative Boruta Feature Selection example. A Boruta run was performed on 20 selected genes. Importance, calculated as loss of accuracy upon omission of the variable in a RF model, is depicted on the y-axis for each gene (identified by their entrezgene ID). Maximum, mean and minimum importance of known random shadow variables are depicted in blue. Color of the boxplot indicates outcome of the Boruta algorithm, where red indicate rejected variables, yellow tentatively relevant variables and green confirmed relevant variables.

The stability of feature selection algorithms is frequently overlooked [270]. Here, we report the Jaccard index of the selection over 200 runs, and generate two gene-sets: genes selected as related to GlycA in each of the 200 runs make up the core gene-set, where the related gene-set consists of all genes selected in at least 1 Boruta run.

This Boruta selection makes up the initial screening and is an example of a so-called 'wrapper' technique which uses a learning technique to evaluate which features are useful. Subsequent model generation techniques can incorporate additional feature

selection techniques like 'filter' and 'embedded' algorithms, to come to an optimal best-smallest subset of features to make their predictions. Filter techniques determine the relevance of variables without the use of learning techniques by e.g. determining which features best correlate with the response variable, whereas in embedded algorithms the feature selection is inherent to the model construction as is the case in Least Absolute Shrinkage and Selection Operator (LASSO) regression.

Response variable

GlycA levels measured in the three datasets are expected to be roughly similar, as each of these datasets was generated in general population samples. However, empirical observations show the GlycA levels for the YFS to be higher than those of the D07 and D14 studies. The only inherent difference between the two datasets is the ages of the sampled subjects, which could have an impact on the GlycA measurements as a positive association between age and GlycA has previously been reported [89]. However, the median age for the YFS dataset, which has a higher GlycA, is the lowest of the three examined datasets. Furthermore, even though the median age of the subjects in the D14 datasets is higher than those in the D07 dataset their GlycA levels are highly comparable, suggesting that the observed differences in the YFS datasets are due to differences in GlycA quantification methods rather than true biological differences (Figure 12). To allow comparison of predicted and observed GlycA levels across these three datasets, GlycA is standardized across each of the three datasets and the resulting z-score of the GlycA concentration is used as the response variable in subsequent model construction. Predicted values are therefore expressed as fractions of the standard deviation (SD) observed in these population cohorts.

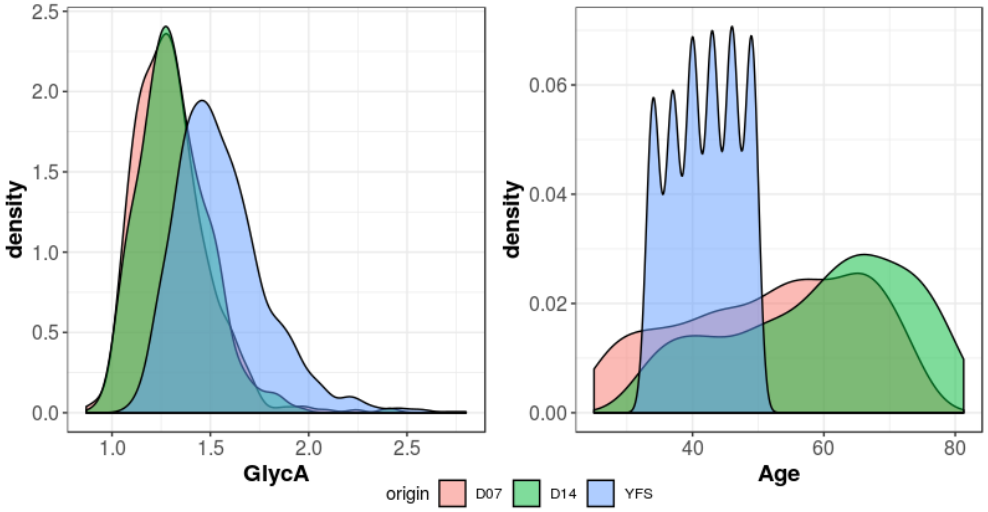


Figure 12: GlycA and Age density plots in D07, D14 and YFS datasets illustrate the differences in study design of the YFS and DILGOM studies and highlight the aberrant GlycA profile measured in the YFS study, necessitating the normalization of the GlycA response variable.

Model Generation

Predictive model construction and model performance assessment is achieved using the caret R package [271]. We compare the performance of linear models (LMs), on all available features (naïve LM) as well as constructed with recursive feature elimination (RFE) or minimum redundancy maximum relevance (mRMR, [272]) feature selection methods, ridge and LASSO regression, non-linear k-nearest neighbor methods (kNN), artificial neural networks (ANN), support vector machines (SVM) with linear or radial basis function (rbf) kernels and recursive regression trees. Detailed below are the chosen values for these models' hyperparameters and how they were determined.

Optimal number of features in RFE was chosen where the average R^2 was highest across 20 repeats of a 10-fold cross-validation, which selected 31 variables. The number of features selected in mRMR regression is chosen based on the information they contribute to the system. Visual inspection of Figure 13 identifies 4, 15 and 32 features as relevant points: the first 4 features contribute to the mutual information in the system far more than subsequent adds, at 15 features a first minimum of the mRMR score is obtained, at 32 features the mRMR score is zero. A second selection

strategy performs 20 repeats of 10-fold cross-validation and selects the optimal number of features where the resulting root mean square error (RMSE) is lowest, at 55 features.

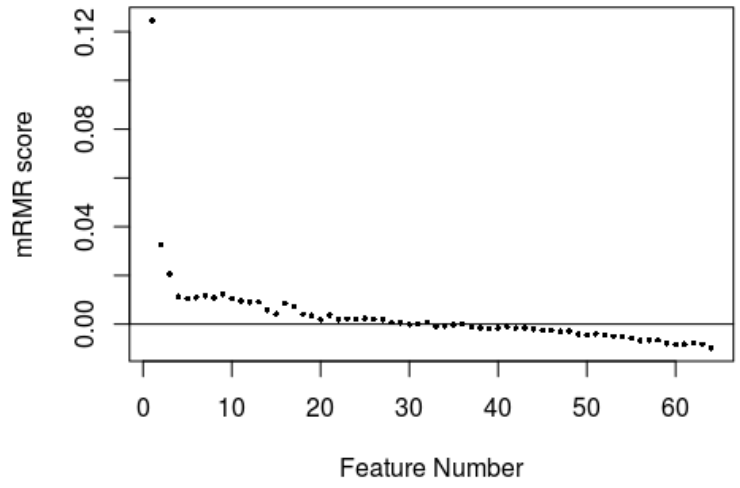


Figure 13: mRMR score of each feature in progressively more complex linear models. The mRMR score represents additional non-redundant information contributed by each additional feature included in a linear model. Salient points are identified at k=4, 15 and 32.

The selection of hyperparameters for other functions is presented in Figure 14. The elastic-net mixing parameter is set to 0 or 1 for ridge and LASSO regression, respectively, and optimal lambda is identified using a grid search between 0 and 0.5 using and selecting the lowest average RMSE in 20 repeats of 10-fold cross-validation (0.4324 for ridge regression, 0.0205 for LASSO). Optimal k for kNN is selected where average RMSE is lowest across 20 repeats of 10-fold cross-validation, k=28, though RMSE difference with k=10 is minimal, so both models are tested. The number of nodes and the weight decay of training hyperparameters for the ANN using a single hidden layer is selected through grid search and lowest RMSE is found using one hidden node and a weight decay of 0.2. Cost parameter for SVM using a linear and RBF kernel is identified through grid search and optimal RMSE is identified at C=0.0025 and 1 for linear and RBF kernel, respectively, with the optimal σ=0.0032, with $\gamma=1/(2\sigma^2)$ or γ=1003.65. The complexity parameter for the recursive regression tree (at minimum 20 observations in a node required for a split, at minimum 7

observations in a terminal node and a maximum depth of the tree capped at 30) identified through grid search and optimal cross-validated RMSE is found at cp=0.007.

Model Selection

To estimate performance on publicly available data with strong inherent biases, model construction is performed in the largest dataset (YFS, n=1651) and the most robust model is selected based on correlation between predicted and observed GlycA levels in the data of the (differently pre-processed) D07 and D14 datasets. We report model performance as cross-validated RMSE and explained variance (R^2) in the YFS training dataset. Performance on D07 and D14 test data is reported as Pearson correlation between predicted and observed GlycA levels and RMSE.

Model Validation

Once constructed, the first step in validating the model is verifying whether it can replicate known GlycA associations from third-party datasets. An overview of all datasets used for model validation can be found in Table 1, full bibliographic details for both the GlycA studies and datasets are listed in Appendix A. These datasets were gathered using Gene Expression Omnibus (GEO) query, using search terms 'PBMC' and 'Whole Blood'. Searches for datasets were performed with great scrutiny in settings where reports on GlycA measurements are available (i.e. SLE, RA, KD, bacterial infection, obesity and psoriasis).

Model Testing

The true test for the model consists using publicly available transcriptomic datasets from settings in which GlycA has not yet been examined, making predictions, and subsequently verifying the predictions using new NMR measurements of GlycA on sample collections in the appropriate settings. NMR measurements are performed by Nightingale Health (Helsinki, Finland)

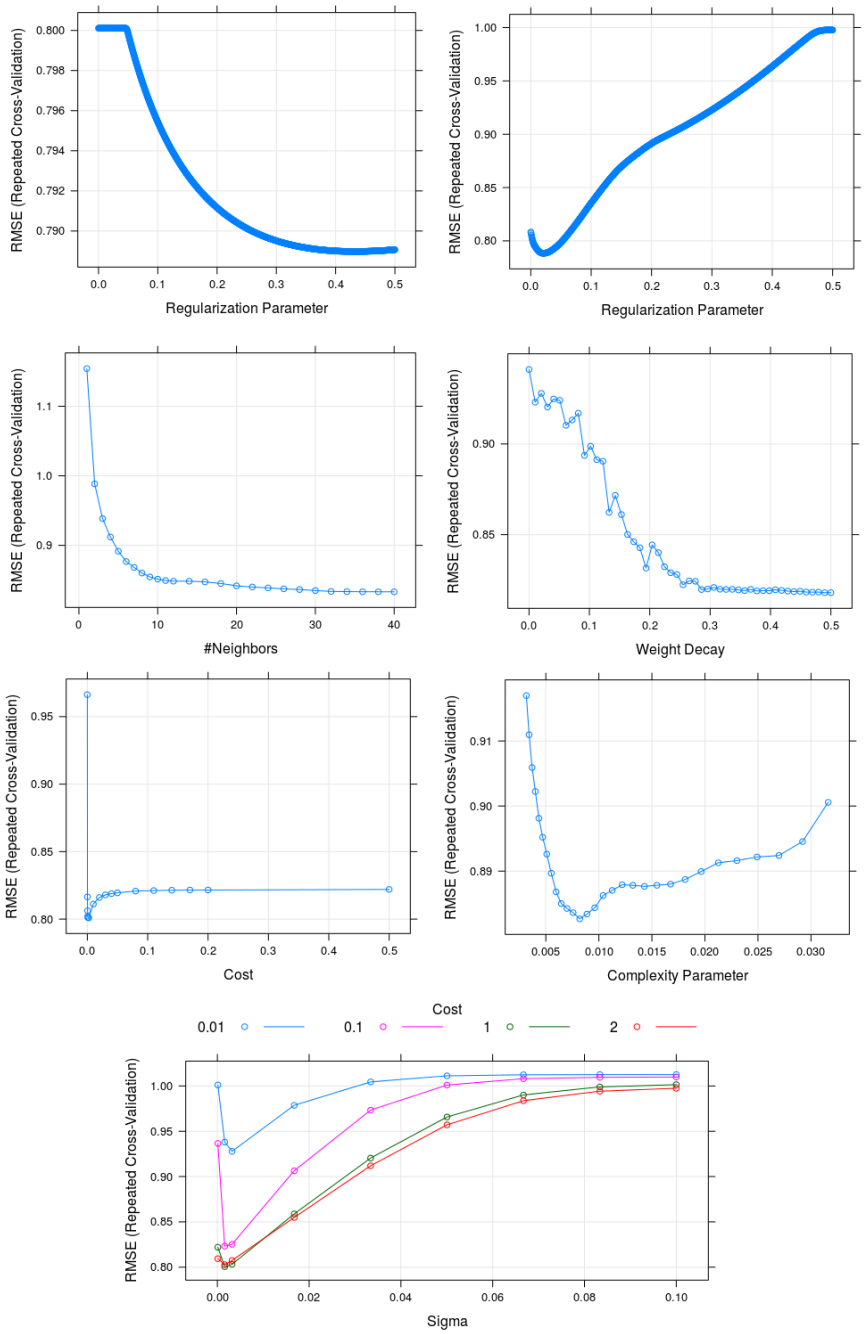


Figure 14: Hyperparameter tuning. Hyperparameters for the different modelling strategies are selected where mean RMSE is lowest across 20 runs of 10-fold cross-validation. Presented from left to right, top to bottom: the regularization parameters for ridge and LASSO regression, the number of neighbours for kNN regression, the weight decay parameter for a single hidden layer ANN, the misclassification cost for a SVM with linear kernel, the complexity (pruning) parameter for the regression tree, and finally the misclassification cost and sigma parameters for a radial basis function kernel SVM.

Disease Context	GlycA Observation	Transcriptomic Dataset	Observed Model Prediction	Replicates known observation?
Diabetes	Increase over HC	GSE46097	Diabetes patient GlycA is 0.276 SD higher than HC	Yes
CVD	Increase over HC	GSE46097	CAD patient GlycA is 0.264 SD higher than HC	Yes
Obesity	Increase over HC	GSE55205	Obese subjects predicted GlycA is 0.263 SD higher than normal weight	Yes
		GSE41505	Pearson correlation between BMI and GlycA = 0.65	Yes
SLE	Increase over HC	GSE50772	SLE patient GlycA is 0.489 SD higher than HC	Yes
		GSE17755	SLE patient GlycA is 1.038 SD higher than HC	Yes
	Correlation with SLEDAI	GSE50772	Pearson correlation between SLEDAI and GlycA = 0.47	Conflicting
		GSE72754	Pearson correlation between SLEDAI and GlycA = 0.28	Conflicting
RA	Increase over HC	GSE17755	RA patient GlycA is 0.489 SD higher than HC	Yes
		GES15573	RA patient GlycA is 0.769 SD higher than HC	Yes
	Correlated with DAS28	GSE15258	Pearson correlation between DAS28 and GlycA = 0.20	Yes
Psoriasis	Increase over HC	GSE61281	Psoriasis patient GlycA is 0.432 SD higher than HC	Yes
		GSE61281	Arthritic psoriasis patient GlycA is 0.20 SD higher than non-arthritic cutaneous psoriasis	NA
		GSE55201	Psoriasis patient GlycA is 0.265 SD higher than HC	Yes
	Decrease after antiTNF treatment	GSE55201	Anti-IL17A monoclonal antibody lowers GlycA by 0.148 SD	NA
KD	Increase over HC	GSE68004	Acute KD patients have GlycA levels 1.07 SD higher than HC	Yes
		GSE57183	KD patients have GlycA levels 1.08 SD higher than HC	Yes
	Increase over OFI	GSE68004	KD patients have GlycA levels 0.145 SD higher than OFI patients	Yes
		GSE63881	Severe Acute KD patients have GlycA levels 0.134 SD higher than OFI patients	Yes
	Acute > Convalescent	GSE63881	Acute KD patients have GlycA levels 0.968 SD higher than convalescent patients	Yes

Table 1: Model validation dataset overview. Dataset identifiers are Gene Expression Omnibus identifiers, full bibliographic details for both the GlycA studies and the publications associated with these datasets are listed in Appendix A. Most of the existing knowledge about GlycA can be replicated using the predictive model in third party data. For some observations, i.e. GlycA's correlation to SLE disease activity, conflicting reports exist in the literature. Not all reported GlycA observations can be tested: for some, appropriate transcriptomic experiments are unavailable (i.e. pre- and post-antiTNF-treatment gene expression profiles of psoriasis patients).

Results

Boruta Feature Selection

For each of 200 Boruta runs in the YFS dataset, each variable is either confirmed to be associated with GlycA or rejected. Over these 200 runs, 1288 different genes are confirmed as relevant to GlycA at least once, while 232 genes are consistently confirmed as relevant across all 200 runs. The mean Jaccard index of similarity between each run is 0.75, and selection frequencies are visualized in Figure 15.

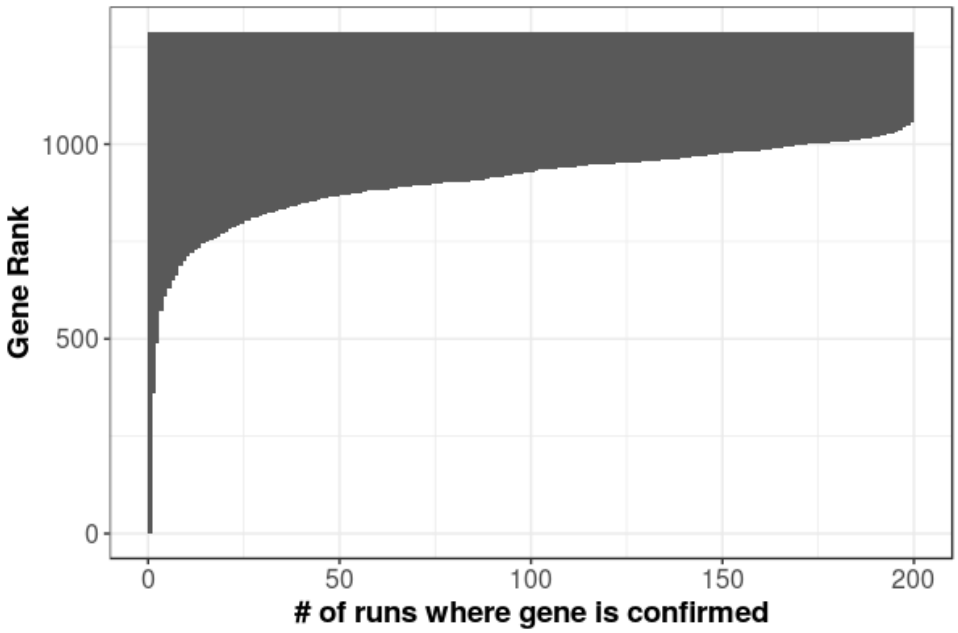


Figure 15: Boruta feature selection results stability. In total, 1288 genes are confirmed as relevant to the response variable at least once. A core set of 232 genes is consistently reported to be relevant in each of these 200 runs.

An overview of all genes in the core set can be found in Appendix B. Functional annotation of the core gene list through overrepresentation analysis using the WebGestalt platform [273] reveals four Gene Ontology (GO) Biological Process gene-sets are statistically enriched with a false discovery rate smaller than 0.05 (Table 2). In addition, specific Gene Set Enrichment Analysis (GSEA) [168] investigation of GO chronic inflammatory response, regulation of acute inflammatory response and granulocyte activation gene sets, as well as the genes of the WGCNA module related to

neutrophil gene expression reported in [76], confirms known GlycA associations (Figure 16).

Gene Set	Description	Size	Expect	Ratio	P Value	FDR
GO:0036230	granulocyte activation	498	5.9237	5.2332	2.7534E-14	2.3404E-11
GO:0009620	response to fungus	48	0.57096	10.509	2.1073E-05	0.0059708
GO:0006959	humoral immune response	237	2.8191	3.9019	0.00012675	0.019693
GO:0050900	leukocyte migration	417	4.9602	3.0241	0.00013901	0.019693

Table 2: Fisher ORA results for GO biological processes, statistically significant after multiple testing correction.

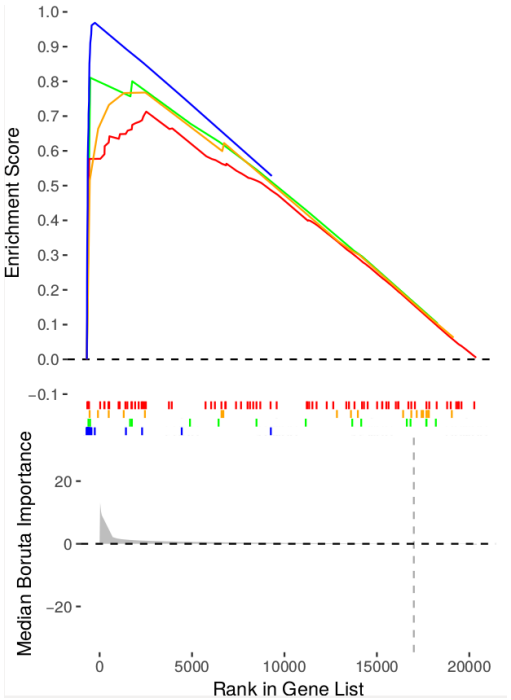


Figure 16 : GSEA results of four selected gene-sets in ranked Boruta feature selection results. Gene Ontology chronic inflammatory response, regulation of acute inflammatory response, granulocyte activation, and the genes from the neutrophil associated WGCNA module reported in [76] are colored green, red, orange and blue, respectively.

Model construction

Cross-validated performance in the YFS training data of the constructed models is presented is Figure 17. To establish a baseline for performance, linear models constructed using a random selection of 4, 15, 32 and 55 variables were constructed and tested. In the training data, the RFE linear model achieves the best accuracy in terms of RMSE and explained variance. It is worth noting that absolute errors of these models are relatively large and that a linear model with four variables boasts a performance which is almost on par with more elaborate models and more sophisticated modelling techniques.

While cross-validated performance in the training data is informative, the application of these models will use third party data with considerable inherent biases owing to the transcriptomic platform and preprocessing steps employed in their generation. Therefore, performance is better judged in the D07 and D14 datasets, each of which was quantified using a distinct processing pipeline. Performance in these test datasets is summarized in Figure 8.

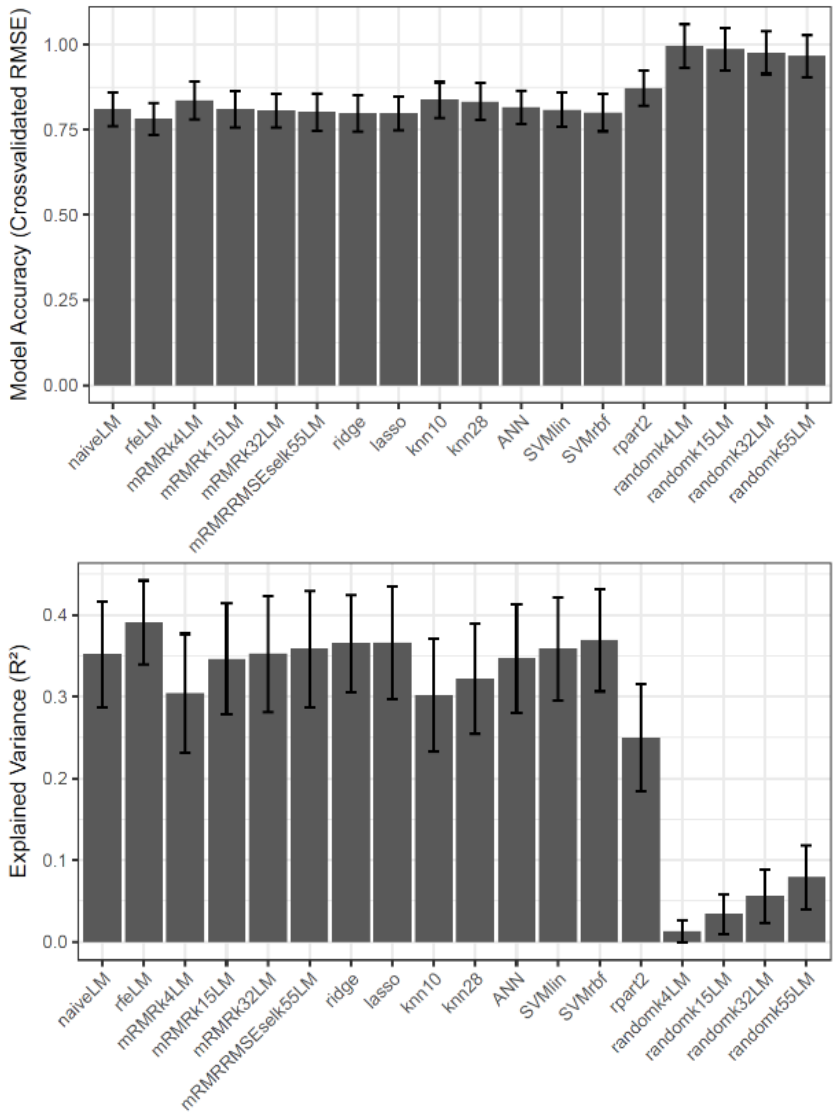


Figure 17: Performance measures of predictive models in YFS training data. On top, 10-fold cross-validated RMSE is compared across all considered models. On the bottom, their explained variance (R^2). The best-performing model in the YFS training data is the RFE linear model.

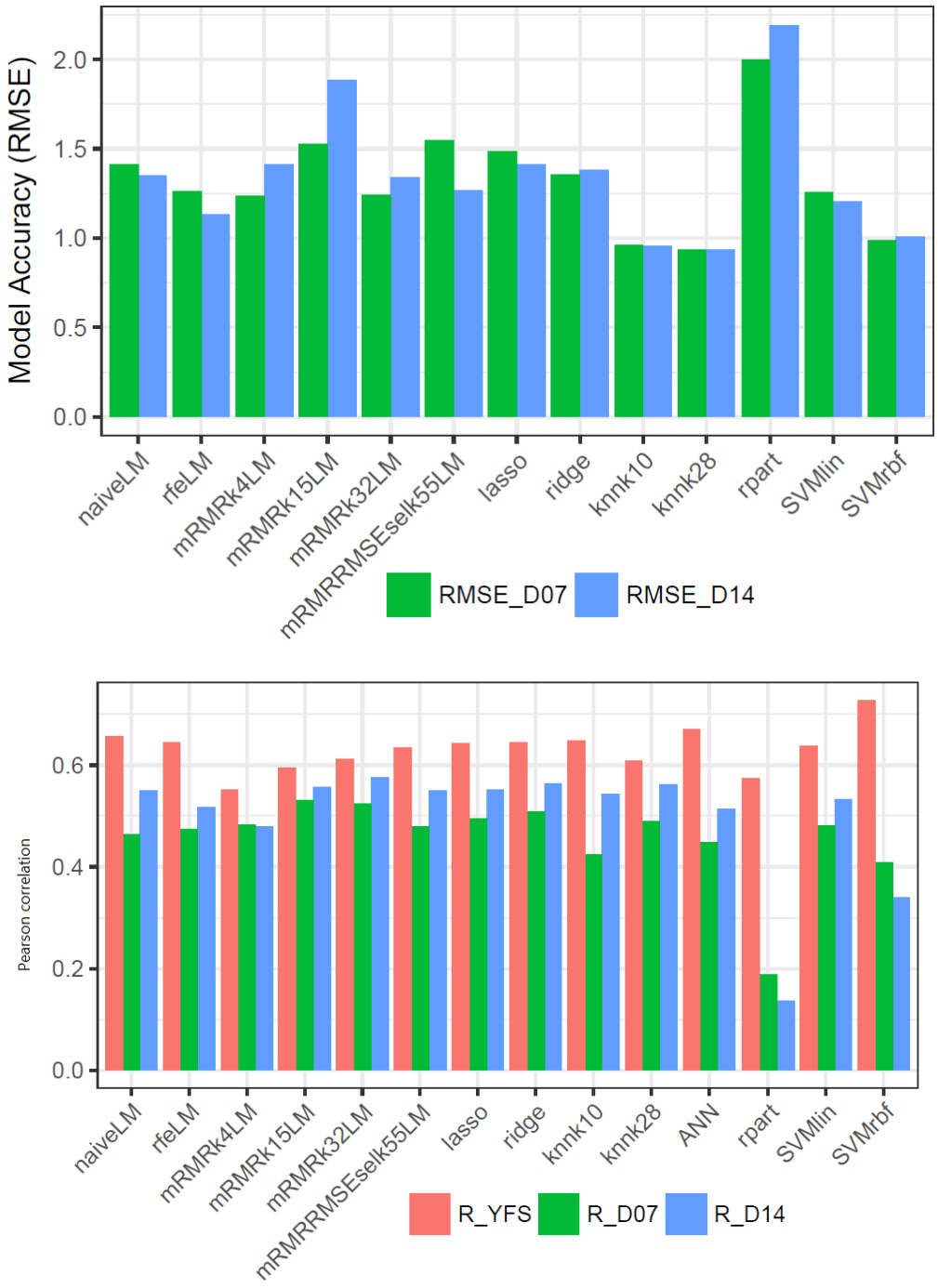


Figure 18: Model performance in unseen test data. Top: RMSE between predicted and observed normalized GlycA levels. Bottom: Pearson correlation coefficient between predicted and observed normalized GlycA levels. Bars are colored according to their dataset: training data, YFS, red, and unseen test data, D07 and D14 datasets, green and blue, respectively.

Unsurprisingly, the overall performance in the unseen test data of the D07 and D14 datasets is considerably lower than the performance in the training YFS dataset. The inherent bias on specific genes is laid bare in the RMSE performance of the 15 feature mRMR model, which is higher than both the 4 feature and 32 feature mRMR models. Inclusion of problematic genes, which are consistently quantified as considerably more or less abundant in the test data then in the training data, has a large impact on a model's performance, in terms of the absolute error. However, correlation between predicted and observed GlycA levels fares better than the RMSE. Comparison between predicted GlycA levels from different datasets using the presented models should be avoided, however within a single dataset the absolute error on the GlycA level is less important than the correlation between observed and predicted outcomes. Therefore, we select our preferred model using the correlation between predicted and observed variables observed in the unseen D07 and D14 data. The best performing models on this metric are the mRMR models using 15 and 32 features, with the latter outperforming the former by the slightest of margins. Predictions made in the following paragraphs were made using the 15 feature mRMR linear model.

Model validation

The first validation of our chosen model is its ability to replicate known GlycA associations. Demographics studies have robustly shown the association between obesity, cardiovascular disease and increased GlycA levels. In addition, there have been disease-specific studies in SLE, RA, KD and psoriasis. Table 1 summarizes known GlycA associations, which publicly available datasets were used to test the model and the outcome of the predictions. Though the effect sizes predicted by the model can differ between multiple datasets obtained from a single setting, overall the model replicates published knowledge about GlycA remarkably well. Reports on GlycA's correlation with disease activity scores in SLE are conflicting [86], [87], though the predictions made by the models are uniformly in favor of SLEDAI correlation.

Novel predictions

Having confirmed our model correctly replicates known associations for GlycA, we generate novel predictions in settings which have not yet been explored and

subsequently perform NMR measurements in appropriate sample collections to test these predictions. Predictions are made in inflammatory bowel disease (IBD) and SLE datasets. In IBD, GlycA prediction in the ArrayExpress dataset MTAB331 suggests a strong correlation between GlycA and Crohn's Disease (CD) activity, measured by the Harvey Bradshaw Severity Index. GEO dataset GSE42296 predicts GlycA of CD responders to infliximab drops twice as much as the GlycA levels of non-responders (0.32 SD, compared to 0.17 SD). In SLE, dataset GE49454 predicts that GlycA levels of proliferative cases of SLE are 0.472 SD higher than non-proliferative SLE cases. In addition, SLE cases with renal involvement are predicted to have 0.355 SD higher GlycA concentrations than those without renal involvement. GlycA measurements by NMR confirm these predictions and are explored in detail in Chapters 5 and 6 for IBD and SLE, respectively.

Discussion

Though many aspects of the work reported in the preceding results section can be improved and expanded upon, it shows conclusively that platform independent imputation of serum GlycA levels from blood gene expression profiles is possible. The methods described here do not yet sufficiently address the biases inherent to each of the transcriptomic platforms and therefore do not allow for comparison of predicted GlycA levels between different transcriptomic datasets. However, within-dataset performance suffices for the recapitulation of published GlycA associations and even for the identification of potential novel GlycA associations. In what follows, I highlight several areas in which further work is necessary to improve the predictive models.

First, our chosen data pre-processing, Quantile Discretization, succeeds in making the range of input data uniform across the different transcriptomic datasets, but does little to address their inherent biases. Quantile Discretized data of the D07, D14 and YFS studies is still separable with a high degree of accuracy using a handful of randomly selected features (data not shown). Furthermore, the arbitrary choice of deciles for the QD method should be expanded upon: though decile discretization was equally effective as percentile discretization of the data (data not shown), a robust assessment of the optimal choice for the number of quantiles has yet to be performed.

Second, it should be possible to refine the feature selection method by performing the Boruta algorithm in all three datasets separately, and creating a core set of related genes that is confirmed across all validation runs in each of the three datasets. The resulting core gene-set should be more robustly informative to the GlycA levels in the third-party datasets. Comprehensive research of these predictive models should also trial multiple strategies for feature selection. Indeed, a comparison of multiple feature selection methods and their stability would be inherently valuable information [270], and will be included when the manuscript of these results is submitted for publication.

Third, it can be argued that hyperparameter tuning by RMSE minimization in the YFS training data is suboptimal if eventual selection of the ideal model is based on correlation between prediction and observed values in unseen data. The performance of some models, e.g. the kNN models, declines sharply when data from a different platform is used as input. If a robust model is preferable, then hyperparameters should be optimized to maximize this robustness. In the results reported here, the models with the highest performance in the training data explain roughly 40% of the variance, even though they have a RMSE error only 25% lower than completely random models (Figure 17), further illustrating optimization to RMSE might be a suboptimal choice. RMSE optimization was a natural choice if the smallest possible error between prediction and reality was the objective, but our results show that our normalization strategy does not eliminate the strong inherent biases in the transcriptomic datasets to make absolute GlycA quantification a tractable goal. Focusing on relative estimates and limiting predictions to within-dataset comparisons and associations is a more realistic application, the potential of which is already evident from the identification of two novel potential contexts where GlycA can be useful (IBD and Lupus Nephritis, explored thoroughly in Chapters 5 and 6, respectively).

Fourth, more methods for generating the predictive models can still be tested. Random Forest (RF) models in particular are an interesting option which has not yet been investigated: considering the feature selection method that was used to generate these results (i.e. Boruta) estimates feature importance through the loss of accuracy by omission in RF regression, it would be worthwhile to test if this favors the subsequent performance of RF models over that of other approaches. In addition,

implementations of state-of-the-art gradient boosting methods have recently become available (e.g. catboost, xgboost [274], lightboost) and should be tested. The results generated thus far do suggest that the relation between GlycA and the blood gene expression profile is linear in nature and the high performance of the linear models in the unseen test data suggests that simpler models are more robust to input data that suffers from large inherent biases than more sophisticated models.

Finally, creating a regression model for a continuous inflammatory biomarker is inherently more challenging than creating a classification model for an active inflammatory state. However, GlycA apparently summarizes such a broad and heterogenous range of inflammatory pathways that quantification is to be preferred over classification: GlycA is a read-out that incorporates the effects of many 'confounding', read: 'contributing', factors. This layered increase of GlycA in the presence of more inflammatory symptoms is neatly illustrated in the higher predicted GlycA levels of arthritic psoriasis compared to the predicted values of cutaneous psoriasis without arthritis (Table 1) and is also observed in our experiments in SLE (Chapter 6). Furthermore, association of gene signatures to binary outcomes are frequently spurious: one study showed that the majority of published gene signatures related to breast cancer outcome do not significantly outperform random signatures [275]. A continuous response variable is inherently less likely to facilitate this problematic behavior, though vigilance is warranted and comparisons to models generated on randomly selected features should always be included (Figure 17).

In conclusion, I have shown that the re-use of transcriptomic data to construct platform independent predictive models for novel biomarkers is feasible, though not without its challenges. Additional tests are required to round out the work reported in this chapter, in order to present a comprehensive study into this topic. Two novel contexts where GlycA may prove useful, i.e. IBD and Lupus Nephritis, are examined in detail in Chapters 5 and 6, respectively.

CHAPTER 5

GlycA, a nuclear magnetic resonance spectroscopy measure for protein glycosylation, is a viable biomarker for disease activity in IBD.

Research Article

Published in *Journal of Crohn's and Colitis* (epub ahead of print)
Dierckx T.*, Verstockt B.*, Vermeire S., Van Weyenbergh J. (2018). **GlycA, a nuclear magnetic resonance spectroscopy measure for protein glycosylation, is a viable biomarker for disease activity in IBD**.

doi: 10.1093/ecco-jcc/jjy162.
PMID: 30312386

* Both authors contributed equally to this work

Abstract

Background and aims

Glycoprotein acetylation (GlycA) is a novel nuclear magnetic resonance (NMR) biomarker measured in serum or plasma, which summarizes signal originating from glycan groups of certain acute phase glycoproteins. This biomarker has been shown to be robustly associated to cardiovascular and short-term all-cause mortality, and to disease severity in several inflammatory conditions. We investigated GlycA levels in a cohort of healthy individuals (HC), Crohn's disease (CD) and ulcerative colitis (UC) patients prior to and after therapeutic control of inflammation.

Methods

Serum samples of 10 HC, 37 CD patients and 21 UC patients before and after biological therapy were subjected to high throughput NMR analysis by Nightingale Health Ltd. Paired C-reactive protein (CRP) and fecal calprotectin (fCal) measurements were used to characterize baseline differences, treatment effects and post-treatment association to endoscopic response (50% SES-CD decrease at week 24) and mucosal healing (SES-CD≤2 for CD, Mayo endoscopic score ≤1 for UC).

Results

GlycA levels were significantly higher in patients with active IBD compared to healthy controls, and accurately reflected the mucosal recovery to a 'healthy' state in both CD and UC patients achieving mucosal healing. In CD patients who experienced an endoscopic response without achieving full mucosal healing, GlycA levels also decreased but did not normalize to HC levels. Overall, GlycA correlated well with CRP and fCal, and accurately tracked disease activity in CRP negative patients (<5 mg/dL).

Conclusion

GlycA holds promise as a viable serological biomarker for disease activity in IBD, even in patients without elevated CRP, and should therefore be tested in large prospective cohorts.

Introduction

Glycans are complex oligosaccharides that play a fundamental role in human health and contribute to the development of many complex inflammatory diseases [276]. Glycoprotein acetylation (GlycA) is a novel nuclear magnetic resonance (NMR) biomarker measured in blood serum or plasma, which summarizes the NMR signal originating from glycan groups of certain acute phase glycoproteins (mainly α-1-acid glycoprotein, haptoglobin, α-1-antichymotrypsin and transferrin) [62], [261]. GlycA has been extensively investigated in cardiovascular disease (CVD), where it was found to be robustly associated to atherosclerosis and risk of future CVD [62], [277]. Large demographic studies have also shown an association to cardiovascular and short-term all-cause mortality [70]. Robust associations to both CVD and diabetes in large cohorts have prompted the conclusion that GlycA summarizes the summative risk resulting from multiple inflammatory pathways [76], [277], [278]. In several inflammatory conditions such as rheumatoid arthritis [83], [84], Kawasaki disease [57] and psoriasis [90], GlycA has been shown to be associated with disease severity even after adjustment for traditional acute inflammation metrics such as C-Reactive Protein (CRP). Furthermore, GlycA levels decrease during successful treatment of psoriatic skin inflammation with anti-TNF therapy [90], suggesting it could be a clinically relevant biomarker for monitoring disease severity.

Despite growing evidence on the role of glycosylation in other immune-mediated entities, including inflammatory bowel disease (IBD) [276], [279], [280], GlycA has not yet been studied in this context. Current IBD treatment algorithms are not solely symptom driven, but additionally guided by biomarkers, an approach which has been shown to improve endoscopic outcomes [281]. However, CRP has little discriminatory power in IBD: many, but not all, patients with Crohn's disease (CD) show a strong CRP response, whereas patients with ulcerative colitis (UC) generally only have a modest or absent CRP response [56]. A better alternative biomarker has been found in fecal calprotectin (fCal), which correlates well with mucosal inflammation and can be used as a surrogate marker for mucosal healing in both UC and CD [282]. However, in

many countries serial fCal measurements are not reimbursed, and patients generally do not prefer (regular) fecal sampling [283].

We therefore investigated GlycA in a cohort of healthy individuals, CD and UC patients. We not only studied differences between healthy and affected individuals, but also collected post-treatment samples allowing associations with mucosal healing relative to established inflammatory biomarkers such as CRP and fCal.

Materials and methods

Study design and patients

We conducted this prospective study at the tertiary IBD referral center of the University Hospitals Leuven (Leuven, Belgium). All patients included in the analysis had given written consent to participate in the Institutional Review Board approved IBD Biobank (B322201213950/S53684), collecting serum and clinical (baseline features, patient reported outcomes (PRO2), Harvey-Bradshaw Index (HBI)) characteristics among other items.

We randomly selected samples from patients with active endoscopic disease initiating biological therapy. Patients with an ostomy were excluded. All patients were prospectively monitored, including clinical and endoscopic assessment, CRP and fCal measurements at predefined outcomes (baseline and 6 months for CD patients, baseline and week 8 (adalimumab, ADM) or week 14 (infliximab, IFX, and vedolizumab, VDM) for UC patients). In addition, serum of 10 gender- and age-matched healthy controls (HC) was collected.

Outcomes

Endoscopic outcomes were assessed 6 months after treatment initiation in CD patients[284], whereas in UC patients timing depended on the national reimbursement criteria of the individual drug (8 weeks for ADM, 14 weeks for IFX and VDM). In UC patients, mucosal healing was defined as a Mayo endoscopic subscore ≤ 1, whereas in CD patients mucosal healing was defined as Simple Endoscopic Score for Crohn's

disease (SES-CD) ≤ 2. Because of the low MH rates in ustekinumab treated patients [285], only endoscopic response (minimal 50% decrease in SES-CD) was evaluated.

Samples

Serum samples were collected at baseline prior to the first administration of the drug, and during maintenance. Samples were centrifuged and stored aliquoted at -20°C. Fecal samples were collected at home, stored at 4°C in the home fridge and transported cooled within 24 hours to the hospital. fCal measurements were performed for all patients with the fCAL ELISA kit (Bühlmann, Schönenbuch, Switzerland). CRP was determined by the routine laboratory of the University Hospitals Leuven. GlycA concentration was quantified using the Nightingale Health Ltd. high-throughput metabolomics platform (Helsinki, Finland) [62], [261].

Statistical analysis

Continuous variables are expressed as median and interquartile range (IQR). The Wilcoxon rank-sum test was used to compare GlycA levels of HC samples to baseline CD and UC samples, and to compare patients showing MH to those without signs of healing in these conditions. All statistical analyses were performed in R, version 3.4.3 (R Development Core Team, Vienna, Austria) [286] using the base stats package. Graphics were generated using the ggplot2 R package [287].

Results

Patient characteristics

Fifty-eight IBD patients (37 CD, 21 UC) were included in this prospective, observational study prior to the initiation of biological therapy (9 ADM, 13 IFX, 24 VDM and 12 UST). Of the examined patients, 34.5% (n=20) did not have an elevated CRP at baseline. In this cohort of patients with a median disease duration of 4.8 (2.4-14.9) years, mucosal healing was obtained in 39.7% (n=23) after a median of 23.7 (22.4-24.7, CD) and 13.7 (11.0-14.3, UC) weeks respectively (Table 1).

GlycA accurately reflects disease activity

GlycA concentrations were significantly increased in both CD and UC patients compared to HCs ($p<10^{-4}$ and $p<10^{-3}$). Variability of GlycA was likewise higher in CD and UC patients than what is observed in HC (coefficient of variation of 14.8%, 15.5% and 8.0%, respectively). At baseline, no significant difference between GlycA levels of CD and UC patients (p=0.92) was observed, and GlycA concentration in both CD and UC patients achieving mucosal healing dropped back to HC levels (p=0.90, p=0.91) (**Figure 19A and 19C**). While CD patients responding well to UST treatment (without achieving mucosal healing) showed a significant decrease of GlycA concentration in comparison to their baseline measurements (p=0.03), their post-treatment GlycA levels remained elevated in comparison to HC levels (one sided wilcox p=0.006) (**Figure 19B**). The observed drop in GlycA levels during maintenance therapy was consistent regardless of treatment used (supplementary figure S6). Analysis of variance across the different treatments found no significant differences between the baseline GlycA, post-treatment GlycA and the difference between these two timepoints of the different treatments.

In patients without elevated CRP (< 5mg/dL) at baseline (n=20, 11 CD, 9 UC), GlycA levels were significantly higher than those observed in healthy controls (p<0.01) (**Figure 20**). In healers, GlycA levels dropped to levels similar to HC, while in non-healers GlycA levels remained elevated compared to HC (p=0.07).

GlycA correlated well with both fCal and CRP (spearman ϱ=0.39, p=0.02 and ϱ=0.65, $p<10^{-8}$, respectively) **(Figure 21**). Across the three considered settings (IFX, ADM or VDZ treated CD, UST treated CD and IFX, ADM or VDZ treated UC), of these three biomarkers, only GlycA post-treatment levels consistently showed a significant difference between responder and non-responder levels (**Figure 19**). Endoscopic activity, quantified by the Mayo endoscopic subscore in UC, correlated well with GlycA (ϱ=0.51, p=0.0016) but not with the Simple Endoscopic Score for CD (SES-CD), though this latter measure was only available for 15 of 37 CD patients. Disease location was associated to GlycA levels, with a Wilcoxon test showing a borderline significant (p=0.06) difference between GlycA levels of L1 and L3 CD patients and

UC patients with E1 disease showing significantly lower GlycA levels than those of E2/E3 afflicted patients (wilcoxon p=0.04). CD patients with prior enterectemy showed similar baseline GlycA levels as patients without occurrence of any resection of a part of the gut, or stricturoplasty for stenosing complications. The greater GlycA decrease observed in treatment responders compared to non-responders was observed in resectioned patients and non-resectioned patients alike, indicating prior enterectemy does not impact GlycA's association to disease activity. Finally, the expected association of GlycA to BMI, robustly shown in large demographic studies [71], [88], was absent in this IBD cohort, analogous to observations in rheumatoid arthritis [84].

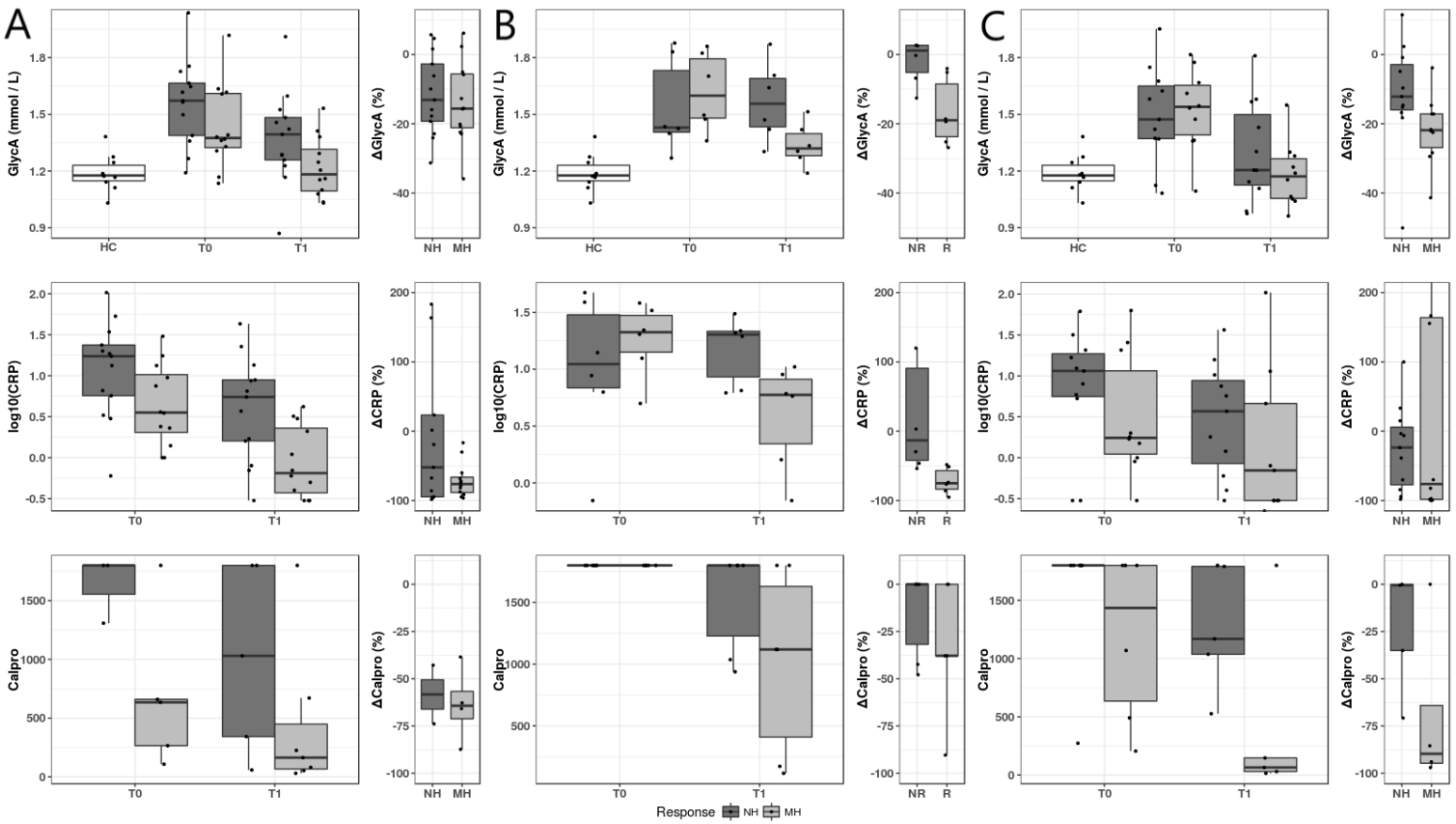


Figure 19: GlycA, CRP and fCal measurements. The top row depicts GlycA concentrations in healthy controls (HC), and IBD patients at baseline (T0) and during maintenance therapy (T1) as well as the observed percentage change after treatment. The middle row depicts CRP measurements, the bottom row depicts fCal levels. From left to right, samples originate from A) IFX, ADM or VDM treated CD, B) UST treated CD and C) IFX, ADM or VDM treated UC. Patients are divided into non-responding (no mucosal healing, NH or no endoscopic response, NR, depicted in dark gray) and responding (mucosal healing, MH or endoscopic response, R, depicted in light gray) groups.

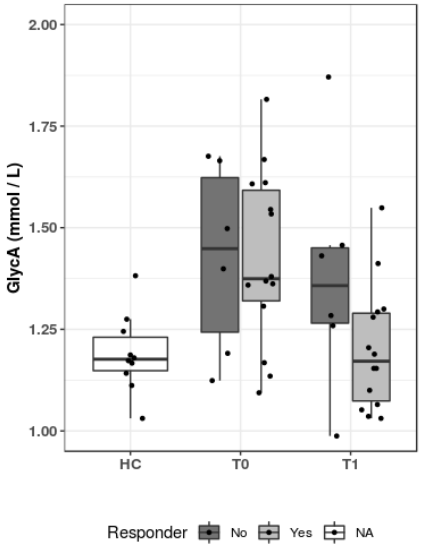


Figure 20: GlycA levels of CRP negative patients. Patients with healthy CRP levels (< 5 mg/dL), (n=20, 11 CD, 9 UC) have significantly increased GlycA levels over healthy controls ($p<0.01$). After treatment (T1), patients that responded to treatment return to GlycA levels observed in HC, while non-responder GlycA levels remain elevated over HC levels ($p=0.07$).

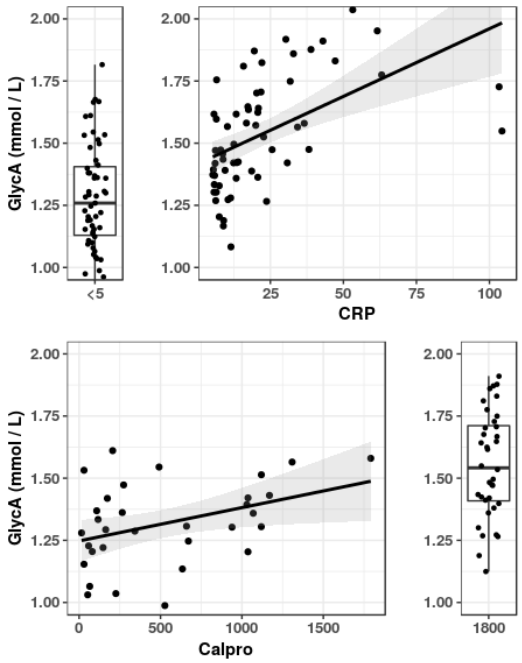


Figure 21: GlycA, CRP and fCal correlations. Correlation between GlycA and CRP (spearman $\varrho=0.40$, $p=0.02$, top) and fCal (spearman $\varrho=0.61$, $p<10^{-8}$, bottom) is illustrated in all available data. Data near the lower detection limit for CRP (<5 mg / L) or the upper detection limit for fCAL (>1800 µg / g) are depicted as boxplots alongside the scatterplot.

Discussion

In this prospective pilot project, we provide the first evidence that GlycA can be a relevant biomarker in the context of IBD. GlycA levels were found to be significantly higher in patients with active IBD, compared to healthy controls. As elevated GlycA levels have previously been associated with atherosclerosis and cardiovascular disease [62], [277], our observations support previous findings regarding increased cardiovascular risk reported in IBD due to the persistence of chronic inflammation [288]. While pre-treatment GlycA levels showed no association to eventual treatment outcome, a treatment-induced decrease in GlycA was strongly associated with successful treatment response and mucosal healing, and associated to treatment response at least as well as traditional inflammatory biomarkers such as CRP or fCal. GlycA accurately reflected the mucosal recovery to a healthy state, as evidenced by the return to healthy GlycA levels in healers, but not in non-healers or partial responders.

The discovery of GlycA as a relevant biomarker for IBD disease activity makes pathophysiological sense, as glycans are important in fine tuning immune responses [280]. Furthermore, recent mass spectrometry data demonstrated that plasma N-glycomes show distinct glycosylation patterns differentiating CD and UC patients, and that these patterns can be associated with disease progression and the need for more potent medication and surgery [289]. The method used in our report summarizes NMR signal originating from acetyl groups of specific glycosylated proteins present in these patterns (specifically haptoglobin and α1-acid-glycoprotein) [289].

Our results suggest that GlycA is a viable serological biomarker to monitor disease activity and treatment response in CRP-negative IBD patients. The advantage of GlycA over CRP likely stems from its greater stability, owing to its composite nature as a summary measure of glycan signal originating from several acute-phase proteins. It has been shown that calculating the true set point for biological homeostasis of CRP requires 33 concurrent measurements while for GlycA only a single measurement is required [57]. Contrary to CRP, GlycA has been shown to be a stable and sensitive marker for chronic inflammation in rheumatoid arthritis[84]. Additionally, replacing CRP by another serological biomarker in patients who cannot be biochemically

monitored using CRP might be a better alternative than a fecal biomarker, as patients prefer regular blood instead of regular stool sampling [283]. The cost of GlycA determination is comparable to that of fCal measurement (20-50 euro, depending on sample size, and approximately 25 euro, respectively). Though both biomarkers are considerably more expensive than a CRP measurement (approximately 5 euro), the NMR spectrum used to quantify GlycA additionally quantifies several metabolic biomarkers such as amino acid levels, glycolysis and fatty acid metabolites and an extensive lipoprotein profile.

We recognize that our study has both strengths and weaknesses. To our knowledge, this is the first study which assesses glycosylation levels of plasma proteins through NMR in IBD patients. The study had access to both baseline pre-treatment samples and post-treatment samples of CD and UC patients, which facilitated characterization of baseline differences, treatment effects and post-treatment association to mucosal healing and treatment response. However, sample size was limited, necessitating the grouping of samples from several treatment regimens for meaningful statistical analysis, and would need to be increased to assess the effects of, and biomarkers for successful treatment with, specific drug regimens. Likewise, our sample numbers necessitated the grouping of CD and UC samples to assess GlycA's potential as a biomarker in pre-treatment CRP negative patients. The limited availability of certain clinical metadata in this pilot study, specifically fCal and SES-CD, presents a strong incentive for a more extensive follow-up study. Finally, the GlycA biomarker summarizes the degree of acetylation of specific acute-phase glycoproteins and, to our knowledge, it is currently not known whether the principal contribution underlying the associations with GlycA can be ascribed to the concentrations of these proteins, their glycosylation or their acetylation profiles.

In conclusion, we identified GlycA as a promising candidate biomarker for monitoring disease activity in IBD patients, even in patients without elevated CRP. Additional experiments in larger cohorts are necessary to confirm our findings, and to elucidate whether measuring the GlycA component glycoproteins directly is a low-cost proxy which can track successful treatment response as accurately as the GlycA measurement by NMR.

Tables

Table 1 Baseline disease characteristics

	Crohn's disease (n=37)	**Ulcerative Colitis (n=21)**
Sex, women, *n* (%)	15 (40.5)	9
Disease duration, *y,* median (IQR)	4.6 (2.6 – 16.3)	5.0 (1.8 – 9.6)
Age at inclusion, *y,* median (IQR)	30.2 (22.9 – 41.6)	37.5 (26.2 – 47.2)
Disease location, *n* (%)	L1 Ileal disease 10 (27) L2 Colonic disease 9 (24.3) L3 Ileocolonic disease 18 (48.6)	E1 Proctitis 3 (14.3) E2 Left sided 8 (38.1) E3 Pancolitis 10 (47.6)
Disease behavior, *n* (%) - Inflammatory (B1) - Stricturing (B2) - Penetrating (B3) - Perianal disease (p)	 22 (59.5) 5 (13.5) 10 (27.0) 12 (32.4)	 NA
Smoking status, *n* (%) - Active smoking - Previously smoking - Never smoked	 7 (18.9) 10 (27.0) 20 (54.1)	 0 (0.0) 6 (28.6) 15 (71.4)
Body Mass Index, kg/m², median (IQR)	21.6 (19.8 – 25.3)	22.9 (22.0 – 26.4)
C-reactive protein, *mg/L*, median (IQR)	12.5 (3.5 – 22.1)	5.9 (1.5 – 20.7)
Fecal calprotectin, *μg/g*, median (IQR)	1800.0 (1554.2 – 1800.0)	1800.0 (924.8 – 1800.0)
PRO2, median (IQR)	15.0 (7.0 – 21.0)	4.0 (4.0 – 5.0)
Initiated biological therapy, n (%) - Adalimumab - Infliximab - Vedolizumab - Ustekinumab	 5 (13.5) 8 (21.6) 12 (32.4) 12 (32.4)	 4 (19.0) 5 (23.8) 12 (57.1) NA
Timing of endoscopic assessment, weeks, median (IQR)	23.7 (22.4 – 24.7)	13.7 (11.0-14.3)

IQR = interquartile range; n = number of patients; PRO2 CD = patient reported outcome for Crohn's disease = 5x abdominal pain score + 2x liquid stool frequency; PRO2 UC = patient reported outcome for ulcerative colitis = stool frequency + rectal bleeding

Supplements

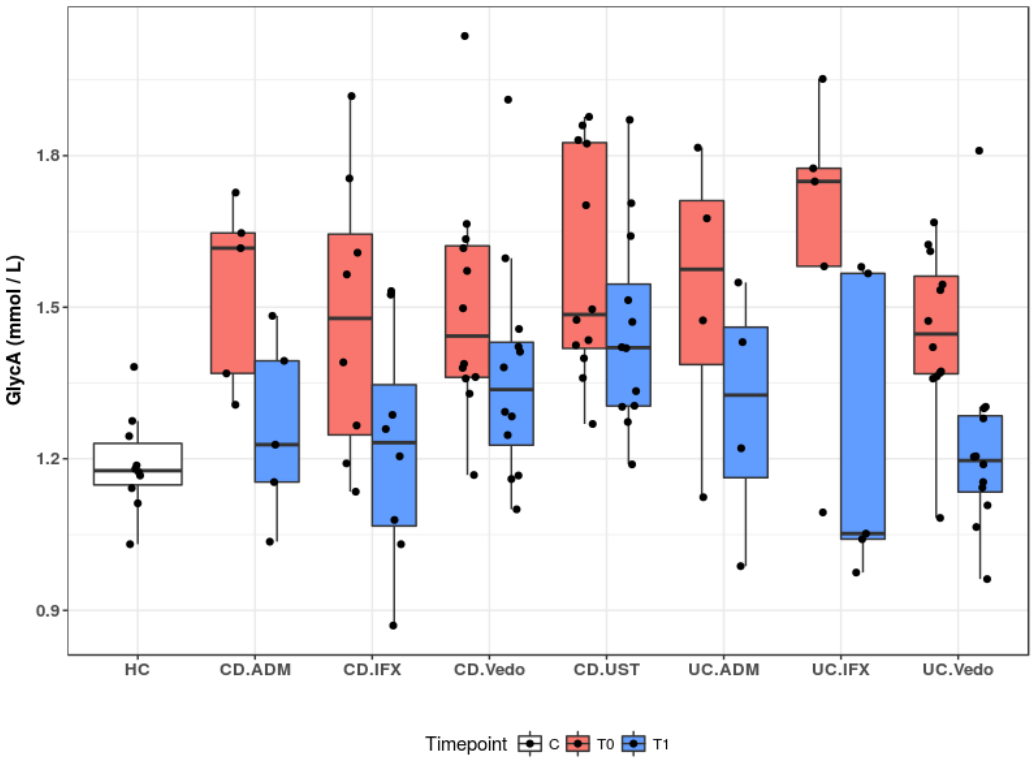


Figure S6: GlycA levels of all patients, prior to (T0) and post treatment (T1), separated by treatment. A consistent decrease in GlycA concentration is observed, regardless of treatment type. ADM, adalimumab, IFX, infliximab, Vedo, vedolizumab.

CHAPTER 6

Serum GlycA level is a candidate biomarker for disease activity in systemic lupus erythematosus and for proliferative status of lupus nephritis, independent of renal function impairment.

Research Article

Under review for *Scientific Reports*, 2019
Dierckx T., Goletti S., Chiche L., Daniel L., Lauwerys B., Jourde-Chiche N., Van Weyenbergh J. **Serum GlycA level is a candidate biomarker for disease activity in systemic lupus erythematosus and for proliferative status of lupus nephritis, independent of renal function impairment.**

bioRxiv doi: 10.1101/493809.

Abstract

Objective

Glycoprotein acetylation (GlycA) is a novel biomarker for chronic inflammation, associated to cardiovascular risk. Serum GlycA levels are increased in several inflammatory diseases, including systemic lupus erythematosus (SLE). We investigated the relevance of serum GlycA measurement in SLE and lupus nephritis (LN).

Methods

GlycA was measured by NMR in 194 serum samples from patients and controls. Comparisons were performed between groups. Clinical and biological parameters were tested for correlation with GlycA levels. The predictive value of GlycA to differentiate proliferative from non-proliferative LN was determined using logistic regression models.

Results

GlycA was correlated to C-reactive protein (CRP), neutrophil count, proteinuria and the SLE disease activity index (SLEDAI), and inversely with serum albumin. GlycA was higher in active (n=105) than in quiescent (n=39) SLE patients, in healthy controls (n=29), and in patients with non-lupus nephritis (n=21), despite a more altered renal function in the latter. In patients with biopsy-proven active LN, GlycA was higher in proliferative (n=32) than non-proliferative (n=11) LN, independent of renal function and proteinuria level. Logistic regression models showed that, in univariate models, GlycA outperforms traditional biomarkers. A bivariate model using GlycA and BMI better predicted the proliferative status of LN than a model comprising CRP, renal function (eGFR), serum albumin, proteinuria, C3 consumption and the presence of anti-dsDNA antibodies.

Conclusion

Serum GlycA is elevated in SLE, and correlates with disease activity and LN. Serum GlycA, which summarizes different inflammatory processes, could be a valuable biomarker to discriminate proliferative from non-proliferative LN and should be tested in large, prospective cohorts.

Introduction

Systemic inflammation is implicated in a wide range of auto-immune diseases such as systemic lupus erythematosus (SLE) [290]. In SLE, a multitude of different inflammatory cytokines and pathways can be activated [291] and these are not always reflected by elevated C-reactive protein (CRP). Lupus nephritis (LN) is one of the most severe complications of SLE which occurs in 20-70% of patients and its occurrence is not reliably predicted by classical inflammatory markers [290]. The occurrence and severity of LN has been associated with the dysregulation of neutrophil extra-cellular traps (NETs) [292] and a blood transcriptional neutrophil signature [293]. Non-invasive biomarkers are needed in SLE to predict the risk of LN in SLE patients, and to determine the severity of LN by differentiating between the proliferative and non-proliferative forms of LN, which determines whether immunosuppressive therapy will be needed to achieve renal remission [294].

Glycoprotein Acetylation (GlycA) is a new nuclear magnetic resonance (NMR) spectroscopy-derived biomarker of systemic inflammation that reflects protein glycosylation which could prove to be promising biomarker in the LN context. GlycA signal, measured in the blood serum or plasma, reflects mainly the glycosylation of acute-phase proteins α1-acid glycoprotein, haptoglobin, α1antitrypsin, α1antichymotrypsin and transferrin [62], [261], as a consequence of inflammatory stimuli. The GlycA signal is thus evaluated as a biomarker of systemic inflammation and cardiovascular risk [295], summarizing the activation of multiple inflammatory pathways [278]. GlycA has been shown to be associated with the risk of cardio-vascular events and mortality in the general population, independently of CRP [71], [89]. In a large Brazilian cohort, GlycA was associated with CRP, age, female gender, tobacco and alcohol consumption, obesity, diabetes, hypertension and dyslipidemia and associated independently with lower estimated glomerular filtration rate (eGFR) and albuminuria [88]. We recently showed that GlycA could be a marker of disease activity in inflammatory bowel disease, even in patients without CRP elevation [296]. Of particular interest to the LN context, large scale gene correlation network analysis has shown that GlycA can also be associated to NET formation [76]. Furthermore, concentrations of GlycA were noted to be higher in SLE patients than in healthy controls, but findings with regards to GlycA's association to disease activity vary between studies [86], [87]. The value of GlycA as a biomarker of LN has not been evaluated yet.

In this work, we investigated if GlycA could be associated with LN severity. We compared patients with proliferative and non-proliferative forms of biopsy-proven LN, and included a control group of patients with non-lupus renal diseases to account for the possible confounding factors of altered eGFR and albuminuria.

Materials and Methods

Patient demographics and Ethics

This study was conducted in accordance with the principles of the declaration of Helsinki. Patients with biopsy-proven LN, and patients with other renal diseases, were included in the biobank DC-2012-1704, approved by the French Ministry of Health, in the Hôpital de la Conception, Marseille, France. All patients gave written informed consent before any procedure. Other samples from patients with SLE were obtained from patients recruited in the LOUvain Lupus Nephritis InCeption (LOULUNIC) cohort, and patients followed-up at the Lupus Clinic of the Université Catholique de Louvain (UCL), Brussels, Belgium, as were samples from age- and sex-matched healthy controls. All patients and controls gave written informed consent before serum samplings. All patients with SLE responded to the SLICC 2012 classification criteria [297].

Clinical metrics

For this study, a total of n=194 samples were analyzed. A graphical overview of assayed samples can be found in **Figure 22**. We grouped the samples as either originating from healthy controls, from patients with biopsy-proven non-lupus renal diseases ("non-lupus nephritic controls", comprising patients with membranous nephropathy, IgA nephropathy, diabetic kidney disease or hypertensive nephropathy) and from SLE patients. SLE patients were further identified as clinically quiescent SLE patients (SLEDAI =< 4 without clinical activity, with or without maintenance therapy, immunological activity authorized), or active SLE patients. Active SLE patients were either SLE patients with an extra-renal flare without nephritic involvement, or patients with active LN (whether there is extra-renal activity or not). For differential analysis between proliferative (class III or IV +/- V, with active lesions, of the ISN/RPS 2003 classification) and non-proliferative (class I, II or isolated class V) flaring LN cases, a subset of samples was used, which originates from patients who were sampled at the time of the biopsy-proven LN. Patient numbers for this cohort, as well as full demographical

and clinical details, can be found in **Supplementary Tables 1 & 2**. For 28 SLE patients, longitudinal samples were available.

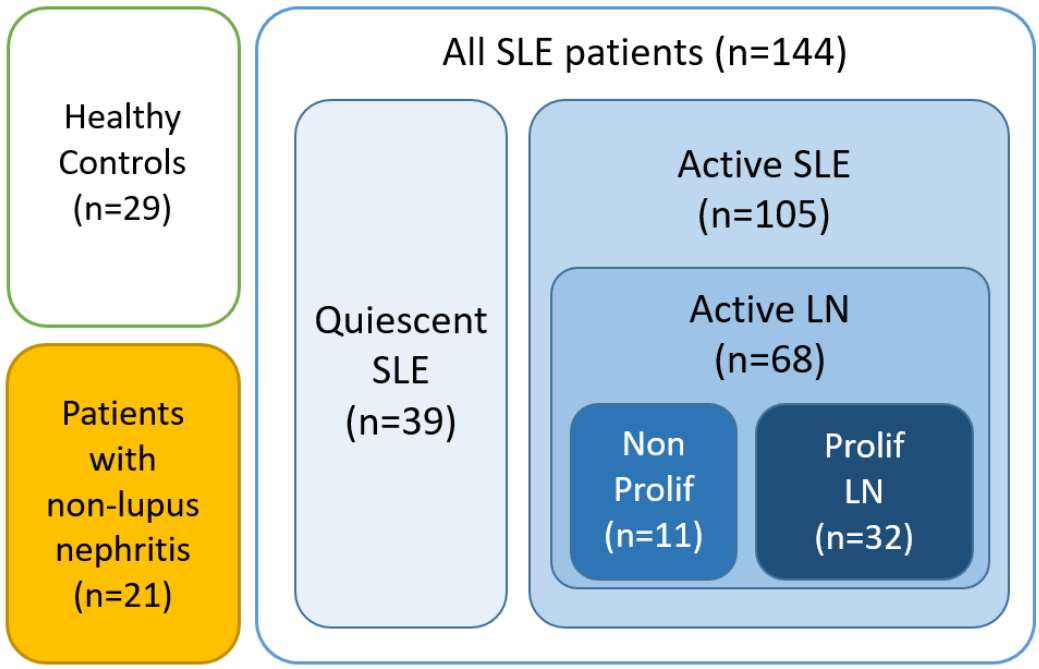


Figure 22: Overview of samples used in this study

Age, Gender, ethnicity, BMI, smoking status, extra-renal lupus activity (ongoing arthritic flare in particular), and diabetic status, were collected and tested for confounding effects on GlycA. Serum CRP level, serum albumin and creatinine levels, estimated glomerular filtration rate (eGFR, calculated with the MDRD equation [298], C3 and C4 concentration, urinary protein/creatinine ratio (UPCR), presence of anti-dsDNA antibodies as a binary variable (above 16 IU/mL by ELIA™, ThermoFisher, MA, USA, or above 10 IU/mL using the Farr assay from Trinity Biotech, Bray, Ireland), daily glucocorticoid dosage, and hydroxychloroquine usage were tested for association to GlycA on all available SLE samples.

GlycA concentration was quantified using the Nightingale Health Ltd. high-throughput metabolomics platform (Helsinki, Finland) [62], [261]. Laboratory and NMR measurements of creatinine and albumin concentrations were found to be highly correlated (ϱ=0.94 and ϱ=0.74, respectively, with p-values<10^{-8}) (**Supplementary Figure 7**).

Statistical analysis

Univariate comparison between the different conditions was performed using unpaired, two-sided Wilcoxon rank sum test. Association between GlycA and all available clinical and demographic data was tested using Spearman's ρ for continuous variables and Mann-Whitney-Wilcoxon's U for categorical variables, p values were corrected for multiple testing using the Benjamini-Hochberg procedure and resulting q values are accepted as significant when smaller than 0.05. Identification of all confounding factors on the relationship between GlycA and proliferative status was achieved with analysis of covariance methods calculating the Type III Sums of Squares in a stepwise forward selection. Starting from the model including only the GlycA response variable and proliferative status, in each step the variable with the most significant contribution (with maximum *p*-value 0.1), as determined in a comparison with an F-test, was included in the model. Discriminatory power of clinical and demographical parameters was assessed in logit-link logistic regression models of LN patients with proliferative status of the patient as the response variable. In addition to univariate and selected multivariate models, a statistically ideal logistic regression model was constructed using an exhaustive best subset algorithm using Aikake's Information Criterion [299]. All statistical analyses were performed in R [286]. C statistics for logistic regression models were calculated using the Epi R package [300]. Leave-one-out cross-validated accuracy measures were calculated using the caret R package [271]. Figures were generated using the ggplot2 R package [287].

Results

GlycA levels in SLE patients and controls

Patients with active SLE showed significantly higher GlycA concentration than healthy controls ($p=0.009$), non-lupus nephritic controls ($p=0.04$) and quiescent SLE patients ($p<10^{-6}$) (**Figure 23**). Quiescent SLE patients also displayed lower CRP levels than patients with active SLE ($p<10^{-3}$). However, in contrast to GlycA levels, CRP concentrations of non-lupus nephritic controls were not significantly different from those of active SLE patients ($p=0.86$). eGFR was significantly lower in non-lupus nephritic controls than in active SLE patients ($p<10^{-3}$), indicating that the increased GlycA levels observed in active SLE can't be solely attributed to a decrease in renal function (**Figure 23**).

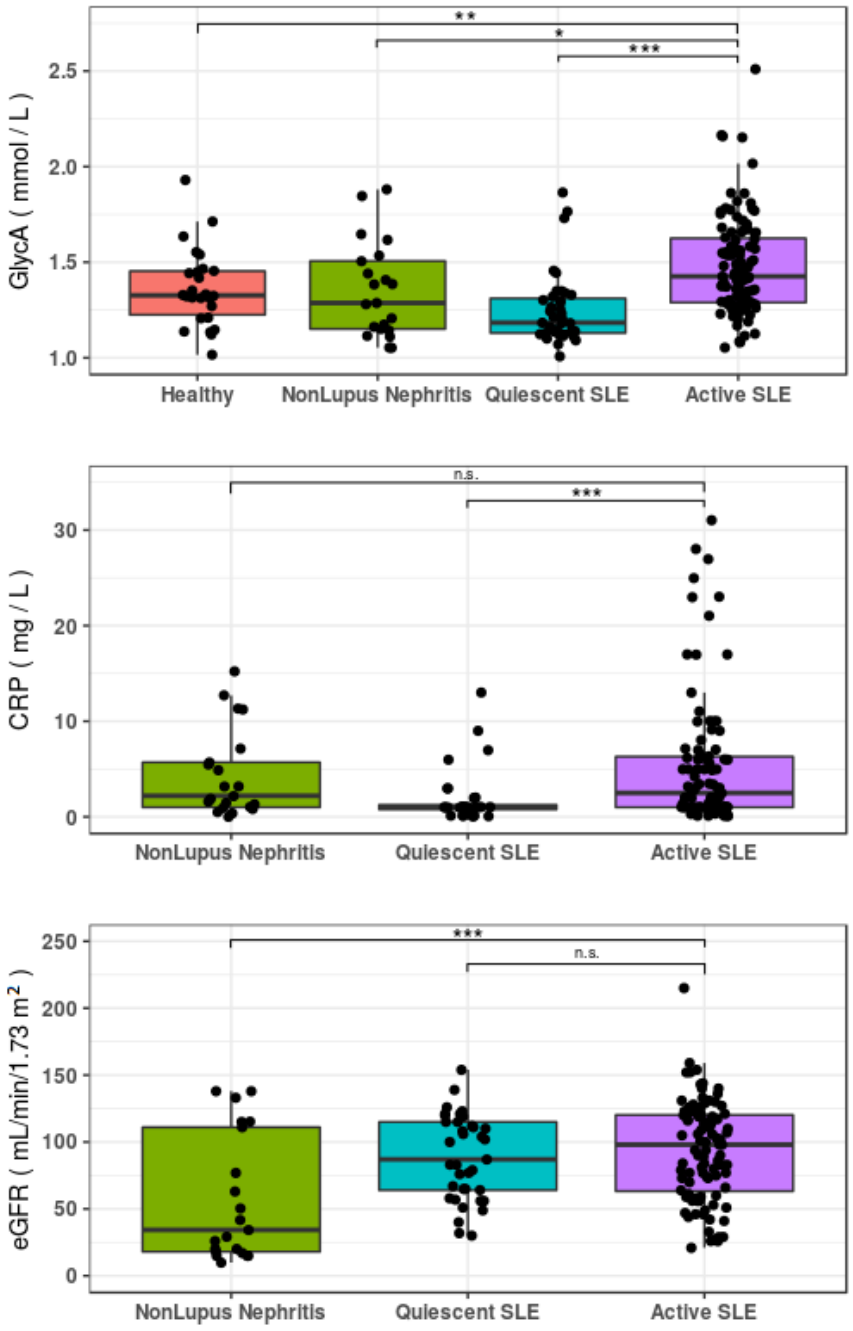


Figure 23: GlycA is elevated in active SLE, compared to healthy controls, non-lupus nephritic controls, and quiescent SLE. C-reactive protein (CRP) level is lower in quiescent SLE than in active SLE and non-lupus nephritis but does not differ between active SLE and non-lupus nephritis. Estimated glomerular filtration rate (eGFR) is lower in patients with non-lupus nephritis than in patients with quiescent or active SLE, excluding the fact that the elevation of GlycA in active SLE may be explained solely by decreased renal function. Significance of a Wilcoxon test comparing the groups to active SLE are indicated (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

GlycA association to clinical measurements

GlycA correlated well with CRP in all samples (ϱ=0.49, q<10^{-8}). Significant correlations to GlycA were observed for neutrophil counts (ϱ=0.32, q=0.04), serum albumin (ϱ=-0.37, q<10^{-4}), and UPCR (ϱ=0.38, q<10^{-4}), and trends for correlation were observed for serum creatinine (ϱ=0.17, q=0.06) and C3 (ϱ=-0.18, q=0.06). GlycA showed no correlation to anti-dsDNA levels measured by ELIA™ or Farr, and no significant difference was found between GlycA concentrations of dsDNA positive SLE patients, when compared to dsDNA negative SLE patients. GlycA levels showed a trend for inverse correlation with eGFR in quiescent SLE patients (ϱ=-0.37, q=0.07) and in non-lupus nephritic controls with a wide range of eGFR and proteinuria levels (ϱ=-0.46, q=0.06). In SLE samples, GlycA showed a significant correlation to the SLEDAI score (ϱ=0.36, q<10^{-4}). This correlation between the SLEDAI score and GlycA was absent if only quiescent SLE samples were considered. The full results of correlational analysis with GlycA can be found in **Supplementary Table 3**. **Figure 24** shows GlycA levels of available longitudinal samples of LN patients with a flare event. While insufficient observations are available to perform robust longitudinal analysis, we note that GlycA level is variable over time in a single patient, and that patients show increased GlycA at or directly following the flare presentation. Most patients show a return to pre-flare GlycA levels following the flare resolution.

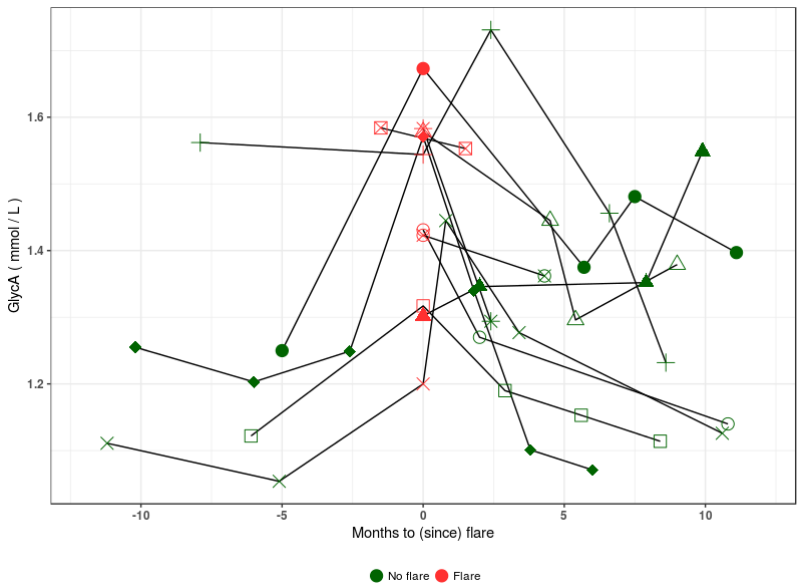


Figure 24: GlycA levels measured in longitudinal samples of flaring LN patients, up to a year prior and after flare event. Patients show increased GlycA levels at or directly following flare presentation. Post flare, most patients show a return to maintenance GlycA levels.

GlycA association to proliferative status in flaring LN

GlycA concentrations in serum samples taken at time of a flare of patients with biopsy-proven active LN were higher in proliferative cases than in non-proliferative cases (p=0.04). Although flaring proliferative LN cases had lower eGFR than non-proliferative cases, this difference was not statistically significant (p=0.28) and non-lupus nephritic controls with considerably lower eGFR than the proliferative LN patients (p<0.01) had GlycA levels comparable to those of healthy controls and non-proliferative flaring LN (**Figure 25**).

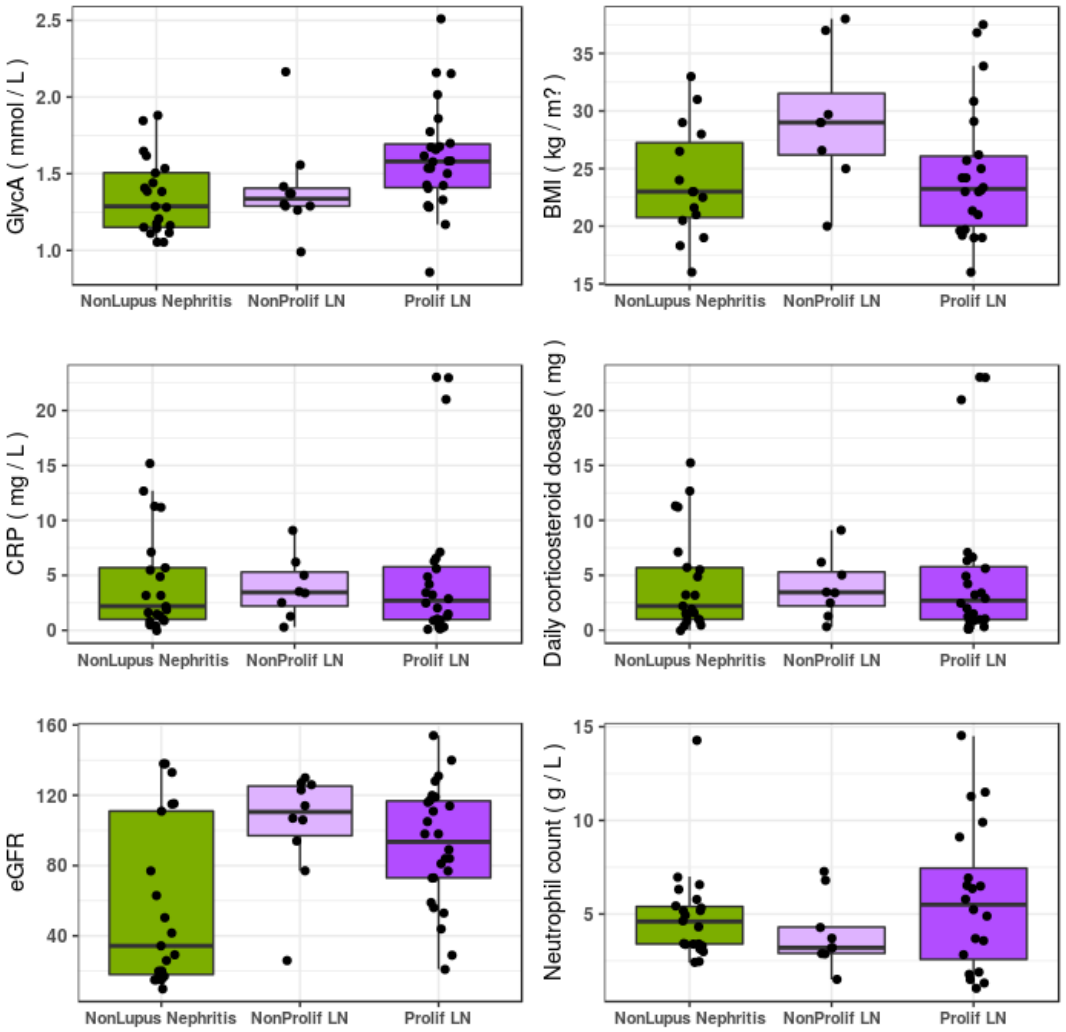


Figure 25: Comparison between LN samples taken at flare with biopsy-proven proliferative (or non-proliferative) status within 30 days of sampling shows significantly increased GlycA in proliferative LN. In addition, CRP and eGFR, as well as identified confounding factors BMI, daily corticosteroid dosage and neutrophil count are visualized in the flaring samples as well as in non-lupus nephritic control samples.

We sought to identify possible confounding factors interfering with GlycA levels' association to proliferative status of patients with flaring LN, such as BMI, which is known to be associated to GlycA, and was significantly higher in patients with non-proliferative than proliferative LN in this cohort ($p=0.04$). Analysis of variance identified confounding effects of BMI, daily corticosteroid dosage and neutrophil count as having independent and significant impacts on proliferative status' association to GlycA levels. After controlling for BMI, daily corticosteroid dosage and neutrophil concentration, proliferative status retained its association to GlycA.

Logistic regression models of proliferative status

We assessed the discriminatory power of GlycA in logistic regression models built to differentiate between proliferative and non-proliferative LN. **Supplementary Table 4** summarizes the predictive power of all univariate and specific extended logistic models, and the performance of select models is visualized in **Figure 26**. We point out that CRP has little to no discriminatory power in this setting (c statistic 0.56) and that the best univariate models are those constructed using GlycA or BMI (c statistic 0.72 and 0.75, respectively). A bivariate model using only GlycA and BMI (c statistic 0.91) outperforms a model constructed of six measures used in current clinical practice (i.e. CRP, serum albumin concentration, eGFR, proteinuria, C3 levels and evidence of anti-dsDNA, c statistic 0.90). We emphasize that GlycA and BMI independently contribute information about the inflammatory burden, evidenced by their independent significance in the multivariate model. Algorithmic construction of a logistic regression model using the exhaustive best subset approach results in a model using BMI, GlycA, presence of anti-dsDNA and daily corticosteroid dosage, capable of perfectly separating all non-proliferative from the proliferative cases.

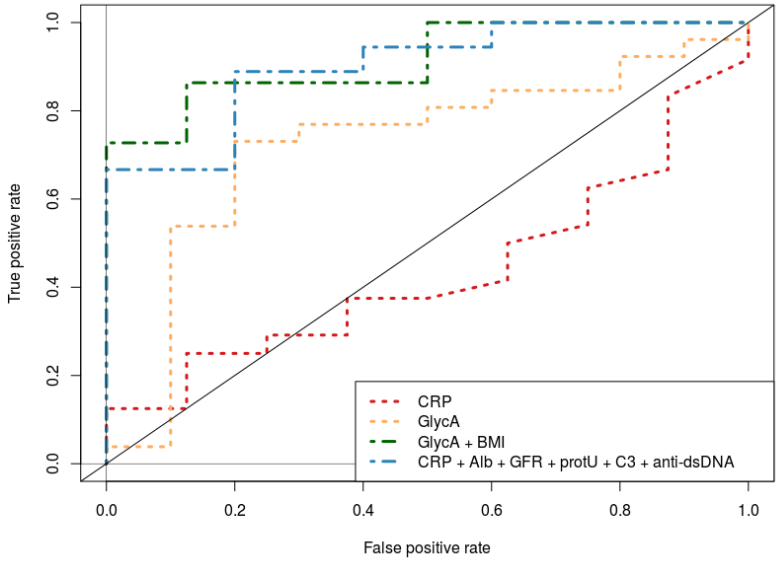


Figure 26: Receiver operating characteristic curve of selected logistic regression models predicting the proliferative nature of LN cases, using the variables indicated in the legend.

Discussion

The results of the present study confirm that GlycA is elevated in SLE patients compared to healthy controls. We extend previous findings and show that GlycA level is correlated with SLEDAI and is higher in active SLE than in quiescent SLE. Moreover, in patients sampled at the time of a biopsy-proven LN flare, we show that GlycA level is higher in patients with proliferative LN than in patients with non-proliferative LN. Interestingly, although we confirm the inverse correlation between GlycA and eGFR (GlycA increasing as renal function declines), we show that the elevation of GlycA in active SLE or in proliferative LN is not solely due to an altered renal function, since non-lupus nephritic controls, who displayed lower eGFR than SLE patients, had lower GlycA levels. Moreover, the association between GlycA level and the proliferative status of LN persisted after correction for eGFR. Interestingly, after correction for other confounding factors, whether they were previously described (BMI [69], [78], [278], [301] and daily corticosteroid dosage [87]) or observed in this work (neutrophil count), GlycA concentration remains significantly associated with the proliferative status of flaring LN patients. The increased GlycA concentration of proliferative LN is not explained by massive proteinuria, as proteinuria levels were comparable between patients with proliferative and non-proliferative LN.

We constructed logistic regression models to determine if the GlycA marker could be of use in the prediction of the proliferative status of LN. We show that GlycA, as a robust biomarker summarizing chronic inflammatory burden from multiple sources, has high predictive value in this context. A bivariate logistic model using GlycA and BMI to predict proliferative status outperforms a model using CRP, eGFR, proteinuria, C3 levels, serum albumin concentration and the presence of anti-dsDNA.

The association between BMI and the non-proliferative status of LN in this cohort is interesting. Although obesity is not a risk factor for the development of SLE [302], the prevalence of obesity in patients with SLE is high, and higher BMI has been associated with disease activity [303], in addition to chronic inflammation and cardiovascular burden [304]. Higher BMI has also been associated with an increased risk of proteinuria in SLE patients [305], but this may either reflect active LN, or be the consequence of glomerular hyperfiltration and glomerulosclerosis induced by obesity itself [306]. Indeed, the detection of a significant proteinuria in a patient with SLE implies that a renal biopsy should be performed to differentiate a severe (proliferative) LN from a less severe (non-proliferative) form or from another kidney disease. The higher BMI observed in the group of patients with non-proliferative LN could be partly explained by the earlier development of proteinuria in obese patients, even in patients without severe active LN. The correlation of GlycA with BMI is well established [69], [78], [278], [301], but we emphasize that GlycA carries additional information relevant to the discrimination between proliferative and non-proliferative LN. This is demonstrated by the fact that both terms make a significant contribution to our logistic regression model and the high performance of the bivariate BMI and GlycA model.

The link between GlycA and LN severity is not altogether unexpected. We have previously shown the involvement of a set of neutrophil activation related genes with disease severity in LN [293], indicating the possible involvement of Neutrophil Extracellular Trap (NET) formation. Large scale gene correlation network analysis has shown that GlycA can also be associated to NET formation [76] and our data shows neutrophil count correlates to GlycA levels. Thus, the independent contribution of both neutrophil concentration and GlycA to the predictive logistic regression model for proliferative status is of particular interest and warrants further research into a mechanistic explanation of this phenomenon.

We point out specific limitations and strengths of this study. The limited sample sizes and unbalanced study design inherent to convenience sampling restrict the hypotheses which can

be statistically tested in this dataset. Specifically, while our observations show evidence that GlycA could be an excellent candidate biomarker for treatment follow-up, the lack of rigorous periodic follow-up sampling leading up to and following the flare event obfuscate the dynamics of the GlycA marker in this period. The strengths of this study lie in the robust phenotypical characterization of patients, the completeness of clinical data recorded, and the availability of samples drawn at the time of a biopsy-proven LN flare.

The GlycA biomarker quantifies the degree to which specific acute-phase glycoproteins are acetylated. Further research is required to elucidate whether the observed GlycA associations can be ascribed to either the concentrations of these proteins, their glycosylation or their acetylation profiles. Recent evidence in cardiovascular disease (CVD) research suggests that immunoglobulin G glycosylation traits can be both positively and negatively associated to CVD risk [307], illustrating our understanding of these processes is still incomplete and inviting further study of these topics.

Our study also provides a plausible explanation of why previous studies have reported conflicting information with regards to the association between GlycA and SLE disease severity. The population studied by Chung et al [86], which reported no association to disease activity, consisted of patients with relatively low disease activity and without nephritis, while the population of Durcan et al [87], where an association to disease activity was reported, was more heterogeneous and included SLE patients with renal involvement.

Taken together, our results indicate that serum GlycA concentration, as a summary measure for multiple inflammatory processes, could be a valuable biomarker to discriminate proliferative from non-proliferative forms of lupus nephritis in flaring patients and could be used to follow-up treatment response. We believe this biomarker should be more extensively tested in large, prospective cohorts of patients with SLE. Our data suggests that GlycA could possibly be used as a globally predictive biomarker and we advocate testing its capacity to identify patients with increased risk of mortality, morbidity and disease flares over longer periods of time.

Supplementary Material

Tables

Supplementary Table 1: Clinical and Demographic overview

	Characteristics	Healthy Controls (n = 29)	Non-Lupus Nephritis (n =21)	Quiescent SLE (n = 39)	All Active SLE (n =105)
Demographic characterictics	Age (median, IQR)		48, 35-56	35, 27.5-47	32, 26-42
	Gender female (%)		38	95	94
	Ethnicity caucasian (%)			67	82
	BMI (median, IQR)		23, 20.75-27.25	21.45, 20.06-22.77	23.42, 20.96-26.58
	Current smoking (%)		19	5	16
	Arthritis (%)		0	0	25
Clinical characteristics	SLEDAI (median, IQR)		NA	4, 2-4	8, 6-12
	Corticosteroid daily dose, mg (median, IQR mg)		0, 0-0	6, 4-8	6, 4-10
	Hydroxychloroquine (%)		0	95	84
Lab values	CRP (median, IQR mg/L)		2.2, 1.0-5.7	1, 0.775-1.25	2.5, 1-6.3
	Serum albumin (median, IQR g/L)		38, 32-41	41.17, 39.96-43.23	36.10, 33-38.12
	Serum creatinin (median, IQR µmol/L)		172, 97-303	64.55, 56.15-83.12	68.04, 54.82-88.20
	Glomerular Filtration Rate (median, IQR mL/min 1.73m²)		34.3, 18-111	87, 64-115	98, 63.25-120.25
	C3 (median, IQR g/L)		NA	0.86, 0.78-0.91	0.72, 0.54-0.96
	C4 (median, IQR g/L)		NA	0.15, 0.13-0.17	0.11, 0.06-0.19
	UPCR (median, IQR g/g)		2, 0.465-5.9	0.303, 0.080-0.555	1.225, 0.635-2.900
NMR measurements	serum albumin (median, IQR signal area)	0.086, 0.083-0.090	0.079, 0.074-0.082	0.086, 0.083-0.088	0.078, 0.070-0.081
	serum creatinin (median, IQR mmol/L)	0.051, 0.041-0.059	0.126, 0.074-0.218	0.057, 0.048-0.068	0.057, 0.047-0.074
	GlycA (median, IQR mmol/L)	1.322, 1.147-1.449	1.287, 1.151-1.506	1.184, 1.13-1.311	1.423, 1.287-1.621

Supplementary Table 2: SLE patient clinical and demographical overview

	Characteristics	Active SLE (n = 37)	Active LN (n = 25)	Biopsy proven Flaring Non-Proliferative LN (n = 11)	Biopsy proven Flaring Proliferative LN (n = 32)
Demographic characterictics	Age (median, IQR)	31.5, 26-44	32, 26-42	35, 27-38.5	32.5, 26.75-41.5
	Gender female (%)	95	100	91	91
	Ethnicity caucasian (%)	68	92	82	90
	BMI (median, IQR)	24.39, 21.45-27.21	23.46, 20.51-26.37	29, 25-29.7	23, 20.325-25.175
	Current smoking (%)	3	12	36	26
	Arthritis (%)	24	0	64	31
Clinical characteristics	SLEDAI (median, IQR)	8, 6-12	8, 8-12	8, 5-13	12, 7-15
	Corticosteroid daily dose, mg (median, IQR mg)	6, 4-8	6, 6-10	5, 0-13.75	8, 5-19
	Hydroxychloroquine (%)	84	96	18	34
Lab values	CRP (median, IQR mg/L)	1, 1-7	3, 1-8	3.5, 2.5-6.2	2.7, 1-5.7
	Serum albumin (median, IQR g/L)	38.24, 35.26-42.56	36.67, 35.91-37.55	34.17, 30-37	32, 27.85-36.00
	Serum creatinin (median, IQR µmol/L)	61.89, 53.94-78.25	68.08, 54.82-91.96	58, 53-66.5	74, 57.50-101.49
	Glomerular Filtration Rate (median, IQR mL/min 1.73m²)	100, 64-123	83, 57-116	114, 100-126.5	84, 62.5-116.5
	C3 (median, IQR g/L)	0.84, 0.58-1.02	0.67, 0.52-0.98	0.81, 0.56-1.22	0.66, 0.54-0.89
	C4 (median, IQR g/L)	0.11, 0.07-0.14	0.08, 0.05-0.18	0.16, 0.07-0.25	0.11, 0.06-0.19
	UPCR (median, IQR g/g)	0.618, 0.402-1.066	1.5, 0.677-2.042	2.1, 0.790-9.925	2.075, 0.99-5.717
NMR measurements	NMR serum albumin (median, IQR signal area)	0.081, 0.076-0.086	0.077, 0.074-0.08	0.07, 0.062-0.075	0.071, 0.067-0.08
	NMR serum creatinin (median, IQR mmol/L)	0.054, 0.046-0.071	0.059, 0.052-0.076	0.051, 0.041-0.055	0.062, 0.052-0.08
	GlycA (median, IQR mmol/L)	1.357, 1.267-1.537	1.474, 1.306-1.753	1.371, 1.289-1.488	1.540, 1.326-1.674

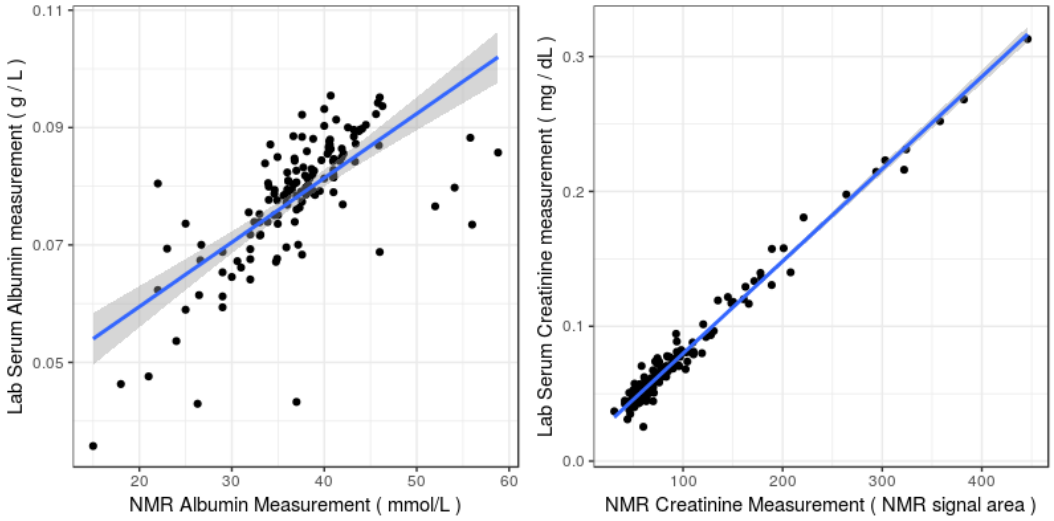
Supplementary Table 3: spearman correlation of GlycA to measured clinical variables. Significant adjusted p values ($p<0.05$ after Benjamini Hochberg multiple testing correction) are bolded.

Characteristics	All Samples N=194			All SLE Patient N=144			Quiescent SLE N=39			Active SLE N=105			Biopsy-proven Flaring LN N=43		
	Spearman ρ	p	adjusted p	Spearman ρ	p	adjusted p	Spearman ρ	p	adjusted p	Spearman ρ	p	adjusted p	Spearman ρ	p	adjusted p
Albumin	-0.38	**5.36E-06**	**1.96E-05**	-0.41	**4.90E-06**	**1.80E-05**	-0.22	2.37E-01	5.92E-01	-0.26	**1.58E-02**	6.99E-02	0.05	7.63E-01	7.63E-01
BMI	0.01	9.50E-01	9.50E-01	0.07	4.14E-01	4.45E-01	0.07	6.58E-01	7.18E-01	-0.12	2.49E-01	3.42E-01	0.07	6.73E-01	7.63E-01
C3	-0.18	**3.73E-02**	5.85E-02	-0.18	**3.73E-02**	8.20E-02	-0.06	7.18E-01	7.18E-01	-0.13	2.13E-01	3.34E-01	-0.15	3.93E-01	7.63E-01
C4	-0.10	2.53E-01	2.79E-01	-0.10	2.53E-01	3.10E-01	-0.08	6.11E-01	7.18E-01	0.04	7.01E-01	7.71E-01	0.06	7.49E-01	7.63E-01
CRP	0.49	**5.31E-10**	**5.84E-09**	0.49	**8.91E-09**	**9.80E-08**	0.57	**6.06E-04**	**6.06E-03**	0.36	**4.38E-04**	**4.82E-03**	0.06	7.32E-01	7.63E-01
CorticoDosage	0.11	1.88E-01	2.30E-01	0.07	4.45E-01	4.45E-01	0.12	5.56E-01	7.18E-01	0.02	8.51E-01	8.51E-01	-0.22	1.71E-01	4.71E-01
Creatinine	0.17	**3.68E-02**	5.85E-02	0.16	5.47E-02	9.83E-02	0.41	**1.22E-02**	6.09E-02	0.13	1.78E-01	3.26E-01	0.34	**2.93E-02**	1.61E-01
GFR	-0.13	9.12E-02	1.25E-01	-0.13	1.40E-01	1.93E-01	-0.37	**2.32E-02**	7.74E-02	-0.14	1.71E-01	3.26E-01	-0.34	**2.75E-02**	1.61E-01
Neutrophil	0.34	**1.67E-02**	**3.68E-02**	0.35	6.25E-02	9.83E-02	*NA*	*NA*	*NA*	0.35	6.25E-02	1.72E-01	0.35	6.25E-02	2.29E-01
SLEDAI	0.36	**1.20E-05**	**3.31E-05**	0.36	**1.20E-05**	**3.31E-05**	0.15	3.53E-01	7.06E-01	0.10	3.38E-01	4.13E-01	-0.09	5.65E-01	7.63E-01
UPCR	0.38	**4.84E-06**	**1.96E-05**	0.40	**4.76E-06**	**1.80E-05**	-0.12	5.68E-01	7.18E-01	0.24	**1.91E-02**	6.99E-02	0.08	6.03E-01	7.63E-01

Supplementary Table 4: c-statistics of all univariate and selected multivariate logistic regression models.

Included variables	**C-statistic**	N=36
protU	0.53	35
CRP	0.56	32
Albumin	0.57	36
C4	0.58	29
Neutrophil	0.61	29
Corticosteroid Dosage	0.62	36
eGFR	0.62	36
Creatinine	0.64	36
Arthritis	0.66	36
dsDNA	0.66	33
C3	0.67	29
GlycA	0.72	36
BMI	0.75	30
CRP + Albumin + eGFR + protU + C3	0.72	26
CRP + Albumin + eGFR + protU + C3 + C4	0.76	26
GlycA + Albumin + eGFR + protU + C3	0.76	29
GlycA + Albumin + eGFR + protU + C3 + C4	0.80	29
GlycA + eGFR	0.73	36
GlycA + dsDNA	0.83	33
GlycA + BMI	0.92	30
GlycA + Arthritis + BMI	0.92	30
GlycA + dsDNA + BMI	0.98	28
CRP + Albumin + eGFR + protU + C3 + dsDNA	0.90	25
GlycA + Albumin + eGFR + protU + C3 + dsDNA	0.90	26
CRP + Albumin + eGFR + protU + C3 + dsDNA + BMI	1.00	20
GlycA + BMI + dsDNA + Corticosteroid Dosage	1.00	26

Figures



Supplementary Figure 7: Measurements of serum creatinine and albumin levels correlate very well between NMR and standard laboratory tests (ϱ=0.74 and ϱ=0.94, respectively, with p-values<10^{-8}).

CHAPTER 7
Discussion and Future Perspectives

7.1 How transcriptomic analyses in ATL revealed the importance of inflammation and how GlycA quantification from gene expression led to discovery of novel applications for GlycA

International meta-analysis has shown IFN-α and AZT combination therapy is associated with improved ATL disease outcomes, and can be considered the standard treatment for leukemic ATL in most of the world [140], [142], [174], though this consensus is not 100% unanimous [308]. In Chapter 2, we showed that IFN-β, in *ex vivo* ATL cell cultures, has greater anti-proliferative and pro-apoptotic effects than IFN-α. Of note, the NF-κB pathway was shown to be differentially regulated by these IFN subtypes: the IFN-α treated cultures showed a modest down-regulation of the NF-κB related gene-set, in accordance with *in vitro* reports [185], while, in contrast, a significant upregulation of this pathway was evident following IFN-β treatment. Considering that persistent NF-κB activation is a hallmark of viral protein Tax's effects [123] and considering NF-κB's strong anti-apoptotic effects [38], [39], it is somewhat surprising that the stronger anti-leukemic effects of IFN-β (that is, greater evidence of anti-proliferative and pro-apoptotic gene transcription) are accompanied by an increase in 'anti-apoptotic' NF-κB activation. Doubly so, considering IFN-α combination therapy is provably effective in treating ATL and causes NF-κB downregulation [185]. While it is possible that the apoptotic effects of IFN-β simply outclass the anti-apoptotic effects of NF-κB, an alternative hypothesis explaining these observations draws similarities to the IFN paradox observed in lymphocytic choriomeningitis virus infection, where IFN has strong antiviral activity early during the early stages of acute viral infection, but knockdown of the viral protection enables a resurgence of the virus and allows the host to clear its chronic infection [309]–[311]. The upregulation of NF-κB activity caused by IFN-β could disturb the proliferative environment established by the virus sufficiently to enable the host immune system to effectively combat the cancer cells.

Further complicating matters, recent reports show that NF-κB's activities are more heterogenous than was originally thought. NF-κB is known for its anti-apoptotic effects [38], [39], and it is firmly established as a pro-inflammatory signaling pathway due to a) its ability to induce transcription of pro-inflammatory genes like IL-1 and TNFα and b) its implication in many inflammatory diseases including RA, MS and IBD [312]. However, like many cytokines, the transcription factor NF-κB seems to have anti-inflammatory properties in addition to its pro-inflammatory function: first, inhibition of NF-κB during the resolution of inflammation prolonged the inflammatory response and inhibited apoptosis in an animal model [313]. Second, mouse studies have shown IkB kinase (IKKβ) (a major component of the NF-κB pathway) has cell-type specific pro- and anti-inflammatory effects: it can have a proapoptotic role in neutrophils, which translates into an anti-inflammatory effect [314], and is an important factor in the development of anti-inflammatory M2 macrophages [315]. A full review of the complexity of NF-κB's activity can be found in [316]. That said, if we do read the NF-κB upregulation by IFN-β as a pro-inflammatory effect, the results generated in cell cultures reported in Chapter 2 suggest that inflammation is beneficial in promoting apoptosis and decreasing proliferation in ATL. The hypothesis suggesting inflammation plays a protective role in ATL also fits the results generated in Chapter 3, where network analysis linked IL-17 genes (i.e. pro-inflammatory cytokines) to genes with anti-proliferative characteristics. A robust negative correlation between IL-17C and PCNA expression, i.e. a biomarker for proliferation, is not only observed in several ATL gene expression datasets, but also in data generated in other Acute Lymphoblastic Leukemias.

This apparent protective role of IL-17 associated inflammation in ATL might seem surprising considering the deleterious effects ascribed to IL-17 in neuro-inflammatory pathologies like Multiple Sclerosis [317]. In MS, IL-17 is increased over levels observed in healthy controls [27], and IFN-β, which is a well-established first-line therapy for MS, exhibits therapeutic anti-inflammatory effects through the inhibition of IL-17 production [318], [319]. However, reports have shown that IFN-β can also exacerbate some subtypes of MS [320] and investigation into IFN-β's effects in other autoimmune and inflammatory conditions has revealed that IFN-β can indeed exert both

detrimental and beneficial influences depending on whether the considered disease is driven by Th1 (where IFN-β is protective) or Th17 cell activity (where IFN-β exacerbates disease) [236]. This is one example of contradictory evidence in MS: IFN-β is an effective anti-inflammatory treatment option which inhibits IL-17 production, but also exacerbates disease in settings with high Th17 cell activity. Our own observations (Chapter 3) have illustrated IL-17 gene expression is very low in ATL, in agreement with other reports, but also in the neuro-inflammatory disease HAM/TSP. Our group has recently demonstrated that IFN-β's more pronounced anti-proliferative effect when compared to IFN-α is evident in HAM/TSP [321], just like it is in the ATL patient cell cultures examined in Chapter 2. It is worth noting that IL-17 represents a gene-family which includes several forms. The protective role of IL-17C reported in Chapter 3 is not yet been examined in MS and HAM/TSP, where reports on IL-17 have primarily examined IL-17A and IL-17F forms [236], [322].

The results in Chapters 2 and 3 rely on the interpretation of gene signatures, and inflammation was not comprehensively characterized at the protein level in these studies. Theoretically, it is possible that the observed upregulation of NF-KB pathway related genes in ATL in response to IFN-β, which coincides with its anti-leukemic effects, is actually anti-inflammatory in nature [313]–[316], though the evidence in Chapter 3 challenges that hypothesis. Clearly, elucidating the role inflammation plays in the HTLV-1 associated pathologies is important, but studies have found no association between CRP levels and disease severity in HAM/TSP [323] and no difference in CRP levels of uninfected and asymptomatic HTLV-1 infected carriers [324]. In ATL, IL-6 and CRP levels have been shown to be elevated in patients, compared to healthy controls, and significantly higher in acute-subtype than in chronic-subtype patients [325].

In several contexts, GlycA has proven to be a more robust measure of inflammation than CRP [57], [84], [89]. To test GlycA's relevance in the HTLV-1 pathologies, NMR spectroscopy in collection of serum or plasma samples is required. However, not only are HTLV-1 sample collections relatively rare, but sample collection and biomarker measurements represent a significant cost. Meanwhile, transcriptomic data from the HTLV-1 setting is already available. If this transcriptomic data could be repurposed to

obtain an estimate of GlycA levels, then its potential relevance could be explored prior to committing to a thorough study. The predictive model described in Chapter 4 is designed specifically for that purpose. It uses GlycA as a response variable to enable the summarization of the chronic inflammatory burden caused by a broad range of inflammatory processes. Preliminary results generated with the model (data not shown) suggest that inflammatory burden in ATL is increased over that of healthy controls but no clinical ATL subtype and disease severity metadata is available for the dataset used to generate these predictions, meaning the protective inflammation hypothesis could not be tested. For HTLV-1's other major associated pathology, HAM/TSP, preliminary data generated by our predictive model suggests a correlation of GlycA with pro-viral load, which measures how prevalent the incorporation of the retrovirus is in the host's DNA and which is considered a biomarker for disease severity. The predictive model also suggests a stepwise increase from healthy controls to asymptomatic carriers and HAM/TSP cases. NMR experiments on HAM/TSP samples to test these hypotheses are planned. Considering HAM/TSP's similarities to MS, the GlycA biomarker should also be more closely examined in this better characterized neuroinflammatory pathology.

The GlycA predictive model was designed to be applicable to any transcriptomic dataset. Re-use of publicly available gene expression data allowed us to correctly identify two novel contexts where GlycA could have clinical usefulness: the inflammatory bowel disease (IBD) context, where GlycA can be used to track treatment response even if patients are CRP negative at baseline (Chapter 5), and SLE, where GlycA is associated to disease activity and can be used to differentiate between proliferative and non-proliferative cases of flaring LN (Chapter 6). While these results show great promise for the applicability of the GlycA biomarker, they do not provide a mechanistic explanation for the observed associations. In what follows, I explore a number of hypotheses and discuss analysis techniques which could help elucidate how and why high GlycA concentrations are so detrimental to human health. The discussion concludes with a brief highlight of the clinical value of the GlycA biomarker before summarizing the experiments planned in future work.

7.2 What causes Glycoprotein Acetylation?

7.2.1 Neutrophilic involvement, observations in SLE

Recent evidence from our collaborators suggests that an IFN and neutrophilic gene expression signature is associated with renal involvement and disease severity in SLE and could possibly be used to predict disease flares [293]. Considering the results from a landmark network analysis in more than 2500 transcriptomes general population blood samples revealed an association between neutrophil gene expression and GlycA [76] and our own data (Chapter 6) confirms that GlycA is associated to SLE disease activity, neutrophilic activity is strongly implicated as a contributor to serum GlycA concentration.

Observations made in a Lupus mouse model with regards to the beneficial effect of AAT protein and gene treatment (with AAT being the protein component of GlycA which best explains its associations to morbidity, according to [63]), suggest GlycA could be associated to B cells: dendritic cell (DC) activation plays a crucial role in SLE pathogenesis [326], and AAT treatment inhibited activation of DCs while also inhibiting the production of pro-inflammatory cytokines, DC-aided B cell proliferation and IgM production [327]. In this sense, one could argue that GlycA is a read-out for DC and B cell activity, though the correlation network analysis into gene expression associations with GlycA levels supports the role of neutrophils to a greater extent than that of DC and B cell gene expression [76]. Considering that the two primary gene expression profiles associated with SLE disease activity are neutrophil and IFN related [293], and a subset of DCs has been identified as major producers of IFN-α in SLE [328], all signs indicate that GlycA is a good biomarker in the SLE disease context. SLE is likely not to be the only interferonopathy in which the GlycA biomarker could be relevant, considering reports that show neutrophils are sources of IFN-α in some circumstances [329] and IFN-β in others [330] (the relevance to SLE of which is explored in [331]).

7.2.2 T cell involvement, observations in UC

Recent reports on glycosylation profiles and the expression level of branched N-glycans on intestinal T cells in UC colonic biopsies have shown a correlation to disease severity [279] and an association to the failure of standard therapy [332]. Furthermore, these biomarkers show promise as clinical targets, as recent reports show that the deficiency in branched N-glycans can be ameliorated through metabolic supplementation with N-Acetylglucosamine (GlcNAc) in mice [280] and in human *ex vivo* $CD3^+$ T cells [333]. Our own results in IBD (Chapter 5) show that disease severity is associated with increased systemic levels of acute phase proteins or their altered glycosylation profile. While these glycosylation profiles at the site of inflammation prove to be accurate biomarkers, their link to the systemic GlycA marker has not been explored. Whether this systemic evidence correlates with the altered glycosylation profile of T cell surface molecules at the site of inflammation is an important question which could help elucidate its root cause.

Recent reports show that T cell development, transformation and can be regulated through enzymatic protein glycosylation, which requires the cellular availability of specific metabolites [334]. It should be noted that these reports quantified total O-linked GlcNAcylation content of purified T cells (in contrast to GlycA measurements which quantify N-linked GlcNAcylation in complete serum or plasma). Still, the link between metabolic pathways and T cell activity [335] suggests that the NMR analysis technique used to quantify GlycA could prove useful in these research settings, as the technique simultaneously quantifies amino acid and glycolysis related metabolites without additional cost.

7.2.3 Gut microbiota and GlycA

When searching for mechanistic explanations for GlycA's effects, the potential involvement of gut microbiota should not be ignored. In a Finnish study of pregnant overweight and obese women, fecal microbiota richness (determined using 16S rRNA sequencing) was associated to decreased levels of GlycA and fiber intake levels above the recommended daily minimum [336]. This association between healthy microbial

diversity and decreased low-grade inflammation was not evident in other inflammatory markers (i.e. hsCRP), though in this study hsCRP showed no correlation to any other quantified metric, illustrating again the high variability inherent in CRP measurements.

We can find further circumstantial evidence supporting the gut microbiome link to serum GlycA levels in the scientific literature. First, the observation that bariatric surgery normalizes GlycA levels of obese patients even if they remain overweight or obese post-surgery [81], can be explained by the impact of the surgery on the gut microbiome [337], [338]. Second, meta-analysis shows that while the effects of regular exercise on traditional inflammatory markers such as CRP are not uniform [339], the observed effects on GlycA across a variety of studies show a clear-cut beneficial effect [82] and it is possible that this protective effect of exercise on GlycA levels is achieved through the microbiome, as aerobic exercise has likewise been shown to positively affect the microbiome [340], [341].

7.2.4 So what is GlycA, really?

Retaking the introduction, we know that the GlycA signal originates from the acetylation of glycan groups of 5 main proteins and that, of these five, AAT best explains the GlycA's associations with morbidity and mortality [63]. While this observation is largely based on imputed results and care should be taken in its interpretation, it is surprising that an inflammation suppressant like AAT [64] carries the bulk of the relevant information for GlycA's associations with morbidity and mortality. Parallels between GlycA and cytokines are hard to ignore: both are soluble serum biomarkers and the presence of anti-inflammatory markers can be considered as evidence of inflammation. Going forward, protein levels should be measured concurrently with GlycA levels to elucidate which contributions to the GlycA signal hold the most explanatory power: the concentrations of the proteins, their degree of glycosylation or precisely which sugar groups are attached to them? Furthermore, it remains to be seen whether increased protein glycosylation causes an inflammatory process or if the glycosylation occurs in response to inflammatory processes: i.e. we can wonder if GlycA represents a druggable target. In mouse models for arthritis [342], lupus [327] and type I diabetes [343], therapies increasing the prevalence of AAT, a

prominent contributor to the GlycA signal, have yielded promising results on both protein and gene therapy levels. Whether the other factors contributing to GlycA signal, i.e. the degree and type of glycosylation, can be targeted remains to be seen.

These seemingly straightforward questions belie the complexity the glycome represents. Just like the genome gives rise to the transcriptome, which in turn gives rise to the proteome, the glycome represents an additional layer of complexity in the regulation of biological processes through protein interaction. The glycome contains information on which proteins are glycosylated, on which positions, and with which sugars. This combinatorial complexity is the reason why the glycome is orders of magnitude more vast than even the proteome. Considering the far reaching effects of glycosylation even on single therapeutic molecules (like high mannose glycans increasing IgG clearance from human serum [344], reviewed in [345]), it is unsurprising that investigating the complete host protein glycome is staggeringly difficult. Fortunately, even limited information in this context can prove relevant to human disease. Examples include the fucosylation of β-haptoglobin as a biomarker for colon cancer [68], the altered glycosylation profiles of transferrin (which coincidentally also contributes to the GlycA signal) in Alzheimer's disease (with an increase in sialylation in blood [346], and a decrease in sialylation in cerebral spinal fluid [347], reviewed in [348]), and the heterogeneous effects of glycosylation on the biological function of immunoglobin (Ig) [349]. In this context, GlycA can be considered a summative measure of a specific set of glycosylations on a specific set of (acute-phase) proteins, which is far easier and cheaper to determine than a detailed glycosylation profile and represents an attractive avenue to fast-track glycoscience into the clinic.

7.3 Clinical Value of serum NMR analysis

While the GlycA NMR measurement is a more expensive way of quantifying inflammation (between 25 and 50 €) than CRP quantification (approx. 5 €), its greater stability and broader applicability make it an attractive option for clinical implementation. In addition, the NMR spectrum used to quantify GlycA concentration simultaneously quantifies 228 metabolites, ranging from a detailed cholesterol and (apo-)lipoprotein profile to phospholipids, fatty acids, amino acids and glycolysis

related metabolites. Several of these metabolites have already been shown to be useful biomarkers for e.g. vascular disease in diabetes [72] or risk for dementia and Alzheimer's disease [350]. Other examples include the recently reported vascular anti-aging effects of β-hydroxybutyrate [351] or the protective anti-inflammatory effects of pyruvate (through inhibition of NF-KB activation) [352]. Combined, this means that clinical implementation of serum analysis using NMR would not only give clinicians access to a superior inflammatory biomarker but also to several other lipidomic and metabolomic markers (each with their own added value), while simultaneously eliminating the need for cholesterol profiling using different tests. In this thesis I have shown that it is possible to screen publicly available transcriptomic data for settings in which the GlycA marker could be relevant (Chapter 4), and I have provided evidence of GlycA's clinical relevance in IBD and SLE disease contexts (Chapters 5 and 6, respectively).

7.4 Future Perspectives

Future research efforts focus on two different aspects: first, more work is needed to further refine the predictive models and to robustly prove their performance. This includes testing the model's ability to identify an anti-inflammatory state, which preliminary GlycA predictions, somewhat surprisingly, suggest exists in cutaneous parasitic leishmaniasis infection.

The second avenue of research is focused on proving the clinical usefulness of GlycA. As a summary measure of chronic inflammation, GlycA *a priori* seems like a biomarker worth investigating in neuroinflammatory pathologies like HAM/TSP and MS. The tentative results generated by the predictive model do seem to support this hypothesis in HAM/TSP, where predicted GlycA is associated to pro-viral load (a biomarker for HAM/TSP severity) and the model suggests the existence of a stepwise increase in GlycA concentration between healthy controls, asymptomatic infection and active HAM/TSP. For HTLV-1's other major pathology, ATL, the predictive model suggests patients with active disease have increased GlycA levels compared to healthy controls, but additional clinical metadata is required to test whether the hypothesis of protective inflammation, following the observations reported in Chapters 2 and 3, bears true.

Outside of the HTLV-1 context, plans exist to test GlycA's clinical usefulness to detect imminent disease flares in SLE. If the neutrophilic gene expression which is associated to disease activity can predict the occurrence of SLE disease flares [293] then it is possible that GlycA concentration, which is also associated with disease activity (Chapter 6) and with neutrophil gene expression [76], could also be used to predict SLE disease flares. A project has been submitted to the Research Foundation Flanders (FWO) to investigate this hypothesis in the LUMIER² study (LUpus Molecular Immuno-monitoring to Evaluate the Risk of Relapse, ClinicalTrials.gov Identifier NCT02811094). In that study, clinically quiescent adult patients with SLE that had no change in treatment in the past 3 months are followed for a period of 12 months. In the absence of flare, blood samples are drawn every 3 months. Patients presenting a flare, are sampled at the time of the flare and again 1 month post-flare. The LUMIER² study was designed to test the flare-predictive value of neutrophil- and IFN-related genes using RNA-Seq experiments. Adding metabolomic measurements to this study not only addresses the hypothesis about GlycA's ability to predict SLE disease flare, but incidentally also provides valuable data currently missing from the literature, namely the short-term stability of the metabolic markers and GlycA in patients with stable and managed disease.

Acknowledgements

We thank T. Marketta, H. Kennet and H. Aki from the THL biobank for their assistance with regards to the DILGOM data transfer. We thank NJ. Perualia, Z. Shkedy and Y. Moreau for valuable discussion on feature selection strategies. We thank M. Imbrechts and P. Matthys for supplying sJIA samples and P. Verschueren for providing RA samples for comparisons.

Scientific contributions

K.S., B.Ve., N.J-C., J.vW. designed experiments. K.S., J.vW. analyzed data. K.S., B.Ve., J.vW., N.J-C, interpreted results and contributed to the writing and correction of the manuscript. S.G., B.Ve., contributed to data acquisition. D.D., R.K., SM.M., performed experiments. L.F., A.B., B.L., S.V., N.J-C. provided patient samples. J.K., GP.T., K.K., S.O., provided crucial transcriptomic and clinical data. L.C., L.D., N.J.-C., B.L., S.V., B.Vr., J.vW., A-M.V. provided valuable expertise and interpretation of analysis results. All authors critically read the manuscripts.

Personal contributions

For what is presented in Chapter 2, I performed the statistical analysis, and wrote and corrected the manuscript. For what is presented in Chapter 3, I designed analysis strategies, processed transcriptomic data and performed modular analysis, and contributed to the writing and correction of the manuscript. For what is presented in Chapter 4, I contributed to the study's conception, designed the analysis strategies, performed the machine learning experiments, managed the sample collection, and wrote the manuscript. For what is presented in Chapter 5, I conceived of the study, contributed to its design, performed statistical analysis, and wrote and corrected the manuscript. For what is presented in Chapter 6, I conceived of the study, contributed to its design, performed statistical analysis, and wrote and corrected the manuscript.

Conflicts of interest

B.V. received lecture fees from Ferring, R-biopharm, Takeda Pharmaceuticals and Janssen; consultancy fee from Janssen. S.V. received financial support for research from MSD, Abbvie, Janssen and UCB Pharma; lecture fees from Abbott, Abbvie, Merck Sharpe & Dohme, Ferring Pharmaceuticals and UCB Pharma; consultancy fees from Pfizer, Ferring Pharmaceuticals, Shire Pharmaceuticals Group, Merck Sharpe & Dohme, and AstraZeneca Pharmaceuticals.

All other authors of the manuscripts in this thesis have declared no competing financial or non-financial conflicts of interest.

Summary

Human T-Cell Leukemia Virus Type 1, HTLV-1, is a pathogenic retrovirus infecting approximately 10 million individuals worldwide. The virus causes two distinct pathologies: Adult T-cell leukemia/lymphoma (ATL) and HTLV-1 Associated Myelopathy / Tropical Spastic Paraparesis (HAM/TSP). The common treatment of ATL currently consists of combination therapy with interferon (IFN) α and zidovudine. However, early reports showed IFN-β was also an effective treatment strategy, though IFN-α treatment became the standard based on empirical results. To explore the potential viability of IFN-β treatment in ATL, we tested the differential effects of IFN-α and -β on short term PBMC cultures of ATL patients and concluded that IFN-β has superior anti-proliferative and pro-apoptotic effects. Additional meta-analysis in four ATL gene expression datasets revealed a consistent decrease in *RORC* transcript abundance. In addition, a robust negative correlation exists between *IL17C* gene expression and proliferative gene expression in ATL and in other lymphoid leukemias. The transcriptomic experiments used in these studies also showed that inflammation could serve a protective role in ATL. As HTLV-1's other major pathology, HAM/TSP, is a neuroinflammatory disorder, we aimed to find a robust way of quantifying the inflammatory burden in transcriptomic experiments. Glycoprotein Acetylation (GlycA) is a novel biomarker for inflammation quantified in blood serum or plasma using Nuclear Magnetic Resonance (NMR) spectroscopy. This marker is a summary measure associated with a broad range of inflammatory processes and can be interpreted as a patient's chronic inflammatory burden. Using various machine learning algorithms on a large collection of paired NMR measurements and blood gene expression profiles, we constructed a predictive model which quantifies relative GlycA concentration from the gene transcript abundance in a patient's blood. This predictive model was first shown to replicate published GlycA associations. Then, novel predictions were made using publicly available third-party data, which were tested, and confirmed to be accurate, using new NMR experiments. The GlycA measurements in Inflammatory Bowel Disease (IBD) and Systemic Lupus Erythematosus (SLE) were studied in greater detail. In IBD, GlycA concentration in patient serum samples was found to be higher than what was measured in healthy controls. In patients that responded to treatment and achieved mucosal healing, GlycA fell back down to the levels observed in healthy controls. Patients that showed an endoscopic treatment response but did not achieve full mucosal healing showed a GlycA decrease but fell short of returning to the healthy control GlycA

levels. Considering our data shows that GlycA tracks disease activity even in patients without elevated C-reactive protein, our results demonstrate that GlycA holds great promise as a serological biomarker for disease activity in IBD. In SLE, our results show that GlycA levels are higher in SLE patients than those observed in healthy controls and even in nephritic controls without lupus, despite the altered renal function of the latter. We find that GlycA is associated to the SLE disease activity index and that proliferative lupus nephritis patients have higher GlycA concentrations than non-proliferative patients at time of renal biopsy. When comparing performance of GlycA to traditional biomarkers, we show that GlycA is the more informative biomarker.

Samenvatting

Human T-Cell Leukemia Virus Type 1 (HTLV-1) is een pathogeen retrovirus waar wereldwijd ongeveer 10 miljoen mensen mee geïnfecteerd zijn. Infectie met het virus kan twee verschillende ziekten veroorzaken: Adult T-cell leukemia/lymphoma (ATL) en HTLV-1 Associated Myelopathy of Tropical Spastic Paraparesis (HAM/TSP). De behandeling van ATL bestaat uit combinatietherapie met interferon (IFN) α en zidovudine. IFN-α is de internationaal aanvaarde behandeling geworden op basis van empirisch behaalde resultaten, ondanks dat gerapporteerde testen met IFN-β aantoonden dat dit ook een potentiële behandelingsoptie was. Om na te gaan of IFN-β een beter alternatief zou kunnen zijn, hebben we de effecten van IFN-α en -β behandeling getest in PBMC celculturen van ATL patiënten. De resultaten tonen aan dat IFN-β sterkere anti-proliferatieve en pro-apoptotische eigenschappen heeft dan IFN-α. Meta-analyse in vier genexpressie datasets van ATL-patiënten toonde een consistente daling van de hoeveelheid *RORC*-transcript. Daarnaast observeerden we ook een robuuste negatieve correlatie tussen *IL17C* genexpressie en proliferatieve genexpressie in ATL én in andere lymfatische leukemieën. Uit de transcriptoomanalyse waarop deze besluiten gebaseerd zijn, bleek daarnaast ook aan dat ontsteking een beschermende rol kan spelen in ATL. Gezien de andere belangrijke pathologie veroorzaakt door HTLV-1, HAM/TSP, een neuro-inflammatoire aandoening is, hebben we een manier uitgewerkt om onstekingslast te kwantificeren in transcriptomen. Glycoproteïne Acetylatie (GlycA) is een nieuwe biomerker voor inflammatie die gemeten wordt in serum of plasma met behulp van Nucleaire Magnetische Resonantie (NMR) spectroscopie. Deze merker is geassocieerd met een brede waaier inflammatoire processen en wordt geïnterpreteerd als de chronische ontstekingslast van een patiënt. Met behulp van 'machine learning' algoritmes hebben we uit een grote collectie gepaarde NMR en genexpressie experimenten een wiskundig model opgesteld dat GlycA kan afschatten op basis van de gemeten gen transcriptie in bloedstalen. Dit model slaagt erin om gekende associaties met GlycA te repliceren in publiek beschikbare data en nieuwe voorspellingen te maken die ook bevestigd werden via NMR-metingen. De GlycA metingen in chronische darmonstekingsziekten (IBD) en Systemische Lupus Erythematodes (SLE) werden nader onderzocht. In IBD tonen we aan dat GlycA concentraties van IBD-patiënten met actieve ziekte, groter zijn dan deze van gezonde personen. Bij patiënten die goed reageerden op hun behandeling en waarvan genezing van de mucosa kon vastgesteld worden, valt de GlycA concentratie terug naar het niveau dat gemeten

wordt bij gezonde personen. Als patiënten goed reageren op behandeling, maar geen volledige genezing van de slijmvliezen vertoonden dan daalt de GlycA concentratie, maar bereikt deze niet hetzelfde niveau als dat van een gezond persoon. Onze data toont aan dat GlycA in IBD een zeer beloftevolle serologische biomerker is, gezien zelfs in patiënten waar de traditionele ontstekingsmerker C-reactief proteïne niet verhoogd is GlycA toch geassocieerd is met ziekteactiviteit. In SLE tonen we aan dat GlycA concentratie hoger is patiënten dan in gezonde personen én dan in nierontstekingspatiënten zonder lupus, ondanks de verminderde nierfunctie in deze laatste groep. GlycA is hier gerelateerd met ziekteactiviteit en patiënten met proliferatieve nierontsteking hebben bij hun nier biopsie een hogere GlycA concentratie in hun serum dan patiënten wiens nierontsteking niet proliferatief is. Als we GlycA vergelijken met traditionele biomerkers voor de proliferatieve status van de nierontsteking, dan zien we dat GlycA veel informatiever is dan de merkers die momenteel gebruikt worden.

Appendices

Appendix A

Disease Context	Study	GlycA Observation	Transcriptomic Dataset	PMID	Observed Model Prediction	Replicates known observation?
Diabetes	[72]	Increase over HC	GSE46097	24563419	Diabetes patient GlycA is 0.276 SD higher than HC	Yes
CVD	[69]–[71]	Increase over HC	GSE46097	24563419	CAD patient GlycA is 0.264 SD higher than HC	Yes
Obesity	[79], [80]	Increase over HC	GSE55205	27501771	Obese subjects predicted GlycA is 0.263 SD higher than normal weight	Yes
			GSE41505	25471305	Pearson correlation between BMI and GlycA = 0.65	Yes
SLE	[86], [87]	Increase over HC	GSE50772	25861459	SLE patient GlycA is 0.489 SD higher than HC	Yes
			GSE17755	21496236	SLE patient GlycA is 1.038 SD higher than HC	Yes
		Correlation with SLEDAI	GSE50772	25861459	Pearson correlation between SLEDAI and GlycA = 0.47	Conflicting
			GSE72754	27354683	Pearson correlation between SLEDAI and GlycA = 0.28	Conflicting
RA	[83], [84]	Increase over HC	GSE17755	21496236	RA patient GlycA is 0.489 SD higher than HC	Yes
			GES15573	19710928	RA patient GlycA is 0.769 SD higher than HC	Yes
		Correlated with DAS28	GSE15258	19699293	Pearson correlation between DAS28 and GlycA = 0.20	Yes
Psoriasis	[90]	Increase over HC	GSE61281	25243786	Psoriasis patient GlycA is 0.432 SD higher than HC	Yes
			GSE61281	25243786	Arthritic psoriasis patient GlycA is 0.20 SD higher than non-arthritic cutaneous psoriasis	NA
			GSE55201	24999591	Psoriasis patient GlycA is 0.265 SD higher than HC	Yes
		Decrease after antiTNF treatment	GSE55201	24999591	Anti-IL17A monoclonal antibody lowers GlycA by 0.148 SD	NA
KD	[57]	Increase over HC	GSE68004	29813106	Acute KD patients have GlycA levels 1.07 SD higher than HC	Yes
			GSE57183	26267155	KD patients have GlycA levels 1.08 SD higher than HC	Yes
		Increase over OFI	GSE68004	29813106	KD patients have GlycA levels 0.145 SD higher than OFI patients	Yes
			GSE63881	25614765	Severe Acute KD patients have GlycA levels 0.134 SD higher than OFI patients	Yes
		Acute > Convalescent	GSE63881	25614765	Acute KD patients have GlycA levels 0.968 SD higher than convalescent patients	Yes

Appendix B

Entrez Gene ID	Gene Symbol	Gene Description	GlycA Pearson.r	GlycA Pearson.p	GlycA Spearman.r	GlycA Spearman.p	Boruta Importance
1359	CPA3	carboxypeptidase A3 (mast cell) (CPA3), mRNA.	-0.51	0.00	-0.37	0.00	43.98
3067	HDC	histidine decarboxylase (HDC), mRNA.	-0.49	0.00	-0.32	0.00	42.18
85414	SLC45A3	solute carrier family 45, member 3 (SLC45A3), mRNA.	-0.41	0.00	-0.37	0.00	39.41
2206	MS4A2	membrane-spanning 4-domains, subfamily A, member 2 (Fc fragment of IgE, high affinity I, receptor for; beta polypeptide) (MS4A2), mRNA.	-0.40	0.00	-0.30	0.00	36.16
2624	GATA2	GATA binding protein 2 (GATA2), mRNA.	-0.39	0.00	-0.33	0.00	35.22
84767	SPRYD5	SPRY domain containing 5 (SPRYD5), mRNA.	-0.37	0.00	-0.29	0.00	29.50
440712	C1ORF186	chromosome 1 open reading frame 186 (C1orf186), mRNA.	-0.38	0.00	-0.31	0.00	27.37
399940	LOC399940	PREDICTED: similar to tripartite motif-containing 51 (LOC399940), mRNA.	-0.31	0.00	-0.25	0.00	20.71
1088	CEACAM8	carcinoembryonic antigen-related cell adhesion molecule 8 (CEACAM8), mRNA.	0.25	0.00	0.28	0.00	20.55
4973	OLR1	oxidized low density lipoprotein (lectin-like) receptor 1 (OLR1), mRNA.	0.25	0.00	0.21	0.00	18.87
22915	MMRN1	multimerin 1 (MMRN1), mRNA.	0.14	0.00	0.13	0.00	18.74
729645	LOC729645	PREDICTED: similar to C-type lectin domain family 4, member g (LOC729645), mRNA.	-0.27	0.00	-0.23	0.00	17.51
4057	LTF	lactotransferrin (LTF), mRNA.	0.27	0.00	0.27	0.00	16.20
339324	ZNF260	zinc finger protein 260 (ZNF260), mRNA.	0.02	0.46	0.01	0.71	15.38
5176	SERPINF1	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1 (SERPINF1), mRNA.	-0.13	0.00	-0.18	0.00	15.17
65008	MRPL1	mitochondrial ribosomal protein L1 (MRPL1), nuclear gene encoding mitochondrial protein, mRNA.	0.05	0.03	0.01	0.58	15.08
7979	SHFM1	split hand/foot malformation (ectrodactyly) type 1 (SHFM1), mRNA.	0.29	0.00	0.21	0.00	14.86
339512	C1ORF110	chromosome 1 open reading frame 110 (C1orf110), mRNA.	0.16	0.00	0.02	0.33	14.76
389517	LOC389517	similar to Williams Beuren syndrome chromosome region 19 (LOC389517), mRNA.	0.08	0.00	0.04	0.08	14.58
1669	DEFA4	defensin, alpha 4, corticostatin (DEFA4), mRNA.	0.20	0.00	0.24	0.00	14.54
2994	GYPB	glycophorin B (MNS blood group) (GYPB), mRNA.	0.19	0.00	0.23	0.00	14.50
28998	MRPL13	mitochondrial ribosomal protein L13 (MRPL13), nuclear gene encoding mitochondrial protein, mRNA.	0.13	0.00	0.14	0.00	14.15
932	MS4A3	membrane-spanning 4-domains, subfamily A, member 3 (hematopoietic cell-specific) (MS4A3), transcript variant 3, mRNA.	-0.14	0.00	-0.09	0.00	14.08
26122	EPC2	enhancer of polycomb homolog 2 (Drosophila) (EPC2), mRNA.	0.03	0.29	-0.01	0.64	13.91
757	TMEM50B	transmembrane protein 50B (TMEM50B), mRNA.	0.15	0.00	0.10	0.00	13.84
4680	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen) (CEACAM6), mRNA.	0.23	0.00	0.26	0.00	13.81
100507547	NA	NA	0.16	0.00	0.13	0.00	13.58
780853	SNORD3C	small nucleolar RNA, C/D box 3C (SNORD3C), small nucleolar RNA.	0.18	0.00	0.13	0.00	13.48
199675	C19ORF59	chromosome 19 open reading frame 59 (C19orf59), mRNA.	0.21	0.00	0.22	0.00	13.47
9962	SLC23A2	solute carrier family 23 (nucleobase transporters), member 2 (SLC23A2), transcript variant 2, mRNA.	-0.18	0.00	-0.23	0.00	13.19
131118	DNAJC19	DnaJ (Hsp40) homolog, subfamily C, member 19 (DNAJC19), mRNA.	0.03	0.31	-0.02	0.52	13.17
669	BPGM	2,3-bisphosphoglycerate mutase (BPGM), transcript variant 2, mRNA.	0.21	0.00	0.25	0.00	13.14
5169	ENPP3	ectonucleotide pyrophosphatase/phosphodiesterase 3 (ENPP3), mRNA.	-0.21	0.00	-0.22	0.00	13.12
7503	XIST	X (inactive)-specific transcript (non-protein coding) (XIST), non-coding RNA.	-0.19	0.00	-0.21	0.00	13.11
5008	OSM	oncostatin M (OSM), mRNA.	0.08	0.00	0.13	0.00	13.05
3934	LCN2	lipocalin 2 (LCN2), mRNA.	0.27	0.00	0.29	0.00	12.94
57605	PITPNM2	phosphatidylinositol transfer protein, membrane-associated 2 (PITPNM2), mRNA.	0.02	0.40	0.01	0.64	12.89
160622	GRASP	GRP1 (general receptor for phosphoinositides 1)-associated scaffold protein (GRASP), mRNA.	0.04	0.11	0.05	0.04	12.87
9846	GAB2	GRB2-associated binding protein 2 (GAB2), transcript variant 1, mRNA.	-0.25	0.00	-0.11	0.00	12.74
654101	LOC654101	PREDICTED: similar to tripartite motif-containing 51 (LOC654101), mRNA.	-0.27	0.00	-0.25	0.00	12.67
84243	ZDHHC18	zinc finger, DHHC-type containing 18 (ZDHHC18), mRNA.	-0.21	0.00	-0.06	0.02	12.62
8079	MLF2	myeloid leukemia factor 2 (MLF2), mRNA.	-0.23	0.00	-0.17	0.00	12.61
642569	TRIM53	PREDICTED: misc_RNA (TRIM53), miscRNA.	-0.04	0.13	-0.11	0.00	12.36
54762	GRAMD1C	GRAM domain containing 1C (GRAMD1C), mRNA.	-0.13	0.00	-0.15	0.00	12.23
1195	CLK1	CDC-like kinase 1 (CLK1), mRNA.	-0.14	0.00	-0.15	0.00	12.16
4317	MMP8	matrix metallopeptidase 8 (neutrophil collagenase) (MMP8), mRNA.	0.26	0.00	0.21	0.00	12.12
51690	LSM7	LSM7 homolog, U6 small nuclear RNA associated (S. cerevisiae) (LSM7), mRNA.	0.19	0.00	0.10	0.00	11.98
1308	COL17A1	collagen, type XVII, alpha 1 (COL17A1), mRNA.	0.21	0.00	0.23	0.00	11.97
2996	GYPE	glycophorin E (GYPE), transcript variant 1, mRNA.	0.18	0.00	0.23	0.00	11.93
154664	ABCA13	ATP-binding cassette, sub-family A (ABC1), member 13 (ABCA13), mRNA.	0.19	0.00	0.15	0.00	11.92
8693	GALNT4	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 4 (GalNAc-T4) (GALNT4), mRNA.	0.11	0.00	0.19	0.00	11.85
653600	LOC653600	PREDICTED: similar to Neutrophil defensin 1 precursor (HNP-1) (HP-1) (HP1) (Defensin, alpha 1) (LOC653600), mRNA.	0.19	0.00	0.25	0.00	11.83
376267	RAB15	RAB15, member RAS onocogene family (RAB15), mRNA.	-0.10	0.00	-0.13	0.00	11.80
160365	CLECL1	C-type lectin-like 1 (CLECL1), mRNA.	0.10	0.00	0.02	0.40	11.74
5140	PDE3B	phosphodiesterase 3B, cGMP-inhibited (PDE3B), mRNA.	-0.11	0.00	-0.16	0.00	11.72

Entrez Gene ID	Gene Symbol	Gene Description	GlycA Pearson.r	GlycA Pearson.p	GlycA Spearman.r	GlycA Spearman.p	Boruta Importance
1511	CTSG	cathepsin G (CTSG), mRNA.	0.16	0.00	0.18	0.00	11.66
340146	SLC35D3	solute carrier family 35, member D3 (SLC35D3), mRNA.	0.06	0.01	0.05	0.03	11.65
55924	C1ORF183	chromosome 1 open reading frame 183 (C1orf183), transcript variant 1, mRNA.	-0.22	0.00	-0.09	0.00	11.52
55889	GOLGA6	golgi autoantigen, golgin subfamily a, 6 (GOLGA6), mRNA.	-0.13	0.00	-0.16	0.00	11.48
2235	FECH	ferrochelatase (protoporphyria) (FECH), nuclear gene encoding mitochondrial protein, transcript variant 2, mRNA.	0.22	0.00	0.25	0.00	11.47
5376	PMP22	peripheral myelin protein 22 (PMP22), transcript variant 2, mRNA.	-0.13	0.00	-0.15	0.00	11.45
2913	GRM3	glutamate receptor, metabotropic 3 (GRM3), mRNA.	0.11	0.00	0.09	0.00	11.45
65124	ANKRD57	ankyrin repeat domain 57 (ANKRD57), mRNA.	0.03	0.28	0.02	0.33	11.39
493856	CISD2	CDGSH iron sulfur domain 2 (CISD2), mRNA.	0.13	0.00	0.19	0.00	11.32
8347	HIST1H2BC	histone cluster 1, H2bc (HIST1H2BC), mRNA.	-0.21	0.00	-0.12	0.00	11.14
732393	LOC732393	PREDICTED: similar to tripartite motif protein 27 (LOC732393), mRNA.	-0.25	0.00	-0.22	0.00	11.10
286	ANK1	ankyrin 1, erythrocytic (ANK1), transcript variant 4, mRNA.	0.11	0.00	0.11	0.00	11.02
2353	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog (FOS), mRNA.	-0.22	0.00	-0.14	0.00	10.95
60592	SCOC	short coiled-coil protein (SCOC), mRNA.	0.06	0.02	0.00	0.91	10.90
85019	C18ORF45	chromosome 18 open reading frame 45 (C18orf45), mRNA.	0.06	0.02	0.04	0.11	10.83
10562	OLFM4	olfactomedin 4 (OLFM4), mRNA.	0.20	0.00	0.22	0.00	10.77
3887	KRT81	keratin 81 (KRT81), mRNA.	-0.22	0.00	-0.23	0.00	10.74
51499	TRIAP1	TP53 regulated inhibitor of apoptosis 1 (TRIAP1), mRNA.	0.14	0.00	0.11	0.00	10.74
9474	ATG5	ATG5 autophagy related 5 homolog (S. cerevisiae) (ATG5), mRNA.	-0.01	0.62	-0.01	0.67	10.74
58476	TP53INP2	tumor protein p53 inducible nuclear protein 2 (TP53INP2), mRNA.	-0.22	0.00	-0.16	0.00	10.72
2205	FCER1A	Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide (FCER1A), mRNA.	-0.23	0.00	-0.22	0.00	10.59
8842	PROM1	prominin 1 (PROM1), mRNA.	0.14	0.00	0.10	0.00	10.57
23365	ARHGEF12	Rho guanine nucleotide exchange factor (GEF) 12 (ARHGEF12), mRNA.	0.22	0.00	0.24	0.00	10.55
84557	MAP1LC3A	microtubule-associated protein 1 light chain 3 alpha (MAP1LC3A), transcript variant 2, mRNA.	-0.28	0.00	-0.16	0.00	10.51
56954	NIT2	nitrilase family, member 2 (NIT2), mRNA.	0.03	0.29	0.04	0.11	10.42
51065	RPS27L	ribosomal protein S27-like (RPS27L), mRNA.	0.27	0.00	0.17	0.00	10.41
653641	LOC653641	PREDICTED: similar to Golgin subfamily A member 6 (Golgin linked to PML) (Golgin-like protein), transcript variant 1 (LOC653641), mRNA.	0.09	0.00	0.06	0.02	10.34
55081	IFT57	intraflagellar transport 57 homolog (Chlamydomonas) (IFT57), mRNA.	0.01	0.62	-0.01	0.65	10.31
4547	MTTP	microsomal triglyceride transfer protein (MTTP), mRNA.	0.14	0.00	0.03	0.17	10.23
199713	NLRP7	NLR family, pyrin domain containing 7 (NLRP7), transcript variant 2, mRNA.	-0.16	0.00	-0.18	0.00	10.22
90871	C9ORF123	chromosome 9 open reading frame 123 (C9orf123), mRNA.	0.04	0.12	0.03	0.16	10.11
101	ADAM8	ADAM metallopeptidase domain 8 (ADAM8), mRNA.	-0.21	0.00	-0.10	0.00	10.11
10745	PHTF1	putative homeodomain transcription factor 1 (PHTF1), mRNA.	0.05	0.03	0.05	0.04	10.09
1991	ELANE	elastase, neutrophil expressed (ELANE), mRNA.	0.18	0.00	0.23	0.00	9.96
434	ASIP	agouti signaling protein, nonagouti homolog (mouse) (ASIP), mRNA.	-0.15	0.00	-0.15	0.00	9.93
522	ATP5J	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6 (ATP5J), nuclear gene encoding mitochondrial protein, transcript variant 1, mRNA.	0.22	0.00	0.18	0.00	9.89
820	CAMP	cathelicidin antimicrobial peptide (CAMP), mRNA.	0.28	0.00	0.30	0.00	9.87
5671	PSG3	pregnancy specific beta-1-glycoprotein 3 (PSG3), mRNA.	-0.13	0.00	-0.01	0.70	9.83
55714	ODZ3	odz, odd Oz/ten-m homolog 3 (Drosophila) (ODZ3), mRNA.	0.15	0.00	0.11	0.00	9.82
9521	EEF1E1	eukaryotic translation elongation factor 1 epsilon 1 (EEF1E1), mRNA.	0.10	0.00	0.01	0.58	9.76
56666	PANX2	pannexin 2 (PANX2), mRNA.	-0.23	0.00	-0.16	0.00	9.70
23318	ZCCHC11	zinc finger, CCHC domain containing 11 (ZCCHC11), transcript variant 3, mRNA.	-0.17	0.00	-0.25	0.00	9.68
54665	RSBN1	round spermatid basic protein 1 (RSBN1), mRNA.	0.11	0.00	0.07	0.00	9.67
645052	LOC645052	PREDICTED: similar to poly (ADP-ribose) polymerase family, member 8 (LOC645052), mRNA.	-0.19	0.00	-0.18	0.00	9.66
27253	PCDH17	protocadherin 17 (PCDH17), mRNA.	0.07	0.01	0.05	0.06	9.63
55704	CCDC88A	coiled-coil domain containing 88A (CCDC88A), mRNA.	-0.02	0.45	-0.06	0.02	9.59
7504	XK	X-linked Kx blood group (McLeod syndrome) (XK), mRNA.	0.18	0.00	0.23	0.00	9.58
149420	PDIK1L	PDLIM1 interacting kinase 1 like (PDIK1L), mRNA.	0.01	0.66	0.00	0.88	9.56
57484	RNF150	ring finger protein 150 (RNF150), mRNA.	0.11	0.00	0.05	0.04	9.55
23607	CD2AP	CD2-associated protein (CD2AP), mRNA.	0.02	0.38	0.00	0.96	9.54
148534	TMEM56	transmembrane protein 56 (TMEM56), mRNA.	0.17	0.00	0.15	0.00	9.46
55769	ZNF83	zinc finger protein 83 (ZNF83), mRNA.	-0.08	0.00	-0.14	0.00	9.45
767	CA8	carbonic anhydrase VIII (CA8), mRNA.	-0.20	0.00	-0.23	0.00	9.45
140032	RPS4Y2	ribosomal protein S4, Y-linked 2 (RPS4Y2), mRNA.	0.18	0.00	0.18	0.00	9.40
8654	PDE5A	phosphodiesterase 5A, cGMP-specific (PDE5A), transcript variant 1, mRNA.	0.11	0.00	0.18	0.00	9.33
6037	RNASE3	ribonuclease, RNase A family, 3 (eosinophil cationic protein) (RNASE3), mRNA.	0.18	0.00	0.19	0.00	9.31
79886	C9ORF82	chromosome 9 open reading frame 82 (C9orf82), mRNA.	0.01	0.67	-0.03	0.20	9.29
729417	LOC729417	PREDICTED: hypothetical LOC729417 (LOC729417), mRNA.	0.04	0.09	0.05	0.04	9.25
5657	PRTN3	proteinase 3 (PRTN3), mRNA.	0.08	0.00	0.07	0.00	9.24
117247	SLC16A10	solute carrier family 16, member 10 (aromatic amino acid transporter) (SLC16A10), mRNA.	-0.10	0.00	-0.15	0.00	9.11

Entrez Gene ID	Gene Symbol	Gene Description	GlycA Pearson.r	GlycA Pearson.p	GlycA Spearman.r	GlycA Spearman.p	Boruta Importance
80173	IFT74	intraflagellar transport 74 homolog (Chlamydomonas) (IFT74), transcript variant 2, mRNA.	0.04	0.11	0.03	0.16	9.10
2867	FFAR2	free fatty acid receptor 2 (FFAR2), mRNA.	-0.15	0.00	0.01	0.60	9.09
2034	EPAS1	endothelial PAS domain protein 1 (EPAS1), mRNA.	-0.16	0.00	-0.17	0.00	9.07
51668	C1ORF41	chromosome 1 open reading frame 41 (C1orf41), mRNA.	0.07	0.01	0.07	0.00	9.05
56959	C9ORF102	chromosome 9 open reading frame 102 (C9orf102), mRNA.	0.02	0.32	0.01	0.59	9.00
6164	RPL34	ribosomal protein L34 (RPL34), transcript variant 2, mRNA.	0.17	0.00	0.11	0.00	9.00
3240	HP	haptoglobin (HP), mRNA.	0.22	0.00	0.24	0.00	8.97
56729	RETN	resistin (RETN), mRNA.	0.18	0.00	0.21	0.00	8.92
200942	KLHDC8B	kelch domain containing 8B (KLHDC8B), mRNA.	-0.21	0.00	-0.15	0.00	8.91
1349	COX7B	cytochrome c oxidase subunit VIIb (COX7B), nuclear gene encoding mitochondrial protein, mRNA.	0.24	0.00	0.16	0.00	8.90
10559	SLC35A1	solute carrier family 35 (CMP-sialic acid transporter), member A1 (SLC35A1), mRNA.	0.09	0.00	0.00	0.98	8.88
51534	VTA1	Vps20-associated 1 homolog (S. cerevisiae) (VTA1), mRNA.	-0.01	0.82	-0.02	0.50	8.87
22838	RNF44	ring finger protein 44 (RNF44), mRNA.	-0.24	0.00	-0.21	0.00	8.86
192683	SCAMP5	secretory carrier membrane protein 5 (SCAMP5), mRNA.	-0.15	0.00	-0.16	0.00	8.81
10321	CRISP3	cysteine-rich secretory protein 3 (CRISP3), mRNA.	0.21	0.00	0.18	0.00	8.71
653061	LOC653061	PREDICTED: similar to golgi autoantigen, golgin subfamily a, 8A (LOC653061), mRNA.	0.19	0.00	0.18	0.00	8.63
671	BPI	bactericidal/permeability-increasing protein (BPI), mRNA.	0.20	0.00	0.19	0.00	8.63
3577	CXCR1	chemokine (C-X-C motif) receptor 1 (CXCR1), mRNA.	-0.22	0.00	-0.05	0.06	8.62
693192	MIR607	microRNA 607 (MIR607), microRNA.	0.13	0.00	0.09	0.00	8.57
246126	CYORF15A	chromosome Y open reading frame 15A (CYorf15A), mRNA.	0.14	0.00	0.13	0.00	8.55
54682	MANSC1	MANSC domain containing 1 (MANSC1), mRNA.	-0.14	0.00	-0.03	0.23	8.54
27258	LSM3	LSM3 homolog, U6 small nuclear RNA associated (S. cerevisiae) (LSM3), mRNA.	0.22	0.00	0.15	0.00	8.51
4128	MAOA	monoamine oxidase A (MAOA), nuclear gene encoding mitochondrial protein, mRNA.	0.16	0.00	0.14	0.00	8.43
100132774	LOC100132774	PREDICTED: hypothetical protein LOC100132774, transcript variant 2 (LOC100132774), mRNA.	0.07	0.01	0.05	0.06	8.40
119710	C11ORF74	chromosome 11 open reading frame 74 (C11orf74), mRNA.	0.15	0.00	0.06	0.01	8.31
4856	NOV	nephroblastoma overexpressed gene (NOV), mRNA.	-0.27	0.00	-0.22	0.00	8.26
59285	CACNG6	calcium channel, voltage-dependent, gamma subunit 6 (CACNG6), transcript variant 1, mRNA.	-0.24	0.00	-0.22	0.00	8.23
8464	SUPT3H	suppressor of Ty 3 homolog (S. cerevisiae) (SUPT3H), transcript variant 2, mRNA.	0.09	0.00	0.09	0.00	8.20
5678	PSG9	pregnancy specific beta-1-glycoprotein 9 (PSG9), mRNA.	-0.14	0.00	-0.05	0.07	8.17
55432	YOD1	YOD1 OTU deubiquinating enzyme 1 homolog (S. cerevisiae) (YOD1), mRNA.	0.22	0.00	0.23	0.00	8.16
306	ANXA3	annexin A3 (ANXA3), mRNA.	0.12	0.00	0.16	0.00	8.15
5616	PRKY	protein kinase, Y-linked (PRKY), mRNA.	0.18	0.00	0.16	0.00	8.07
892	CCNC	cyclin C (CCNC), transcript variant 2, mRNA.	0.01	0.65	0.00	0.97	8.06
162466	PHOSPHO1	phosphatase, orphan 1 (PHOSPHO1), mRNA.	-0.11	0.00	-0.02	0.51	7.99
8178	ELL	elongation factor RNA polymerase II (ELL), mRNA.	-0.17	0.00	-0.09	0.00	7.95
26577	PCOLCE2	procollagen C-endopeptidase enhancer 2 (PCOLCE2), mRNA.	0.14	0.00	0.09	0.00	7.90
8334	HIST1H2AC	histone cluster 1, H2ac (HIST1H2AC), mRNA.	-0.13	0.00	-0.07	0.00	7.89
100129034	LOC100129034	PREDICTED: hypothetical protein LOC100129034 (LOC100129034), mRNA.	-0.11	0.00	-0.03	0.18	7.89
2358	FPR2	formyl peptide receptor 2 (FPR2), transcript variant 1, mRNA.	-0.20	0.00	-0.09	0.00	7.82
439996	IFIT1L	interferon-induced protein with tetratricopeptide repeats 1-like (IFIT1L), mRNA.	0.16	0.00	0.20	0.00	7.80
58526	MID1IP1	MID1 interacting protein 1 (gastrulation specific G12 homolog (zebrafish)) (MID1IP1), mRNA.	-0.29	0.00	-0.22	0.00	7.73
23645	PPP1R15A	protein phosphatase 1, regulatory (inhibitor) subunit 15A (PPP1R15A), mRNA.	-0.19	0.00	-0.09	0.00	7.64
126308	MOBKL2A	MOB1, Mps One Binder kinase activator-like 2A (yeast) (MOBKL2A), mRNA.	-0.20	0.00	-0.08	0.00	7.61
100131546	LOC100131546	PREDICTED: misc_RNA (LOC100131546), miscRNA.	0.05	0.05	0.01	0.59	7.57
84520	C14ORF142	chromosome 14 open reading frame 142 (C14orf142), mRNA.	0.03	0.29	0.01	0.55	7.56
133	ADM	adrenomedullin (ADM), mRNA.	-0.11	0.00	0.01	0.60	7.54
9086	EIF1AY	eukaryotic translation initiation factor 1A, Y-linked (EIF1AY), mRNA.	0.18	0.00	0.20	0.00	7.51
6518	SLC2A5	solute carrier family 2 (facilitated glucose/fructose transporter), member 5 (SLC2A5), mRNA.	0.16	0.00	0.19	0.00	7.47
80127	C14ORF45	chromosome 14 open reading frame 45 (C14orf45), mRNA.	0.18	0.00	0.21	0.00	7.42
595097	SNORD16	small nucleolar RNA, C/D box 16 (SNORD16), small nucleolar RNA.	0.17	0.00	0.17	0.00	7.42
9185	REPS2	RALBP1 associated Eps domain containing 2 (REPS2), transcript variant 2, mRNA.	-0.17	0.00	-0.03	0.22	7.40
66035	SLC2A11	solute carrier family 2 (facilitated glucose transporter), member 11 (SLC2A11), transcript variant 3, mRNA.	-0.15	0.00	-0.07	0.00	7.37
79009	DDX50	DEAD (Asp-Glu-Ala-Asp) box polypeptide 50 (DDX50), mRNA.	0.04	0.13	0.01	0.61	7.34
2791	GNG11	guanine nucleotide binding protein (G protein), gamma 11 (GNG11), mRNA.	0.17	0.00	0.18	0.00	7.31
818	CAMK2G	calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma (CAMK2G), transcript variant 5, mRNA.	-0.20	0.00	-0.15	0.00	7.29
5411	PNN	pinin, desmosome associated protein (PNN), mRNA.	-0.16	0.00	-0.20	0.00	7.29
83639	TEX101	testis expressed 101 (TEX101), mRNA.	-0.21	0.00	-0.22	0.00	7.28
5724	PTAFR	platelet-activating factor receptor (PTAFR), mRNA.	-0.25	0.00	-0.14	0.00	7.28
84844	PHF5A	PHD finger protein 5A (PHF5A), mRNA.	0.26	0.00	0.19	0.00	7.27
84316	LSMD1	LSM domain containing 1 (LSMD1), mRNA.	0.18	0.00	0.16	0.00	7.26
64864	RFX7	regulatory factor X, 7 (RFX7), mRNA.	0.11	0.00	0.06	0.01	7.24

Entrez Gene ID	Gene Symbol	Gene Description	GlycA Pearson.r	GlycA Pearson.p	GlycA Spearman.r	GlycA Spearman.p	Boruta Importance
56945	MRPS22	mitochondrial ribosomal protein S22 (MRPS22), nuclear gene encoding mitochondrial protein, mRNA.	0.17	0.00	0.10	0.00	7.17
100133058	LOC100133058	PREDICTED: misc_RNA (LOC100133058), miscRNA.	0.08	0.00	0.04	0.10	7.16
55281	TMEM140	transmembrane protein 140 (TMEM140), mRNA.	-0.19	0.00	-0.07	0.01	7.14
23421	ITGB3BP	integrin beta 3 binding protein (beta3-endonexin) (ITGB3BP), mRNA.	0.06	0.01	0.00	0.92	7.12
11057	ABHD2	abhydrolase domain containing 2 (ABHD2), transcript variant 2, mRNA.	-0.04	0.13	0.04	0.07	7.11
84437	KIAA1826	KIAA1826 (KIAA1826), mRNA.	0.00	0.97	0.00	0.98	7.10
26135	SERBP1	SERPINE1 mRNA binding protein 1 (SERBP1), transcript variant 3, mRNA.	0.15	0.00	0.07	0.01	7.09
55711	FAR2	fatty acyl CoA reductase 2 (FAR2), mRNA.	-0.17	0.00	-0.07	0.00	7.01
401466	C8ORF59	chromosome 8 open reading frame 59 (C8orf59), transcript variant 3, mRNA.	0.14	0.00	0.04	0.08	6.99
26095	PTPN20B	protein tyrosine phosphatase, non-receptor type 20B (PTPN20B), transcript variant 8, mRNA.	0.16	0.00	0.17	0.00	6.98
157567	ANKRD46	ankyrin repeat domain 46 (ANKRD46), mRNA.	-0.01	0.57	-0.02	0.32	6.95
650155	LOC650155	PREDICTED: similar to hCG1812818 (LOC650155), mRNA.	-0.19	0.00	-0.16	0.00	6.95
84329	HVCN1	hydrogen voltage-gated channel 1 (HVCN1), transcript variant 2, mRNA.	-0.25	0.00	-0.23	0.00	6.94
51307	FAM53C	family with sequence similarity 53, member C (FAM53C), mRNA.	-0.13	0.00	0.00	0.85	6.93
170622	COMMD6	COMM domain containing 6 (COMMD6), transcript variant 1, mRNA.	0.20	0.00	0.11	0.00	6.93
8344	HIST1H2BE	histone cluster 1, H2be (HIST1H2BE), mRNA.	-0.20	0.00	-0.13	0.00	6.90
6119	RPA3	replication protein A3, 14kDa (RPA3), mRNA.	0.19	0.00	0.13	0.00	6.83
29992	PILRA	paired immunoglobin-like type 2 receptor alpha (PILRA), transcript variant 1, mRNA.	-0.23	0.00	-0.13	0.00	6.77
4353	MPO	myeloperoxidase (MPO), nuclear gene encoding mitochondrial protein, mRNA.	0.15	0.00	0.16	0.00	6.73
134430	WDR36	WD repeat domain 36 (WDR36), mRNA.	0.12	0.00	0.10	0.00	6.71
4724	NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase) (NDUFS4), mRNA.	0.17	0.00	0.08	0.00	6.66
285955	WBSCR19	PREDICTED: Williams Beuren syndrome chromosome region 19, transcript variant 4 (WBSCR19), mRNA.	0.12	0.00	0.13	0.00	6.66
9497	SLC4A7	solute carrier family 4, sodium bicarbonate cotransporter, member 7 (SLC4A7), mRNA.	0.04	0.11	-0.02	0.54	6.64
7411	VBP1	von Hippel-Lindau binding protein 1 (VBP1), mRNA.	0.09	0.00	0.01	0.61	6.57
100132199	LOC100132199	PREDICTED: misc_RNA (LOC100132199), miscRNA.	0.11	0.00	0.05	0.03	6.56
768211	RELL1	RELT-like 1 (RELL1), transcript variant 1, mRNA.	-0.24	0.00	-0.22	0.00	6.55
100505659	NA	NA	0.05	0.05	0.03	0.18	6.52
91272	FAM44B	family with sequence similarity 44, member B (FAM44B), mRNA.	0.02	0.42	-0.01	0.74	6.52
514	ATP5E	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit (ATP5E), nuclear gene encoding mitochondrial protein, mRNA.	0.25	0.00	0.22	0.00	6.52
22889	KIAA0907	KIAA0907 (KIAA0907), mRNA.	-0.15	0.00	-0.19	0.00	6.51
51362	CDC40	cell division cycle 40 homolog (S. cerevisiae) (CDC40), mRNA.	0.11	0.00	0.08	0.00	6.51
566	AZU1	azurocidin 1 (AZU1), mRNA.	0.12	0.00	0.15	0.00	6.48
51187	RSL24D1	ribosomal L24 domain containing 1 (RSL24D1), mRNA.	0.14	0.00	0.09	0.00	6.47
154467	C6ORF129	chromosome 6 open reading frame 129 (C6orf129), mRNA.	0.17	0.00	0.14	0.00	6.43
128338	TMEM77	transmembrane protein 77 (TMEM77), mRNA.	-0.01	0.59	-0.04	0.09	6.41
9556	C14ORF2	chromosome 14 open reading frame 2 (C14orf2), mRNA.	0.26	0.00	0.22	0.00	6.40
51115	FAM82B	PREDICTED: family with sequence similarity 82, member B (FAM82B), mRNA.	0.12	0.00	0.11	0.00	6.34
8284	JARID1D	jumonji, AT rich interactive domain 1D (JARID1D), mRNA.	0.16	0.00	0.13	0.00	6.31
100302207	MIR1974	microRNA 1974 (MIR1974), microRNA.	0.20	0.00	0.20	0.00	6.28
6227	RPS21	ribosomal protein S21 (RPS21), mRNA.	0.19	0.00	0.18	0.00	6.15
4116	MAGOH	mago-nashi homolog, proliferation-associated (Drosophila) (MAGOH), mRNA.	0.09	0.00	0.10	0.00	6.13
7337	UBE3A	ubiquitin protein ligase E3A (UBE3A), transcript variant 3, mRNA.	0.07	0.00	0.02	0.38	6.10
8648	NCOA1	nuclear receptor coactivator 1 (NCOA1), transcript variant 3, mRNA.	-0.15	0.00	-0.04	0.15	6.10
57210	SLC45A4	solute carrier family 45, member 4 (SLC45A4), mRNA.	-0.15	0.00	-0.05	0.04	6.09
9874	TLK1	tousled-like kinase 1 (TLK1), mRNA.	0.00	0.88	0.03	0.30	6.07
723790	HIST2H2AA4	histone cluster 2, H2aa4 (HIST2H2AA4), mRNA.	-0.13	0.00	0.00	0.94	6.05
8539	API5	apoptosis inhibitor 5 (API5), mRNA.	0.08	0.00	0.00	0.89	6.02
780854	SNORD3D	small nucleolar RNA, C/D box 3D (SNORD3D), small nucleolar RNA.	0.19	0.00	0.17	0.00	5.94
25843	MOBKL3	MOB1, Mps One Binder kinase activator-like 3 (yeast) (MOBKL3), transcript variant 2, mRNA.	0.11	0.00	0.08	0.00	5.90
642909	LOC642909	PREDICTED: misc_RNA (LOC642909), miscRNA.	0.11	0.00	0.07	0.00	5.83
56548	CHST7	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 7 (CHST7), mRNA.	-0.24	0.00	-0.18	0.00	5.78
54529	ASNSD1	asparagine synthetase domain containing 1 (ASNSD1), mRNA.	0.06	0.02	0.02	0.51	5.72
64801	ARV1	ARV1 homolog (S. cerevisiae) (ARV1), mRNA.	0.02	0.48	0.02	0.32	5.65
81542	TMX1	thioredoxin-related transmembrane protein 1 (TMX1), mRNA.	0.14	0.00	0.09	0.00	5.62
9807	IP6K1	inositol hexakisphosphate kinase 1 (IP6K1), transcript variant 1, mRNA.	-0.19	0.00	-0.07	0.00	5.60
4851	NOTCH1	Notch homolog 1, translocation-associated (Drosophila) (NOTCH1), mRNA.	-0.12	0.00	-0.04	0.14	5.55

Bibliography

[1] P. M. Odorizzi and E. J. Wherry, "An interferon paradox.," *Science*, vol. 340, no. 6129, pp. 155–6, Apr. 2013.

[2] R. A. Settipane and C. Schwindt, "Chapter 15: Allergic rhinitis.," *Am. J. Rhinol. Allergy*, vol. 27 Suppl 1, pp. S52-5.

[3] J. Anderson, L. Caplan, J. Yazdany, M. L. Robbins, T. Neogi, K. Michaud, K. G. Saag, J. R. O'dell, and S. Kazi, "Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice," *Arthritis Care Res. (Hoboken).*, vol. 64, no. 5, pp. 640–647, May 2012.

[4] P. M. Ridker, C. H. Hennekens, J. E. Buring, and N. Rifai, "C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women.," *N. Engl. J. Med.*, vol. 342, no. 12, pp. 836–43, Mar. 2000.

[5] D. Aran, A. Lasry, A. Zinger, M. Biton, E. Pikarsky, A. Hellman, A. J. Butte, and Y. Ben-Neriah, "Widespread parainflammation in human cancer," *Genome Biol.*, vol. 17, no. 1, p. 145, Dec. 2016.

[6] P. L. Minciullo, A. Catalano, G. Mandraffino, M. Casciaro, A. Crucitti, G. Maltese, N. Morabito, A. Lasco, S. Gangemi, and G. Basile, "Inflammaging and Anti-Inflammaging: The Role of Cytokines in Extreme Longevity.," *Arch. Immunol. Ther. Exp. (Warsz).*, vol. 64, no. 2, pp. 111–26, Apr. 2016.

[7] C. Franceschi, M. Bonafè, S. Valensin, F. Olivieri, M. De Luca, E. Ottaviani, and G. De Benedictis, "Inflamm-aging. An evolutionary perspective on immunosenescence.," *Ann. N. Y. Acad. Sci.*, vol. 908, no. 1, pp. 244–54, Jun. 2000.

[8] A. H. J. Van Den Biggelaar, A. J. M. De Craen, J. Gussekloo, T. W. J. Huizinga, B. T. Heijmans, M. Frölich, T. B. L. Kirkwood, and R. G. J. Westendorp, "Inflammation underlying cardiovascular mortality is a late consequence of evolutionary programming.," *FASEB J.*, vol. 18, no. 9, pp. 1022–4, Jun. 2004.

[9] L. Cavallone, M. Bonafè, F. Olivieri, M. Cardelli, F. Marchegiani, S. Giovagnetti, G. Di Stasio, C. Giampieri, E. Mugianesi, R. Stecconi, F. Sciacca, L. M. Grimaldi, G. De Benedictis, D. Lio, C. Caruso, and C. Franceschi, "The role of IL-1 gene cluster in longevity: a study in Italian population.," *Mech. Ageing Dev.*, vol. 124, no. 4, pp. 533–8, Apr. 2003.

[10] T. E. Sweeney, A. Shidham, H. R. Wong, and P. Khatri, "A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set," *Sci. Transl. Med.*, vol. 7, no. 287, p. 287ra71-287ra71, May 2015.

[11] C. Nathan, "Points of control in inflammation.," *Nature*, vol. 420, no. 6917, pp. 846–52, 2002.

[12] V. Brinkmann, U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann, D. S. Weiss, Y. Weinrauch, and A. Zychlinsky, "Neutrophil extracellular traps kill bacteria.," *Science*, vol. 303, no. 5663, pp. 1532–5, Mar. 2004.

[13] B. G. Yipp and P. Kubes, "NETosis: how vital is it?," *Blood*, vol. 122, no. 16, pp. 2784–94, Oct. 2013.

[14] M. J. Kaplan and M. Radic, "Neutrophil extracellular traps: double-edged swords of innate immunity.," *J. Immunol.*, vol. 189, no. 6, pp. 2689–95, Sep. 2012.

[15] F. O. Martinez and S. Gordon, "The M1 and M2 paradigm of macrophage activation: time for reassessment.," *F1000Prime Rep.*, vol. 6, no. March, p. 13, 2014.

[16] Y.-C. Liu, X.-B. Zou, Y.-F. Chai, and Y.-M. Yao, "Macrophage polarization in inflammatory diseases.," *Int. J. Biol. Sci.*, vol. 10, no. 5, pp. 520–9, 2014.

[17] L. Gattinoni, E. Lugli, Y. Ji, Z. Pos, C. M. Paulos, M. F. Quigley, J. R. Almeida, E. Gostick, Z. Yu, C. Carpenito, E. Wang, D. C. Douek, D. A. Price, C. H. June, F. M. Marincola, M. Roederer, and N. P. Restifo, "A human memory T cell subset with stem cell–like properties," *Nat. Med.*, vol. 17, no. 10, pp. 1290–1297, Sep. 2011.

[18] L. Gattinoni, D. E. Speiser, M. Lichterfeld, and C. Bonini, "T memory stem cells in health and disease.," *Nat. Med.*, vol. 23, no. 1, pp. 18–27, 2017.

[19] P. Monti, A.-K. Heninger, and E. Bonifacio, "Differentiation, expansion, and homeostasis of autoreactive T cells in type 1 diabetes mellitus.," *Curr. Diab. Rep.*, vol. 9, no. 2, pp. 113–8, Apr. 2009.

[20] C. O. Tabler, M. B. Lucera, A. A. Haqqani, D. J. McDonald, S. A. Migueles, M. Connors, and J. C. Tilton, "CD4+ memory stem cells are infected by HIV-1 in a manner regulated in part by SAMHD1 expression.," *J. Virol.*, vol. 88, no. 9, pp. 4976–86, May 2014.

[21] S. Jaafoura, M. G. de Goër de Herve, E. A. Hernandez-Vargas, H. Hendel-Chavez, M. Abdoh, M. C. Mateo, R. Krzysiek, M. Merad, R. Seng, M. Tardieu, J. F. Delfraissy, C. Goujard, and Y. Taoufik, "Progressive contraction of the latent HIV reservoir around a core of less-differentiated CD4$^+$ memory T Cells.," *Nat. Commun.*, vol. 5, p. 5407, Nov. 2014.

[22] M. J. Buzon, H. Sun, C. Li, A. Shaw, K. Seiss, Z. Ouyang, E. Martin-Gayo, J. Leng, T. J. Henrich, J. Z. Li, F. Pereyra, R. Zurakowski, B. D. Walker, E. S. Rosenberg, X. G. Yu, and M. Lichterfeld, "HIV-1 persistence in CD4+ T cells with stem cell-like properties.," *Nat. Med.*, vol. 20, no. 2, pp. 139–42, Feb. 2014.

[23] Y. Nagai, M. Kawahara, M. Hishizawa, Y. Shimazu, N. Sugino, S. Fujii, N. Kadowaki, and A. Takaori-Kondo, "T memory stem cells are the hierarchical apex of adult T-cell leukemia," *Blood*, vol. 125, no. 23, pp. 3527–3535, Jun. 2015.

[24] Y. Iwakura, H. Ishigame, S. Saijo, and S. Nakae, "Functional specialization of interleukin-17 family members.," *Immunity*, vol. 34, no. 2, pp. 149–62, Feb. 2011.

[25] H. R. Conti and S. L. Gaffen, "IL-17-Mediated Immunity to the Opportunistic Fungal Pathogen Candida albicans.," *J. Immunol.*, vol. 195, no. 3, pp. 780–8, Aug. 2015.

[26] K. Kisand, A. S. Bøe Wolff, K. T. Podkrajsek, L. Tserel, M. Link, K. V Kisand, E. Ersvaer, J. Perheentupa, M. M. Erichsen, N. Bratanic, A. Meloni, F. Cetani, R. Perniola, B. Ergun-Longmire, N. Maclaren, K. J. E. Krohn, M. Pura, B. Schalke, P. Ströbel, M. I. Leite, T. Battelino, E. S. Husebye, P. Peterson, N. Willcox, and A. Meager, "Chronic mucocutaneous candidiasis in APECED or thymoma patients correlates with autoimmunity to Th17-associated

cytokines.," *J. Exp. Med.*, vol. 207, no. 2, pp. 299–308, Feb. 2010.

[27] Y. Cao, B. A. Goods, K. Raddassi, G. T. Nepom, W. W. Kwok, J. C. Love, and D. A. Hafler, "Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple sclerosis.," *Sci. Transl. Med.*, vol. 7, no. 287, p. 287ra74, May 2015.

[28] G. Allam, A. Abdel-Moneim, and A. M. Gaber, "The pleiotropic role of interleukin-17 in atherosclerosis.," *Biomed. Pharmacother.*, vol. 106, no. April, pp. 1412–1418, Oct. 2018.

[29] C. A. Hunter and S. A. Jones, "IL-6 as a keystone cytokine in health and disease.," *Nat. Immunol.*, vol. 16, no. 5, pp. 448–57, May 2015.

[30] H. Yasukawa, M. Ohishi, H. Mori, M. Murakami, T. Chinen, D. Aki, T. Hanada, K. Takeda, S. Akira, M. Hoshijima, T. Hirano, K. R. Chien, and A. Yoshimura, "IL-6 induces an anti-inflammatory response in the absence of SOCS3 in macrophages.," *Nat. Immunol.*, vol. 4, no. 6, pp. 551–6, Jun. 2003.

[31] H. B. Richards, M. Satoh, M. Shaw, C. Libert, V. Poli, and W. H. Reeves, "Interleukin 6 dependence of anti-DNA antibody production: evidence for two pathways of autoantibody formation in pristane-induced lupus.," *J. Exp. Med.*, vol. 188, no. 5, pp. 985–90, Sep. 1998.

[32] T. Alonzi, E. Fattori, D. Lazzaro, P. Costa, L. Probert, G. Kollias, F. De Benedetti, V. Poli, and G. Ciliberto, "Interleukin 6 is required for the development of collagen-induced arthritis.," *J. Exp. Med.*, vol. 187, no. 4, pp. 461–8, Feb. 1998.

[33] C. Gealy, M. Denson, C. Humphreys, B. McSharry, G. Wilkinson, and R. Caswell, "Posttranscriptional suppression of interleukin-6 production by human cytomegalovirus.," *J. Virol.*, vol. 79, no. 1, pp. 472–85, Jan. 2005.

[34] C. A. Fielding, R. M. McLoughlin, C. S. Colmont, M. Kovaleva, D. A. Harris, S. Rose-John, N. Topley, and S. A. Jones, "Viral IL-6 blocks neutrophil infiltration during acute inflammation.," *J. Immunol.*, vol. 175, no. 6, pp. 4024–9, Sep. 2005.

[35] A. P. Hutchins, D. Diez, and D. Miranda-Saavedra, "The IL-10/STAT3-mediated anti-inflammatory response: recent developments and future challenges.," *Brief. Funct. Genomics*, vol. 12, no. 6, pp. 489–98, Nov. 2013.

[36] S. S. Iyer and G. Cheng, "Role of interleukin 10 transcriptional regulation in inflammation and autoimmune disease.," *Crit. Rev. Immunol.*, vol. 32, no. 1, pp. 23–63, 2012.

[37] C. M. Hedrich and J. H. Bream, "Cell type-specific regulation of IL-10 expression in inflammation and disease.," *Immunol. Res.*, vol. 47, no. 1–3, pp. 185–206, Jul. 2010.

[38] H. Wajant, K. Pfizenmaier, and P. Scheurich, "Tumor necrosis factor signaling.," *Cell Death Differ.*, vol. 10, no. 1, pp. 45–65, Jan. 2003.

[39] G. Chen and D. V Goeddel, "TNF-R1 signaling: a beautiful pathway.," *Science*, vol. 296, no. 5573, pp. 1634–5, May 2002.

[40] F. Ciccarelli, M. De Martinis, M. M. Sirufo, and L. Ginaldi, "Psoriasis Induced by Anti-Tumor Necrosis Factor Alpha Agents: A Comprehensive Review of the Literature.," *Acta Dermatovenerol. Croat.*, vol. 24, no. 3, pp. 169–74, Aug. 2016.

[41] E. Esposito and S. Cuzzocrea, "TNF-alpha as a therapeutic target in inflammatory diseases, ischemia-reperfusion injury and trauma.," *Curr. Med. Chem.*, vol. 16, no. 24, pp. 3152–67, 2009.

[42] S. M. Gallitano, L. McDermott, K. Brar, and E. Lowenstein, "Use of tumor necrosis factor (TNF) inhibitors in patients with HIV/AIDS.," *J. Am. Acad. Dermatol.*, vol. 74, no. 5, pp. 974–80, May 2016.

[43] A. ISAACS and J. LINDENMANN, "Virus interference. I. The interferon.," *Proc. R. Soc. London. Ser. B, Biol. Sci.*, vol. 147, no. 927, pp. 258–67, Sep. 1957.

[44] S. Pestka, C. D. Krause, and M. R. Walter, "Interferons, interferon-like cytokines, and their receptors.," *Immunol. Rev.*, vol. 202, pp. 8–32, Dec. 2004.

[45] P. Hermant and T. Michiels, "Interferon-λ in the Context of Viral Infections: Production, Response and Therapeutic Implications.," *J. Innate Immun.*, pp. 563–574, Apr. 2014.

[46] D. Levin, W. M. Schneider, H.-H. Hoffmann, G. Yarden, A. G. Busetto, O. Manor, N. Sharma, C. M. Rice, and G. Schreiber, "Multifaceted Activities of Type I Interferon Are Revealed by a Receptor Antagonist," *Sci. Signal.*, vol. 7, no. 327, pp. ra50-ra50, May 2014.

[47] A. Billiau and P. Matthys, "Interferon-gamma: a historical perspective.," *Cytokine Growth Factor Rev.*, vol. 20, no. 2, pp. 97–113, Apr. 2009.

[48] K. Schroder, P. J. Hertzog, T. Ravasi, and D. A. Hume, "Interferon-gamma: an overview of signals, mechanisms and functions.," *J. Leukoc. Biol.*, vol. 75, no. 2, pp. 163–89, Feb. 2004.

[49] X. Hu and L. B. Ivashkiv, "Cross-regulation of signaling pathways by interferon-gamma: implications for immune responses and autoimmune diseases.," *Immunity*, vol. 31, no. 4, pp. 539–50, Oct. 2009.

[50] P. R. Errante, J. B. Frazao, and A. Condino-Neto, "The Use of Interferon-Gamma Therapy in Chronic Granulomatous Disease," *Recent Patents on Anti-Infective Drug Discovery*, vol. 3. pp. 225–230, 2008.

[51] "Safety Study of Pegylated Interferon Lambda Plus Single or 2 Direct Antiviral Agents With Ribavirin (D-LITE)." [Online]. Available: http://clinicaltrials.gov/show/NCT01309932.

[52] S. Sleijfer, M. Bannink, A. R. Van Gool, W. H. J. Kruit, and G. Stoter, "Side effects of interferon-alpha therapy.," *Pharm. World Sci.*, vol. 27, no. 6, pp. 423–31, Dec. 2005.

[53] E. U. Walther and R. Hohlfeld, "Multiple sclerosis: side effects of interferon beta therapy and their management.," *Neurology*, vol. 53, no. 8, pp. 1622–7, Nov. 1999.

[54] M. B. Pepys and M. L. Baltz, "Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein.," *Adv. Immunol.*, vol. 34, pp. 141–212, 1983.

[55] M. B. Pepys and G. M. Hirschfield, "C-reactive protein: a critical update.," *J. Clin. Invest.*, vol. 111, no. 12, pp. 1805–12, Jun. 2003.

[56] S. Vermeire, G. Van Assche, and P. Rutgeerts, "Laboratory markers in IBD: useful, magic, or unnecessary toys?," *Gut*, vol. 55, no. 3, pp. 426–31, Mar. 2006.

[57] M. A. Connelly, C. Shimizu, D. A. Winegar, I. Shalaurova, R. Pourfarzib, J. D. Otvos, J. T. Kanegaye, A. H. Tremoulet, and J. C. Burns, "Differences in GlycA and lipoprotein particle parameters may help distinguish acute kawasaki disease from other febrile illnesses in children.," *BMC Pediatr.*, vol. 16, no. 1, p. 151, Dec. 2016.

[58] J. D. Otvos, I. Shalaurova, J. Wolak-Dinsmore, M. A. Connelly, R. H. Mackey, J. H. Stein, and R. P. Tracy, "GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation," *Clin. Chem.*, vol. 61, no. 5, pp. 714–723, May 2015.

[59] J. D. Bell, J. C. C. Brown, J. K. Nicholson, and P. J. Sadler, "Assignment of resonances for 'acute-phase' glycoproteins in high resolution proton NMR spectra of human blood plasma," *FEBS Lett.*, vol. 215, no. 2, pp. 311–315, 1987.

[60] J. L. Ebersole and D. Cappelli, "Acute-phase reactants in infections and inflammatory diseases.," *Periodontol. 2000*, vol. 23, no. 1, pp. 19–49, Jun. 2000.

[61] K. F. Aronsen, G. Ekelund, C. Kindmark, and C. B. Laurell, "Sequential changes of plasma proteins after surgical trauma," *Scand. J. Clin. Lab. Invest.*, vol. 29, no. S124, pp. 127–136, 1972.

[62] P. Soininen, A. J. Kangas, P. Würtz, T. Suna, and M. Ala-Korpela, "Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics," *Circ. Cardiovasc. Genet.*, vol. 8, no. 1, pp. 192–206, Feb. 2015.

[63] S. C. Ritchie, J. Kettunen, M. Brozynska, A. P. Nath, A. S. Havulinna, S. Männistö, M. Perola, V. Salomaa, M. Ala-Korpela, G. Abraham, P. Würtz, and I. Michael, "Elevated alpha-1 antitrypsin is a major component of GlycA-associated risk for future morbidity and mortality," *bioRxiv Prepr.*, 2018.

[64] N. Kalsheker, "Alpha-1-Antitrypsin Gene Regulation," in *Alpha-1-antitrypsin Deficiency*, Elsevier, 2017, pp. 25–35.

[65] H. Koami, Y. Sakamoto, T. Miyasho, R. Noguchi, N. Sato, K. Kai, K. Chris Yamada, and S. Inoue, "Haptoglobin Reduces Inflammatory Cytokine INF-γ and Facilitates Clot Formation in Acute Severe Burn Rat Model.," *J. Nippon Med. Sch.*, vol. 84, no. 2, pp. 64–72, 2017.

[66] S. Thompson, E. Dargan, I. D. Griffiths, C. A. Kelly, and G. A. Turner, "The glycosylation of haptoglobin in rheumatoid arthritis.," *Clin. Chim. Acta.*, vol. 220, no. 1, pp. 107–14, Oct. 1993.

[67] R. Saldova, M. R. Wormald, R. A. Dwek, and P. M. Rudd, "Glycosylation changes on serum glycoproteins in ovarian cancer may contribute to disease pathogenesis.," *Dis. Markers*, vol. 25, no. 4–5, pp. 219–32, 2008.

[68] S.-Y. Park, S.-H. Lee, N. Kawasaki, S. Itoh, K. Kang, S. Hee Ryu, N. Hashii, J.-M. Kim, J.-Y. Kim, and J. Hoe Kim, "α1-3/4 fucosylation at Asn 241 of β-haptoglobin is a novel marker for colon cancer: a combinatorial approach for development of glycan biomarkers.," *Int. J. cancer*, vol. 130, no. 10, pp. 2366–76, May 2012.

[69] A. O. Akinkuolie, J. E. Buring, P. M. Ridker, and S. Mora, "A Novel Protein Glycan Biomarker and Future Cardiovascular Disease Events," *J. Am. Heart Assoc.*, vol. 3, no. 5, pp. e001221–e001221, Sep. 2014.

[70] K. Fischer, J. Kettunen, P. Würtz, T. Haller, A. S. Havulinna, A. J. Kangas, P. Soininen, T. Esko, M.-L. Tammesoo, R. Mägi, S. Smit, A. Palotie, S. Ripatti, V. Salomaa, M. Ala-Korpela, M. Perola, and A. Metspalu, "Biomarker Profiling by Nuclear Magnetic Resonance Spectroscopy for the Prediction of All-Cause Mortality: An Observational Study of 17,345 Persons," *PLoS Med.*, vol. 11, no. 2, p. e1001606, Feb. 2014.

[71] D. A. Duprez, J. Otvos, O. A. Sanchez, R. H. Mackey, R. Tracy, and D. R. Jacobs, "Comparison of the Predictive Value of GlycA and Other Biomarkers of Inflammation for Total Death, Incident Cardiovascular Events, Noncardiovascular and Noncancer Inflammatory-Related Events, and Total Cancer Events," *Clin. Chem.*, vol. 62, no. 7, pp. 1020–1031, Jul. 2016.

[72] A. O. Akinkuolie, A. D. Pradhan, J. E. Buring, P. M. Ridker, and S. Mora, "Novel Protein Glycan Side-Chain Biomarker and Risk of Incident Type 2 Diabetes Mellitus," *Arterioscler. Thromb. Vasc. Biol.*, vol. 35, no. 6, pp. 1544–1550, Jun. 2015.

[73] J. E. Kaikkonen, P. Würtz, E. Suomela, M. Lehtovirta, A. J. Kangas, A. Jula, V. Mikkilä, J. S. A. Viikari, M. Juonala, T. Rönnemaa, N. Hutri-Kähönen, M. Kähönen, T. Lehtimäki, P. Soininen, M. Ala-Korpela, and O. T. Raitakari, "Metabolic profiling of fatty liver in young and middle-aged adults: Cross-sectional and prospective analyses of the Young Finns Study.," *Hepatology*, vol. 65, no. 2, pp. 491–500, 2017.

[74] J. B. Muhlestein, H. May, D. Winegar, J. Rollo, M. Connelly, J. Otvos, and J. Anderson, "Differential Association of High-Density Lipoprotein Particle Subclasses and Glyca, a Novel Inflammatory Marker, in Predicting Cardiac Death Among Patients Undergoing Angiography: the Intermountain Heart Collaborative Study," *J. Am. Coll. Cardiol.*, vol. 67, no. 13, p. 162, 2016.

[75] B. Doran, C. Haynes, R. McGarrah, W. E. Kraus, D. Craig, and S. H. Shah, "GlycA, a Novel Inflammatory Marker, Is Strongly Associated With Mortality in Both Women and Men," *Circulation*, vol. 134, p. A17882, 2016.

[76] S. C. Ritchie, P. Würtz, A. P. Nath, G. Abraham, A. S. Havulinna, L. G. Fearnley, A.-P. Sarin, A. J. Kangas, P. Soininen, K. Aalto, I. Seppälä, E. Raitoharju, M. Salmi, M. Maksimow, S. Männistö, M. Kähönen, M. Juonala, S. Ripatti, T. Lehtimäki, S. Jalkanen, M. Perola, O. Raitakari, V. Salomaa, M. Ala-Korpela, J. Kettunen, and M. Inouye, "The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection," *Cell Syst.*, vol. 1, no. 4, pp. 293–301, Oct. 2015.

[77] R. W. McGarrah, J. P. Kelly, D. M. Craig, C. Haynes, R. C. Jessee, K. M. Huffman, W. E. Kraus, and S. H. Shah, "A Novel Protein Glycan-Derived Inflammation Biomarker Independently Predicts Cardiovascular Disease and Modifies the Association of HDL Subclasses with Mortality.," *Clin. Chem.*, vol. 63, no. 1, pp. 288–296, Jan. 2017.

[78] M. Fizelova, R. Jauhiainen, A. J. Kangas, P. Soininen, M. Ala-korpela, J. Kuusisto, M. Laakso, and A. Stan, "Differential Associations of Inflammatory Markers With Insulin Sensitivity and Secretion: The Prospective METSIM Study," *J. Clin. Endocrinol. Metab.*, vol. 102, no. September, pp. 3600–3609, 2017.

[79] R. P. F. Dullaart, E. G. Gruppen, M. A. Connelly, and J. D. Lefrandt, "A pro-inflammatory glycoprotein biomarker is associated with lower bilirubin in metabolic syndrome," *Clin. Biochem.*, vol. 48, no. 16–17, pp. 1045–1047, Nov. 2015.

[80] R. P. F. Dullaart, E. G. Gruppen, M. A. Connelly, J. D. Otvos, and J. D. Lefrandt, "GlycA, a biomarker of inflammatory glycoproteins, is more closely related to the leptin/adiponectin ratio than to glucose tolerance status," *Clin. Biochem.*, vol. 48, no. 12, pp. 811–814, Aug. 2015.

[81] A. Manmadhan, B.-X. Lin, J. Zhong, M. Parikh, J. S. Berger, E. A. Fisher, and S. P. Heffron, "Elevated GlycA in severe obesity is normalized by bariatric surgery.," *Diabetes. Obes. Metab.*, vol. 21, no. 1, pp. 178–182, Jan. 2019.

[82] J. L. Barber, W. E. Kraus, T. S. Church, J. M. Hagberg, P. D. Thompson, D. B. Bartlett, M. W. Beets, C. P. Earnest, K. M. Huffman, R. Q. Landers-Ramos, A. S. Leon, D. C. Rao, R. L. Seip, J. S. Skinner, C. A. Slentz, K. R. Wilund, C. Bouchard, and M. A. Sarzynski, "Effects of regular endurance exercise on GlycA: Combined analysis of 14 exercise interventions," *Atherosclerosis*, vol. 277, pp. 1–6, Oct. 2018.

[83] D. B. Bartlett, M. A. Connelly, H. AbouAssi, L. A. Bateman, K. N. Tune, J. L. Huebner, V. B. Kraus, D. A. Winegar, J. D. Otvos, W. E. Kraus, and K. M. Huffman, "A novel inflammatory biomarker, GlycA, associates with disease activity in rheumatoid arthritis and cardio-metabolic risk in BMI-matched controls.," *Arthritis Res. Ther.*, vol. 18, no. 1, p. 86, Apr. 2016.

[84] M. J. Ormseth, C. P. Chung, A. M. Oeser, M. A. Connelly, T. Sokka, P. Raggi, J. F. Solus, J. D. Otvos, and C. M. Stein, "Utility of a novel inflammatory marker, GlycA, for assessment of rheumatoid arthritis disease activity and coronary atherosclerosis," *Arthritis Res. Ther.*, vol. 17, no. 1, p. 117, Dec. 2015.

[85] R. Fuertes-Martín, D. Taverner, J.-C. Vallvé, S. Paredes, L. Masana, X. Correig Blanchar, and N. Amigó Grau, "Characterization of 1H NMR Plasma Glycoproteins as a New Strategy To Identify Inflammatory Patterns in Rheumatoid Arthritis.," *J. Proteome Res.*, vol. 17, no. 11, pp. 3730–3739, Nov. 2018.

[86] C. P. Chung, M. J. Ormseth, M. A. Connelly, A. Oeser, J. F. Solus, J. D. Otvos, P. Raggi, and C. M. Stein, "GlycA, a novel marker of inflammation, is elevated in systemic lupus erythematosus," *Lupus*, vol. 25, no. 3, pp. 296–300, Mar. 2016.

[87] L. Durcan, D. A. Winegar, M. A. Connelly, J. D. Otvos, L. S. Magder, and M. Petri, "Longitudinal Evaluation of Lipoprotein Variables in Systemic Lupus Erythematosus Reveals Adverse Changes with Disease Activity and Prednisone and More Favorable Profiles with Hydroxychloroquine Therapy," *J. Rheumatol.*, vol. 43, no. 4, pp. 745–750, Apr. 2016.

[88] S. M. Titan, R. Pecoits-Filho, S. M. Barreto, A. A. Lopes, I. J. Bensenor, and P. A. Lotufo, "GlycA, a marker of protein glycosylation, is related to albuminuria and estimated glomerular filtration rate: the ELSA-Brasil study," *BMC Nephrol.*, vol. 18, no. 1, p. 367, Dec. 2017.

[89] E. G. Gruppen, I. J. Riphagen, M. A. Connelly, J. D. Otvos, S. J. L. Bakker, and R. P. F. Dullaart, "GlycA, a Pro-Inflammatory Glycoprotein Biomarker, and Incident Cardiovascular Disease: Relationship with C-Reactive Protein and Renal Function.," *PLoS One*, vol. 10, no. 9, p. e0139057, Sep. 2015.

[90] A. A. Joshi, J. B. Lerman, T. M. Aberra, M. Afshar, H. L. Teague, J. A. Rodante, P. Krishnamoorthy, Q. Ng, T. Z. Aridi, T. Salahuddin, B. Natarajan, B. N. Lockshin, M. A. Ahlman, M. Y. Chen, D. J. Rader, M. P. Reilly, A. T. Remaley, D. A. Bluemke, M. P. Playford, J. M. Gelfand, and N. N. Mehta, "GlycA Is a Novel Biomarker of Inflammation and Subclinical Cardiovascular Disease in Psoriasis," *Circ. Res.*, vol. 119, no. 11, pp. 1242–1253, Nov. 2016.

[91] T. Kawasaki, F. Kosaki, S. Okawa, I. Shigematsu, and H. Yanagawa, "A new infantile acute febrile mucocutaneous lymph node syndrome (MLNS) prevailing in Japan.," *Pediatrics*, vol. 54, no. 3, pp. 271–6, Sep. 1974.

[92] J. C. Burns and M. P. Glodé, "Kawasaki syndrome," *Lancet*, vol. 364, no. 9433, pp. 533–544, Aug. 2004.

[93] J. K. Weisman, D. Meeks, L. Mendelsohn, A. T. Remaley, M. Sampson, D. T. Allen, J. Nichols, A. S. Shet, and S. L. Thein, "GlycA is not a useful biomarker of inflammation in sickle cell disease.," *Int. J. Lab. Hematol.*, no. June, pp. 1–6, Aug. 2018.

[94] M. Tibuakuu, O. E. Fashanu, D. Zhao, J. D. Otvos, T. T. Brown, S. A. Haberlen, E. Guallar, M. J. Budoff, F. J. Palella, J. J. Martinson, A. O. Akinkuolie, S. Mora, W. S. Post, and E. D. Michos, "GlycA, a novel inflammatory marker, is associated with subclinical coronary disease in the multicenter AIDS cohort study.," *AIDS*, p. 1, Nov. 2018.

[95] B. J. Poiesz, F. W. Ruscetti, A. F. Gazdar, P. A. Bunn, J. D. Minna, and R. C. Gallo, "Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma," *Proc. Natl. Acad. Sci.*, vol. 77, no. 12, pp. 7415–7419, 1980.

[96] V. S. Kalyanaraman, M. G. Sarngadharan, M. Robert-Guroff, I. Miyoshi, D. Golde, and R. C. Gallo, "A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia.," *Science*, vol. 218, no. 4572, pp. 571–3, Nov. 1982.

[97] R. C. Gallo, L. Willems, and Y. Tagaya, "Time to Go Back to the Original Name.," *Front. Microbiol.*, vol. 8, p. 1800, 2017.

[98] A. Gessain and O. Cassar, "Epidemiological Aspects and World Distribution of HTLV-1 Infection.," *Front. Microbiol.*, vol. 3, no. November, p. 388, Jan. 2012.

[99] C. Hlela, S. Shepperd, N. P. Khumalo, and G. P. Taylor, "The prevalence of human T-cell lymphotropic virus type 1 in the general population is unknown.," *AIDS Rev.*, vol. 11, no. 4, pp. 205–14.

[100] R. J. Biggar, C. Saxinger, C. Gardiner, W. E. Collins, P. H. Levine, J. W. Clark, F. K. Nkrumah, and W. A. Blattner, "Type-I HTLV antibody in urban and rural Ghana, West Africa.," *Int. J. cancer*, vol. 34, no. 2, pp. 215–9, Aug. 1984.

[101] H. Rafatpanah, M. R. Hedayati-Moghaddam, F. Fathimoghadam, H. R. Bidkhori, S. K. Shamsian, S. Ahmadi, L. Sohgandi, M. R. Azarpazhooh, S. A. Rezaee, R. Farid, and A. Bazarbachi, "High prevalence of HTLV-I infection in Mashhad, Northeast Iran: a population-based seroepidemiology survey.," *J. Clin. Virol.*, vol. 52, no. 3, pp. 172–6, Nov. 2011.

[102] L. Einsiedel, R. J. Woodman, M. Flynn, K. Wilson, O. Cassar, and A. Gessain, "Human T-Lymphotropic Virus type 1 infection in an Indigenous Australian population: epidemiological insights from a hospital-based cohort study.," *BMC Public Health*, vol. 16, p. 787, 2016.

[103] V. Zaninovic, F. Sanzon, F. Lopez, G. Velandia, A. Blank, M. Blank, C. Fujiyama, S. Yashiki, D. Matsumoto, and Y. Katahira, "Geographic independence of HTLV-I and HTLV-II foci in the Andes highland, the Atlantic coast, and the Orinoco of Colombia.," *AIDS Res. Hum. Retroviruses*, vol. 10, no. 1, pp. 97–101, Jan. 1994.

[104] I. Dourado, L. C. J. Alcantara, M. L. Barreto, M. da Gloria Teixeira, and B. Galvão-Castro, "HTLV-I in the general population of Salvador, Brazil: a city with African ethnic and sociodemographic characteristics.," *J. Acquir. Immune Defic. Syndr.*, vol. 34, no. 5, pp. 527–31, Dec. 2003.

[105] C. R. M. Bangham, A. Araujo, Y. Yamano, and G. P. Taylor, "HTLV-1-associated myelopathy/tropical spastic paraparesis," *Nat. Rev. Dis. Prim.*, vol. 1, no. August, p. 15012, Jun. 2015.

[106] A. Paiva, J. Smid, M. E. J. Haziot, T. Assone, S. Pinheiro, L. A. M. Fonseca, A. C. P. de Oliveira, and J. Casseb, "High risk of heterosexual transmission of human T-cell lymphotropic virus type 1 infection in Brazil.," *J. Med. Virol.*, vol. 89, no. 7, pp. 1287–1294, 2017.

[107] A. M. Paiva, T. Assone, M. E. J. Haziot, J. Smid, L. A. M. Fonseca, O. do C. Luiz, A. C. P. de Oliveira, and J. Casseb, "Risk factors associated with HTLV-1 vertical transmission in Brazil: longer breastfeeding, higher maternal proviral load and previous HTLV-1-infected offspring.," *Sci. Rep.*, vol. 8, no. 1, p. 7742, May 2018.

[108] S. Hino, K. Yamaguchi, S. Katamine, H. Sugiyama, T. Amagasaki, K. Kinoshita, Y. Yoshida, H. Doi, Y. Tsuji, and T. Miyamoto, "Mother-to-child transmission of human T-cell leukemia virus type-I.," *Jpn. J. Cancer Res.*, vol. 76, no. 6, pp. 474–80, Jun. 1985.

[109] F. A. Proietti, A. B. F. Carneiro-Proietti, B. C. Catalan-Soares, and E. L. Murphy, "Global epidemiology of HTLV-I infection and associated diseases.," *Oncogene*, vol. 24, no. 39, pp. 6058–68, Sep. 2005.

[110] G. Marano, S. Vaglio, S. Pupella, G. Facco, L. Catalano, V. Piccinini, G. M. Liumbruno, and G. Grazzini, "Human T-lymphotropic virus and transfusion safety: does one size fit all?," *Transfusion*, vol. 56, no. 1, pp. 249–60, Jan. 2016.

[111] D. Moreno-Ajona, J. R. Yuste, P. Martín, and J. Gállego Pérez-Larraya, "HTLV-1 myelopathy after renal transplant and antiviral prophylaxis: the need for screening.," *J. Neurovirol.*, vol. 24, no. 4, pp. 523–525, Aug. 2018.

[112] A. Manns, R. J. Wilks, E. L. Murphy, G. Haynes, J. P. Figueroa, M. Barnett, B. Hanchard, and W. A. Blattner, "A prospective study of transmission by transfusion of HTLV-I and risk factors associated with seroconversion.," *Int. J. cancer*, vol. 51, no. 6, pp. 886–91, Jul. 1992.

[113] O. Cassar, L. Einsiedel, P. V Afonso, and A. Gessain, "Human T-cell lymphotropic virus type 1 subtype C molecular variants among indigenous australians: new insights into the molecular epidemiology of HTLV-1 in Australo-Melanesia.," *PLoS Negl. Trop. Dis.*, vol. 7, no. 9, p. e2418, 2013.

[114] M. Salemi, S. Van Dooren, E. Audenaert, E. Delaporte, P. Goubau, J. Desmyter, and A. M. Vandamme, "Two new human T-lymphotropic virus type I phylogenetic subtypes in seroindeterminates, a Mbuti pygmy and a Gabonese, have closest relatives among African STLV-I strains.," *Virology*, vol. 246, no. 2, pp. 277–87, Jul. 1998.

[115] S. Van Dooren, M. Salemi, and A. M. Vandamme, "Dating the origin of the African human T-cell lymphotropic virus type-i (HTLV-I) subtypes.," *Mol. Biol. Evol.*, vol. 18, no. 4, pp. 661–71, Apr. 2001.

[116] A. M. Vandamme, H. F. Liu, P. Goubau, and J. Desmyter, "Primate T-lymphotropic virus type I LTR sequence variation and its phylogenetic analysis: compatibility with an African origin of PTLV-I.," *Virology*, vol. 202, no. 1, pp. 212–23, Jul. 1994.

[117] S. Van Dooren, "The Low Evolutionary Rate of Human T-Cell Lymphotropic Virus Type-1 Confirmed by Analysis of Vertical Transmission Chains," *Mol. Biol. Evol.*, vol. 21, no. 3, pp. 603–611, Dec. 2003.

[118] K. Verdonck, E. González, S. Van Dooren, A. Vandamme, G. Vanham, and E. Gotuzzo, "Human T-lymphotropic virus 1 : recent knowledge about an ancient infection," *Lancet Infect. Dis.*, vol. 7, no. April, pp. 266–281, 2007.

[119] M. Matsuoka and J. Yasunaga, "Human T-cell leukemia virus type 1: replication, proliferation and propagation by Tax

and HTLV-1 bZIP factor.," *Curr. Opin. Virol.*, vol. 3, no. 6, pp. 684–91, Dec. 2013.

[120] M. Boxus and L. Willems, "Mechanisms of HTLV-1 persistence and transformation.," *Br. J. Cancer*, vol. 101, no. 9, pp. 1497–501, Nov. 2009.

[121] M. Yoshida, "Multiple viral strategies of HTLV-1 for dysregulation of cell growth control," *Annu. Rev. Immunol.*, vol. 19, no. 1, pp. 475–496, Apr. 2001.

[122] M. Matsuoka and K.-T. Jeang, "Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: viral infectivity, Tax, HBZ and therapy.," *Oncogene*, vol. 30, no. 12, pp. 1379–89, Mar. 2011.

[123] S.-C. Sun and S. Yamaoka, "Activation of NF-kappaB by HTLV-I and implications for cell transformation.," *Oncogene*, vol. 24, no. 39, pp. 5952–64, Sep. 2005.

[124] Y. Satou, J. Yasunaga, M. Yoshida, and M. Matsuoka, "HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 3, pp. 720–5, Jan. 2006.

[125] C. Journo and R. Mahieux, "HTLV-1 and innate immunity.," *Viruses*, vol. 3, no. 8, pp. 1374–94, 2011.

[126] R. Furuta, J.-I. Yasunaga, M. Miura, K. Sugata, A. Saito, H. Akari, T. Ueno, N. Takenouchi, J.-I. Fujisawa, K.-R. Koh, Y. Higuchi, M. Mahgoub, M. Shimizu, F. Matsuda, A. Melamed, C. R. Bangham, and M. Matsuoka, "Human T-cell leukemia virus type 1 infects multiple lineage hematopoietic cells in vivo.," *PLoS Pathog.*, vol. 13, no. 11, p. e1006722, Nov. 2017.

[127] A. Q. C. Araujo and M. T. T. Silva, "The HTLV-1 neurological complex.," *Lancet. Neurol.*, vol. 5, no. 12, pp. 1068–76, Dec. 2006.

[128] F. Martin, G. P. Taylor, and S. Jacobson, "Inflammatory manifestations of HTLV-1 and their therapeutic options.," *Expert Rev. Clin. Immunol.*, vol. 10, no. 11, pp. 1531–46, Nov. 2014.

[129] C. R. Bangham, "The immune response to HTLV-I.," *Curr. Opin. Immunol.*, vol. 12, no. 4, pp. 397–402, Aug. 2000.

[130] R. Pessôa, J. T. Watanabe, Y. Nukui, J. Pereira, J. Casseb, J. Kasseb, A. C. P. de Oliveira, A. C. Segurado, and S. S. Sanabani, "Molecular characterization of human T-cell lymphotropic virus type 1 full and partial genomes by Illumina massively parallel sequencing technology.," *PLoS One*, vol. 9, no. 3, p. e93374, 2014.

[131] G. P. Taylor, "Human T-lymphotropic virus type 1 infection and solid organ transplantation.," *Rev. Med. Virol.*, vol. 28, no. 2, Mar. 2018.

[132] L. B. M. Cook, A. Melamed, M. A. Demontis, D. J. Laydon, J. M. Fox, J. H. C. Tosswill, D. de Freitas, A. D. Price, J. F. Medcalf, F. Martin, J. M. Neuberger, C. R. M. Bangham, and G. P. Taylor, "Rapid dissemination of human T-lymphotropic virus type 1 during primary infection in transplant recipients.," *Retrovirology*, vol. 13, p. 3, Jan. 2016.

[133] M. Osame, R. Janssen, H. Kubota, H. Nishitani, A. Igata, S. Nagataki, M. Mori, I. Goto, H. Shimabukuro, and R. Khabbaz, "Nationwide survey of HTLV-I-associated myelopathy in Japan: association with blood transfusion.," *Ann. Neurol.*, vol. 28, no. 1, pp. 50–6, Jul. 1990.

[134] M. Kannagi, T. Ohashi, S. Hanabuchi, H. Kato, Y. Koya, A. Hasegawa, T. Masuda, and T. Yoshiki, "Immunological aspects of rat models of HTLV type 1-infected T lymphoproliferative disease.," *AIDS Res. Hum. Retroviruses*, vol. 16, no. 16, pp. 1737–40, Nov. 2000.

[135] K. Barmak, E. Harhaj, C. Grant, T. Alefantis, and B. Wigdahl, "Human T cell leukemia virus type I-induced disease: pathways to cancer and neurodegeneration.," *Virology*, vol. 308, no. 1, pp. 1–12, Mar. 2003.

[136] M. Shimoyama, "Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the Lymphoma Study Group (1984-87).," *Br. J. Haematol.*, vol. 79, no. 3, pp. 428–37, Nov. 1991.

[137] K. Tsukasaki, O. Hermine, A. Bazarbachi, L. Ratner, J. C. Ramos, W. Harrington, D. O'Mahony, J. E. Janik, A. L. Bittencourt, G. P. Taylor, K. Yamaguchi, A. Utsunomiya, K. Tobinai, and T. Watanabe, "Definition, Prognostic Factors, Treatment, and Response Criteria of Adult T-Cell Leukemia-Lymphoma: A Proposal From an International Consensus Meeting," *J. Clin. Oncol.*, vol. 27, no. 3, pp. 453–459, Dec. 2008.

[138] K. Manivannan, A. G. Rowan, Y. Tanaka, G. P. Taylor, and C. R. M. Bangham, "CADM1/TSLC1 Identifies HTLV-1-Infected Cells and Determines Their Susceptibility to CTL-Mediated Lysis," *PLOS Pathog.*, vol. 12, no. 4, p. e1005560, Apr. 2016.

[139] S. Nakahata and K. Morishita, "CADM1/TSLC1 is a Novel Cell Surface Marker for Adult T-Cell Leukemia/Lymphoma," *J. Clin. Exp. Hematop.*, vol. 52, no. 1, pp. 17–22, 2012.

[140] H. Katsuya and K. Ishitsuka, "Treatment advances and prognosis for patients with adult T-cell leukemia-lymphoma.," *J. Clin. Exp. Hematop.*, vol. 57, no. 3, pp. 87–97, Dec. 2017.

[141] A. Bazarbachi, Y. Plumelle, J. Carlos Ramos, P. Tortevoye, Z. Otrock, G. Taylor, A. Gessain, W. Harrington, G. Panelatti, and O. Hermine, "Meta-Analysis on the Use of Zidovudine and Interferon-Alfa in Adult T-Cell Leukemia/Lymphoma Showing Improved Survival in the Leukemic Subtypes," *J. Clin. Oncol.*, vol. 28, no. 27, pp. 4177–4183, Sep. 2010.

[142] R. Nasr, A. Marçais, O. Hermine, and A. Bazarbachi, "Overview of Targeted Therapies for Adult T-Cell Leukemia/Lymphoma," in *Human T-Lymphotropic Viruses: Methods and Protocols*, C. Casoli, Ed. New York, NY: Springer New York, 2017, pp. 197–216.

[143] K. Ishitsuka and K. Tamura, "Human T-cell leukaemia virus type I and adult T-cell leukaemia-lymphoma," *Lancet Oncol.*, vol. 15, no. 11, pp. e517–e526, Oct. 2014.

[144] M. Osame, K. Usuku, S. Izumo, N. Ijichi, H. Amitani, A. Igata, M. Matsumoto, and M. Tara, "HTLV-I associated

myelopathy, a new clinical entity.," *Lancet (London, England)*, vol. 1, no. 8488, pp. 1031–2, May 1986.

[145] A. Gessain, F. Barin, J. C. Vernant, O. Gout, L. Maurs, A. Calender, and G. de Thé, "Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis.," *Lancet (London, England)*, vol. 2, no. 8452, pp. 407–10, Aug. 1985.

[146] M. C. Choy, K. Visvanathan, and P. De Cruz, "An Overview of the Innate and Adaptive Immune System in Inflammatory Bowel Disease.," *Inflamm. Bowel Dis.*, vol. 23, no. 1, pp. 2–13, 2017.

[147] C. Abraham and J. H. Cho, "IL-23 and autoimmunity: new insights into the pathogenesis of inflammatory bowel disease.," *Annu. Rev. Med.*, vol. 60, pp. 97–110, 2009.

[148] G. Rogler and L. Biedermann, "Clinical Utility of Biomarkers in IBD.," *Curr. Gastroenterol. Rep.*, vol. 17, no. 7, p. 26, Jul. 2015.

[149] S. H. Saverymuttu, H. J. Hodgson, V. S. Chadwick, and M. B. Pepys, "Differing acute phase responses in Crohn's disease and ulcerative colitis.," *Gut*, vol. 27, no. 7, pp. 809–13, Jul. 1986.

[150] C. S. Carlson, S. F. Aldred, P. K. Lee, R. P. Tracy, S. M. Schwartz, M. Rieder, K. Liu, O. D. Williams, C. Iribarren, E. C. Lewis, M. Fornage, E. Boerwinkle, M. Gross, C. Jaquish, D. A. Nickerson, R. M. Myers, D. S. Siscovick, and A. P. Reiner, "Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels.," *Am. J. Hum. Genet.*, vol. 77, no. 1, pp. 64–77, Jul. 2005.

[151] F. Durchschein, W. Petritsch, and H. F. Hammer, "Diet therapy for inflammatory bowel diseases: The established and the new.," *World J. Gastroenterol.*, vol. 22, no. 7, pp. 2179–94, Feb. 2016.

[152] M. F. Neurath, "Cytokines in inflammatory bowel disease.," *Nat. Rev. Immunol.*, vol. 14, no. 5, pp. 329–42, May 2014.

[153] M. Nikpour, M. B. Urowitz, D. Ibañez, and D. D. Gladman, "Frequency and determinants of flare and persistently active disease in systemic lupus erythematosus.," *Arthritis Rheum.*, vol. 61, no. 9, pp. 1152–8, Sep. 2009.

[154] L. M. Ortega, D. R. Schultz, O. Lenz, V. Pardo, and G. N. Contreras, "Review: Lupus nephritis: pathologic features, epidemiology and a guide to therapeutic decisions.," *Lupus*, vol. 19, no. 5, pp. 557–74, Apr. 2010.

[155] R. Cervera, M. A. Khamashta, and G. R. V Hughes, "The Euro-lupus project: epidemiology of systemic lupus erythematosus in Europe.," *Lupus*, vol. 18, no. 10, pp. 869–74, Sep. 2009.

[156] G. Thomas, J. Mancini, N. Jourde-Chiche, G. Sarlon, Z. Amoura, J.-R. Harlé, E. Jougla, and L. Chiche, "Mortality associated with systemic lupus erythematosus in France assessed by multiple-cause-of-death analysis.," *Arthritis Rheumatol. (Hoboken, N.J.)*, vol. 66, no. 9, pp. 2503–11, Sep. 2014.

[157] N. Berghen, J.-B. Vulsteke, R. Westhovens, J. Lenaerts, and E. De Langhe, "Rituximab in systemic autoimmune rheumatic diseases: indications and practical use.," *Acta Clin. Belg.*, vol. 00, no. 00, pp. 1–8, Sep. 2018.

[158] J. M. Ahearn, C.-C. Liu, A. H. Kao, and S. Manzi, "Biomarkers for systemic lupus erythematosus.," *Transl. Res.*, vol. 159, no. 4, pp. 326–42, Apr. 2012.

[159] E. Clough and T. Barrett, "The Gene Expression Omnibus Database.," *Methods Mol. Biol.*, vol. 1418, pp. 93–110, 2016.

[160] C. Soneson and M. Delorenzi, "A comparison of methods for differential expression analysis of RNA-seq data.," *BMC Bioinformatics*, vol. 14, no. 1, p. 91, Jan. 2013.

[161] L. Wang and P. C. H. Li, "Microfluidic DNA microarray analysis: a review.," *Anal. Chim. Acta*, vol. 687, no. 1, pp. 12–27, Feb. 2011.

[162] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, 2009.

[163] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for Illumina sequence data," *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, Aug. 2014.

[164] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads," *EMBnet.journal*, vol. 17, no. 1, p. 10, May 2011.

[165] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szcześniak, D. J. Gaffney, L. L. Elo, X. Zhang, and A. Mortazavi, "A survey of best practices for RNA-seq data analysis," *Genome Biol.*, vol. 17, no. 1, p. 13, Dec. 2016.

[166] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," *Nat. Biotechnol.*, vol. 34, no. 5, pp. 525–527, May 2016.

[167] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis.," *BMC Bioinformatics*, vol. 9, p. 559, Jan. 2008.

[168] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proc. Natl. Acad. Sci.*, vol. 102, no. 43, pp. 15545–15550, Oct. 2005.

[169] M. J. Pencina, R. B. D'Agostino, and R. S. Vasan, "Statistical methods for assessment of added usefulness of new biomarkers.," *Clin. Chem. Lab. Med.*, vol. 48, no. 12, pp. 1703–11, Dec. 2010.

[170] N. R. Cook, "Use and misuse of the receiver operating characteristic curve in risk prediction.," *Circulation*, vol. 115, no. 7, pp. 928–35, Feb. 2007.

[171] C.-H. Chang, "Cohen's kappa for capturing discrimination.," *Int. Health*, vol. 6, no. 2, pp. 125–9, Jun. 2014.

[172] W. Vach, "The dependence of Cohen's kappa on the prevalence does not matter.," *J. Clin. Epidemiol.*, vol. 58, no. 7, pp.

655–61, Jul. 2005.

[173] S. Zec, N. Soriani, R. Comoretto, and I. Baldi, "High Agreement and High Prevalence: The Paradox of Cohen's Kappa.," *Open Nurs. J.*, vol. 11, no. Suppl-1, M5, pp. 211–218, 2017.

[174] A. Marçais, F. Suarez, D. Sibon, L. Frenzel, O. Hermine, and A. Bazarbachi, "Therapeutic Options for Adult T-Cell Leukemia/Lymphoma," *Curr. Oncol. Rep.*, vol. 15, no. 5, pp. 457–464, Oct. 2013.

[175] T. G. Johns, I. R. Mackay, K. A. Callister, P. J. Hertzog, R. J. Devenish, and A. W. Linnane, "Antiproliferative potencies of interferons on melanoma cell lines and xenografts: higher efficacy of interferon beta.," *J. Natl. Cancer Inst.*, vol. 84, no. 15, pp. 1185–90, Aug. 1992.

[176] S. Kinpara, A. Hasegawa, A. Utsunomiya, H. Nishitsuji, H. Furukawa, T. Masuda, and M. Kannagi, "Stromal Cell-Mediated Suppression of Human T-Cell Leukemia Virus Type 1 Expression In Vitro and In Vivo by Type I Interferon," *J. Virol.*, vol. 83, no. 10, pp. 5101–5108, May 2009.

[177] T. Oka, J. Iwata, M. Furihata, H. Sonobe, I. Miyoshi, and Y. Ohtsuki, "Inhibitory effects of human interferons on the immortalization of human, but not rabbit, T lymphocytes by human T-lymphotropic virus type-I (HTLV-I)," *Int J Cancer*, vol. 51, no. 6, pp. 915–920, 1992.

[178] B. Moens, C. Pannecouque, G. López, M. Talledo, E. Gotuzzo, R. Khouri, A. Bittencourt, L. Farré, B. Galvão-Castro, A.-M. Vandamme, and J. Van Weyenbergh, "Simultaneous RNA quantification of human and retroviral genomes reveals intact interferon signaling in HTLV-1-infected CD4+ T cell lines.," *Virol. J.*, vol. 9, no. 1, p. 171, Jan. 2012.

[179] A. Bazarbachi, R. Nasr, M. E. El-Sabban, A. Mahé, R. Mahieux, A. Gessain, N. Darwiche, G. Dbaibo, J. Kersual, Y. Zermati, L. Dianoux, M. K. Chelbi-Alix, H. de Thé, and O. Hermine, "Evidence against a direct cytotoxic effect of alpha interferon and zidovudine in HTLV-I associated adult T cell leukemia/lymphoma.," *Leukemia*, vol. 14, no. 4, pp. 716–21, Apr. 2000.

[180] D. Smith, G. J. Buckle, D. a Hafler, D. a Frank, and P. Höllsberg, "HTLV-I-Infected T Cells Evade the Antiproliferative Action of IFN-β," *Virology*, vol. 257, no. 2, pp. 314–321, May 1999.

[181] K. Tamura, S. Makino, Y. Araki, T. Imamura, and M. Seita, "Recombinant interferon beta and gamma in the treatment of adult T-cell leukemia," *Cancer*, vol. 59, no. 6, pp. 1059–1062, Mar. 1987.

[182] P. S. Gill, W. Harrington, M. H. Kaplan, R. C. Ribeiro, J. M. Bennett, H. A. Liebman, M. Bernstein-Singer, B. M. Espina, L. Cabral, S. Allen, S. Kornblau, M. C. Pike, and A. M. Levine, "Treatment of Adult T-Cell Leukemia–Lymphoma with a Combination of Interferon Alfa and Zidovudine," *N. Engl. J. Med.*, vol. 332, no. 26, pp. 1744–1748, Jun. 1995.

[183] S. Nakahata, Y. Saito, K. Marutsuka, T. Hidaka, K. Maeda, K. Hatakeyama, T. Shiraga, A. Goto, N. Takamatsu, Y. Asada, A. Utsunomiya, A. Okayama, Y. Kubuki, K. Shimoda, Y. Ukai, G. Kurosawa, and K. Morishita, "Clinical significance of CADM1/TSLC1/IgSF4 expression in adult T-cell leukemia/lymphoma.," *Leukemia*, vol. 26, no. 6, pp. 1238–46, Jun. 2012.

[184] A. a Alizadeh, S. P. Bohen, C. Lossos, J. a Martinez-Climent, J. C. Ramos, E. Cubedo-Gil, W. J. Harrington, and I. S. Lossos, "Expression profiles of adult T-cell leukemia-lymphoma and associations with clinical responses to zidovudine and interferon alpha.," *Leuk. Lymphoma*, vol. 51, no. 7, pp. 1200–16, Jul. 2010.

[185] S. Kinpara, M. Kijiyama, A. Takamori, A. Hasegawa, A. Sasada, T. Masuda, Y. Tanaka, A. Utsunomiya, and M. Kannagi, "Interferon-α (IFN-α) suppresses HTLV-1 gene expression and cell cycling, while IFN-α combined with zidovudin induces p53 signaling and apoptosis in HTLV-1-infected cells," *Retrovirology*, vol. 10, no. 1, p. 1, 2013.

[186] A. Datta, M. Bellon, U. Sinha-Datta, A. Bazarbachi, Y. Lepelletier, D. Canioni, T. A. Waldmann, O. Hermine, and C. Nicot, "Persistent inhibition of telomerase reprograms adult T-cell leukemia to p53-dependent senescence," *Blood*, vol. 108, no. 3, pp. 1021–1029, Aug. 2006.

[187] K. Kataoka, Y. Nagata, A. Kitanaka, Y. Shiraishi, T. Shimamura, J. Yasunaga, Y. Totoki, K. Chiba, A. Sato-Otsubo, G. Nagae, R. Ishii, S. Muto, S. Kotani, Y. Watatani, J. Takeda, M. Sanada, H. Tanaka, H. Suzuki, Y. Sato, Y. Shiozawa, T. Yoshizato, K. Yoshida, H. Makishima, M. Iwanaga, G. Ma, K. Nosaka, M. Hishizawa, H. Itonaga, Y. Imaizumi, W. Munakata, H. Ogasawara, T. Sato, K. Sasai, K. Muramoto, M. Penova, T. Kawaguchi, H. Nakamura, N. Hama, K. Shide, Y. Kubuki, T. Hidaka, T. Kameda, T. Nakamaki, K. Ishiyama, S. Miyawaki, S.-S. Yoon, K. Tobinai, Y. Miyazaki, A. Takaori-Kondo, F. Matsuda, K. Takeuchi, O. Nureki, H. Aburatani, T. Watanabe, T. Shibata, M. Matsuoka, S. Miyano, K. Shimoda, and S. Ogawa, "Integrated molecular analysis of adult T cell leukemia/lymphoma," *Nat. Genet.*, vol. 47, no. 11, pp. 1304–1315, Oct. 2015.

[188] J. Sancéau, J. Hiscott, O. Delattre, and J. Wietzerbin, "IFN-β induces serine phosphorylation of Stat-1 in Ewing's sarcoma cells and mediates apoptosis via induction of IRF-1 and activation of caspase-7," *Oncogene*, vol. 19, no. 30, pp. 3372–3383, Jul. 2000.

[189] A. L. Bittencourt, M. D. G. Vieira, C. R. Brites, L. Farre, and H. S. Barbosa, "Adult T-Cell Leukemia/Lymphoma in Bahia, Brazil," *Am. J. Clin. Pathol.*, vol. 128, no. 5, pp. 875–882, Nov. 2007.

[190] L. Farre, A. L. Bittencourt, G. Silva-Santos, A. Almeida, A. C. Silva, D. Decanine, G. M. Soares, L. C. Alcantara, S. Van Dooren, B. Galvão-Castro, A. M. Vandamme, and J. Van Weyenbergh, "Fas-670 promoter polymorphism is associated to susceptibility, clinical presentation, and survival in adult T cell leukemia," *J. Leukoc. Biol.*, vol. 83, no. 1, pp. 220–222, Jan. 2008.

[191] A. L. Bittencourt, H. S. Barbosa, C. Requião, A. C. da Silva, A.-M. Vandamme, J. Van Weyenbergh, and L. Farré, "Adult T-cell leukemia/lymphoma with a mixed CD4+ and CD8+ phenotype and indolent course.," *J. Clin. Oncol.*, vol. 25, no. 17, pp. 2480–2, Jun. 2007.

[192] B. Moens, D. Decanine, S. M. Menezes, R. Khouri, G. Silva-Santos, G. Lopez, C. Alvarez, M. Talledo, E. Gotuzzo, R. de Almeida Kruschewsky, B. Galvão-Castro, A.-M. Vandamme, and J. Van Weyenbergh, "Ascorbic Acid Has Superior Ex Vivo Antiproliferative, Cell Death-Inducing and Immunomodulatory Effects over IFN-α in HTLV-1-Associated Myelopathy," *PLoS Negl. Trop. Dis.*, vol. 6, no. 7, p. e1729, Jul. 2012.

[193] B. Macchi, C. D'Onofrio, R. A. Labianca, and E. Bonmassar, "Mononuclear cells from peripheral blood of adult donors and from cord blood are equally protected by ??- and ??-interferons against infection with HTLV-I," *Pharmacol. Res.*, vol. 22, no. 4, pp. 503–514, 1990.

[194] X. Feng, N. Vander Heyden, and L. Ratner, "Alpha interferon inhibits human T-cell leukemia virus type 1 assembly by preventing Gag interaction with rafts.," *J. Virol.*, vol. 77, no. 24, pp. 13389–95, Dec. 2003.

[195] R. Massoud, Y. Enose-Akahata, Y. Tagaya, N. Azimi, A. Basheer, and S. Jacobson, "Common γ-chain blocking peptide reduces in vitro immune activation markers in HTLV-1-associated myelopathy/tropical spastic paraparesis," *Proc. Natl. Acad. Sci.*, vol. 112, no. 35, pp. 11030–11035, Sep. 2015.

[196] J. Van Weyenbergh, J. Wietzerbin, D. Rouillard, M. Barral-Netto, and R. Liblau, "Treatment of multiple sclerosis patients with interferon-beta primes monocyte-derived macrophages for apoptotic cell death.," *J. Leukoc. Biol.*, vol. 70, no. 5, pp. 745–8, 2001.

[197] B. S. Carvalho and R. A. Irizarry, "A framework for oligonucleotide microarray preprocessing," *Bioinformatics*, vol. 26, no. 19, pp. 2363–2367, Oct. 2010.

[198] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Res.*, vol. 43, no. 7, pp. e47–e47, Apr. 2015.

[199] L. Varemo, J. Nielsen, and I. Nookaew, "Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods," *Nucleic Acids Res.*, vol. 41, no. 8, pp. 4378–4391, Apr. 2013.

[200] Y. Tagaya and R. C. Gallo, "The Exceptional Oncogenicity of HTLV-1.," *Front. Microbiol.*, vol. 8, no. AUG, p. 1425, 2017.

[201] K. Arisawa, M. Soda, S. Endo, K. Kurokawa, S. Katamine, I. Shimokawa, T. Koba, T. Takahashi, H. Saito, H. Doi, and S. Shirahama, "Evaluation of adult T-cell leukemia/lymphoma incidence and its impact on non-Hodgkin lymphoma incidence in southwestern Japan," *Int. J. Cancer*, vol. 85, no. 3, pp. 319–324, Feb. 2000.

[202] K. Takatsuki, "Discovery of adult T-cell leukemia.," *Retrovirology*, vol. 2, no. 16, p. 16, Mar. 2005.

[203] M. Iwanaga, T. Watanabe, and K. Yamaguchi, "Adult T-cell leukemia: a review of epidemiological evidence.," *Front. Microbiol.*, vol. 3, p. 322, 2012.

[204] H. Sasaki, "Overexpression of a cell adhesion molecule, TSLC1, as a possible molecular marker for acute-type adult T-cell leukemia," *Blood*, vol. 105, no. 3, pp. 1204–1213, Sep. 2004.

[205] Y. Sawada, M. Nakamura, R. Kabashima-Kubo, T. Shimauchi, M. Kobayashi, and Y. Tokura, "Defective Epidermal Innate Immunity and Resultant Superficial Dermatophytosis in Adult T-cell Leukemia/Lymphoma," *Clin. Cancer Res.*, vol. 18, no. 14, pp. 3772–3779, Jul. 2012.

[206] F. E. Leal, L. C. Ndhlovu, A. M. Hasenkrug, F. R. Bruno, K. I. Carvalho, H. Wynn-Williams, W. K. Neto, S. S. Sanabani, A. C. Segurado, D. F. Nixon, and E. G. Kallas, "Expansion in CD39+ CD4+ Immunoregulatory T Cells and Rarity of Th17 Cells in HTLV-1 Infected Patients Is Associated with Neurological Complications," *PLoS Negl. Trop. Dis.*, vol. 7, no. 2, p. e2028, Feb. 2013.

[207] H. R. Conti, A. C. Peterson, L. Brane, A. R. Huppler, N. Hernández-Santos, N. Whibley, A. V Garg, M. R. Simpson-Abelson, G. A. Gibson, A. J. Mamo, L. C. Osborne, S. Bishu, N. Ghilardi, U. Siebenlist, S. C. Watkins, D. Artis, M. J. McGeachy, and S. L. Gaffen, "Oral-resident natural Th17 cells and γδ T cells control opportunistic Candida albicans infections," *J. Exp. Med.*, vol. 211, no. 10, pp. 2075–2084, Sep. 2014.

[208] L. A. Tesmer, S. K. Lundy, S. Sarkar, and D. A. Fox, "Th17 cells in human disease," *Immunol. Rev.*, vol. 223, no. 1, pp. 87–113, Jun. 2008.

[209] M. M. Curtis and S. S. Way, "Interleukin-17 in host defence against bacterial, mycobacterial and fungal pathogens," *Immunology*, vol. 126, no. 2, pp. 177–185, Feb. 2009.

[210] A. Y. Swaims, F. Khani, Y. Zhang, A. I. Roberts, S. Devadas, Y. Shi, and A. B. Rabson, "Immune activation induces immortalization of HTLV-1 LTR-Tax transgenic CD4+ T cells," *Blood*, vol. 116, no. 16, pp. 2994–3003, Oct. 2010.

[211] M. D. Dodon, "Tax protein of human T-cell leukaemia virus type 1 induces interleukin 17 gene expression in T cells," *J. Gen. Virol.*, vol. 85, no. 7, pp. 1921–1932, Jul. 2004.

[212] K. M. Elias, A. Laurence, T. S. Davidson, G. Stephens, Y. Kanno, E. M. Shevach, and J. J. O'Shea, "Retinoic acid inhibits Th17 polarization and enhances FoxP3 expression through a Stat-3/Stat-5 independent signaling pathway," *Blood*, vol. 111, no. 3, pp. 1013–1020, Oct. 2007.

[213] C. R. M. Bangham, L. B. Cook, and A. Melamed, "HTLV-1 clonality in adult T-cell leukaemia and non-malignant HTLV-1 infection," *Semin. Cancer Biol.*, vol. 26, pp. 89–98, Jun. 2014.

[214] J. Lin, H. Tang, X. Jin, G. Jia, and J.-T. Hsieh, "p53 regulates Stat3 phosphorylation and DNA binding activity in human prostate cancer cells expressing constitutively active Stat3," *Oncogene*, vol. 21, no. 19, pp. 3082–3088, May 2002.

[215] E. Ueda, S. Kurebayashi, M. Sakaue, M. Backlund, B. Koller, and A. M. Jetten, "High incidence of T-cell lymphomas in mice deficient in the retinoid-related orphan receptor RORgamma.," *Cancer Res.*, vol. 62, no. 3, pp. 901–9, Feb. 2002.

[216] M. Liljevald, M. Rehnberg, M. Söderberg, M. Ramnegård, J. Börjesson, D. Luciani, N. Krutrök, L. Brändén, C. Johansson, X. Xu, M. Bjursell, A.-K. Sjögren, J. Hornberg, U. Andersson, D. Keeling, and J. Jirholt, "Retinoid-related orphan receptor γ (RORγ) adult induced knockout mice develop lymphoblastic lymphoma," *Autoimmun. Rev.*, vol. 15, no. 11, pp. 1062–1070, Nov. 2016.

[217] S. Tattermusch, J. a Skinner, D. Chaussabel, J. Banchereau, M. P. Berry, F. W. McNab, A. O'Garra, G. P. Taylor, and C. R. M. Bangham, "Systems biology approaches reveal a specific interferon-inducible signature in HTLV-1 associated myelopathy.," *PLoS Pathog.*, vol. 8, no. 1, p. e1002480, Jan. 2012.

[218] C. Ohnmacht, J.-H. Park, S. Cording, J. B. Wing, K. Atarashi, Y. Obata, V. Gaboriau-Routhiau, R. Marques, S. Dulauroy, M. Fedoseeva, M. Busslinger, N. Cerf-Bensussan, I. G. Boneca, D. Voehringer, K. Hase, K. Honda, S. Sakaguchi, and G. Eberl, "The microbiota regulates type 2 immunity through ROR t+ T cells," *Science (80-.).*, vol. 349, no. 6251, pp. 989–993, Aug. 2015.

[219] H. J. Motulsky and R. E. Brown, "Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate.," *BMC Bioinformatics*, vol. 7, no. 123, p. 123, Mar. 2006.

[220] J. C. Ramos, P. Ruiz, Jr., L. Ratner, I. M. Reis, C. Brites, C. Pedroso, G. E. Byrne, Jr., N. L. Toomey, V. Andela, E. W. Harhaj, I. Lossos, and W. J. Harrington, Jr., "IRF-4 and c-Rel expression in antiviral resistant adult T-cell leukemia/lymphoma," *Blood*, vol. 109, no. 7, pp. 3060–3068, Nov. 2006.

[221] H. Yu, D. Pardoll, and R. Jove, "STATs in cancer inflammation and immunity: a leading role for STAT3," *Nat. Rev. Cancer*, vol. 9, no. 11, pp. 798–809, Nov. 2009.

[222] P. Muranski and N. P. Restifo, "Essentials of Th17 cell commitment and plasticity," *Blood*, vol. 121, no. 13, pp. 2402–2414, Mar. 2013.

[223] R. Khouri, G. Silva-Santos, T. Dierckx, S. M. Menezes, D. Decanine, K. Theys, A. C. Silva, L. Farré, A. Bittencourt, M. Mangino, M. Roederer, A.-M. Vandamme, and J. Van Weyenbergh, "A genetic IFN/STAT1/FAS axis determines CD4 T stem cell memory levels and apoptosis in healthy controls and Adult T-cell Leukemia patients," *Oncoimmunology*, vol. 7, no. 5, p. e1426423, May 2018.

[224] R. Khouri, G. Silva-Santos, L. Farre, A. Bittencourt, A.-M. Vandamme, and J. Van Weyenbergh, "Superior antiviral and antiproliferative activity of IFN-beta vs. IFN-alpha in primary ATL cells occurs downstream of STAT1 signaling," *Retrovirology*, vol. 11, no. Suppl 1, p. O22, 2014.

[225] A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, and A. A. Alizadeh, "Robust enumeration of cell subsets from tissue expression profiles," *Nat. Methods*, vol. 12, no. 5, pp. 453–457, Mar. 2015.

[226] H. Kagdi, M. A. Demontis, J. C. Ramos, and G. P. Taylor, "Switching and loss of cellular cytokine producing capacity characterize in vivo viral infection and malignant transformation in human T- lymphotropic virus type 1 infection," *PLOS Pathog.*, vol. 14, no. 2, p. e1006861, Feb. 2018.

[227] Y. Satou, J. Yasunaga, T. Zhao, M. Yoshida, P. Miyazato, K. Takai, K. Shimizu, K. Ohshima, P. L. Green, N. Ohkura, T. Yamaguchi, M. Ono, S. Sakaguchi, and M. Matsuoka, "HTLV-1 bZIP Factor Induces T-Cell Lymphoma and Systemic Inflammation In Vivo," *PLoS Pathog.*, vol. 7, no. 2, p. e1001274, Feb. 2011.

[228] M. Mizuguchi, Y. Sasaki, T. Hara, M. Higuchi, Y. Tanaka, N. Funato, N. Tanaka, M. Fujii, and M. Nakamura, "Induction of Cell Death in Growing Human T-Cells and Cell Survival in Resting Cells in Response to the Human T-Cell Leukemia Virus Type 1 Tax," *PLoS One*, vol. 11, no. 2, p. e0148217, Feb. 2016.

[229] T. Dierckx, R. Khouri, S. M. Menezes, D. Decanine, L. Farre, A. Bittencourt, A. M. Vandamme, and J. Van Weyenbergh, "IFN-β induces greater antiproliferative and proapoptotic effects and increased p53 signaling compared with IFN-α in PBMCs of Adult T-cell Leukemia/Lymphoma patients," *Blood Cancer J.*, vol. 7, no. 1, pp. e519–e519, Jan. 2017.

[230] L. Farre, M. d. F. Paim de Oliveira, J. Primo, A.-M. Vandamme, J. Van Weyenbergh, and A. L. Bittencourt, "Early Sequential Development of Infective Dermatitis, Human T Cell Lymphotropic Virus Type 1-Associated Myelopathy, and Adult T Cell Leukemia/Lymphoma," *Clin. Infect. Dis.*, vol. 46, no. 3, pp. 440–442, Feb. 2008.

[231] B. Di Martino Ortiz, R. Riveros, R. Medina, and M. Morel, "Infective Dermatitis in an Adult Patient With HTLV-1," *Am. J. Dermatopathol.*, vol. 37, no. 12, pp. 944–948, Dec. 2015.

[232] M. d. F. S. P. de Oliveira, P. L. Fatal, J. R. L. Primo, J. L. S. da Silva, E. d. S. Batista, L. Farre, and A. L. Bittencourt, "Infective Dermatitis Associated With Human T-Cell Lymphotropic Virus Type 1: Evaluation of 42 Cases Observed in Bahia, Brazil," *Clin. Infect. Dis.*, vol. 54, no. 12, pp. 1714–1719, Jun. 2012.

[233] B. Hanchard, L. Lagrenade, C. Carberry, V. Fletcher, E. Williams, B. Cranston, W. Blattner, and A. Manns, "Childhood infective dermatitis evolving into adult T-cell leukaemia after 17 years," *Lancet*, vol. 338, no. 8782–8783, pp. 1593–1594, Dec. 1991.

[234] C. E. Zielinski, F. Mele, D. Aschenbrenner, D. Jarrossay, F. Ronchi, M. Gattorno, S. Monticelli, A. Lanzavecchia, and F. Sallusto, "Pathogen-induced human TH17 cells produce IFN-γ or IL-10 and are regulated by IL-1β," *Nature*, vol. 484, no. 7395, pp. 514–518, Apr. 2012.

[235] J. S. Bowers, M. H. Nelson, K. Majchrzak, S. R. Bailey, B. Rohrer, A. D. M. Kaiser, C. Atkinson, L. Gattinoni, and C. M. Paulos, "Th17 cells are refractory to senescence and retain robust antitumor activity after long-term ex vivo expansion," *JCI Insight*, vol. 2, no. 5, Mar. 2017.

[236] R. C. Axtell, C. Raman, and L. Steinman, "Interferon-β exacerbates Th17-mediated inflammatory disease.," *Trends*

Immunol., vol. 32, no. 6, pp. 272–7, Jun. 2011.

[237] I. Melero, D. M. Berman, M. A. Aznar, A. J. Korman, J. L. Pérez Gracia, and J. Haanen, "Evolving synergistic combinations of targeted immunotherapies to combat cancer.," *Nat. Rev. Cancer*, vol. 15, no. 8, pp. 457–72, Aug. 2015.

[238] K. Palucka and J. Banchereau, "Cancer immunotherapy via dendritic cells.," *Nat. Rev. Cancer*, vol. 12, no. 4, pp. 265–77, Mar. 2012.

[239] H. Zhang, L. Chen, S.-H. Cai, and H. Cheng, "Identification of TBK1 and IKKε, the non-canonical IκB kinases, as crucial pro-survival factors in HTLV-1-transformed T lymphocytes.," *Leuk. Res.*, vol. 46, pp. 37–44, 2016.

[240] X. Hu, X. Liu, J. Moisan, Y. Wang, C. A. Lesch, C. Spooner, R. W. Morgan, E. M. Zawidzka, D. Mertz, D. Bousley, K. Majchrzak, I. Kryczek, C. Taylor, C. Van Huis, D. Skalitzky, A. Hurd, T. D. Aicher, P. L. Toogood, G. D. Glick, C. M. Paulos, W. Zou, and L. L. Carter, "Synthetic RORγ agonists regulate multiple pathways to enhance antitumor immunity," *Oncoimmunology*, vol. 5, no. 12, p. e1254854, Dec. 2016.

[241] K. Kataoka, M. Iwanaga, J. Yasunaga, Y. Nagata, A. Kitanaka, T. Kameda, M. Yoshimitsu, Y. Shiraishi, A. Sato-Otsubo, M. Sanada, K. Chiba, H. Tanaka, Y. Ochi, K. Aoki, H. Suzuki, Y. Shiozawa, T. Yoshizato, Y. Sato, K. Yoshida, K. Nosaka, M. Hishizawa, H. Itonaga, Y. Imaizumi, W. Munakata, K. Shide, Y. Kubuki, T. Hidaka, T. Nakamaki, K. Ishiyama, S. Miyawaki, R. Ishii, O. Nureki, K. Tobinai, Y. Miyazaki, A. Takaori-Kondo, T. Shibata, S. Miyano, K. Ishitsuka, A. Utsunomiya, K. Shimoda, M. Matsuoka, T. Watanabe, and S. Ogawa, "Prognostic relevance of integrated genetic profiling in adult T-cell leukemia/lymphoma," *Blood*, vol. 131, no. 2, pp. 215–225, Jan. 2018.

[242] J. D. Wolchok, H. Kluger, M. K. Callahan, M. A. Postow, N. A. Rizvi, A. M. Lesokhin, N. H. Segal, C. E. Ariyan, R.-A. Gordon, K. Reed, M. M. Burke, A. Caldwell, S. A. Kronenberg, B. U. Agunwamba, X. Zhang, I. Lowy, H. D. Inzunza, W. Feely, C. E. Horak, Q. Hong, A. J. Korman, J. M. Wigginton, A. Gupta, and M. Sznol, "Nivolumab plus ipilimumab in advanced melanoma.," *N. Engl. J. Med.*, vol. 369, no. 2, pp. 122–33, Jul. 2013.

[243] S. Durinck, P. T. Spellman, E. Birney, and W. Huber, "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt," *Nat. Protoc.*, vol. 4, no. 8, pp. 1184–1191, Aug. 2009.

[244] S. Durinck, Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, and W. Huber, "BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis," *Bioinformatics*, vol. 21, no. 16, pp. 3439–3440, Aug. 2005.

[245] P. Warnat, R. Eils, and B. Brors, "Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes," *BMC Bioinformatics*, vol. 6, no. 1, p. 265, 2005.

[246] T. E. Sweeney, H. R. Wong, and P. Khatri, "Robust classification of bacterial and viral infections via integrated host gene expression diagnostics," *Sci. Transl. Med.*, vol. 8, no. 346, p. 346ra91-346ra91, Jul. 2016.

[247] E. L. Tsalik, R. Henao, M. Nichols, T. Burke, E. R. Ko, M. T. McClain, L. L. Hudson, A. Mazur, D. H. Freeman, T. Veldman, R. J. Langley, E. B. Quackenbush, S. W. Glickman, C. B. Cairns, A. K. Jaehne, E. P. Rivers, R. M. Otero, A. K. Zaas, S. F. Kingsmore, J. Lucas, V. G. Fowler, L. Carin, G. S. Ginsburg, and C. W. Woods, "Host gene expression classifiers diagnose acute respiratory illness etiology," *Sci. Transl. Med.*, vol. 8, no. 322, p. 322ra11-322ra11, 2016.

[248] S. Reardon, "Machine learning helps to hunt down the cause of a paralysing illness," *Nature*, vol. 564, no. 7735, pp. 170–171, Dec. 2018.

[249] P. G. Murray, A. Stevens, C. De Leonibus, E. Koledova, P. Chatelain, and P. E. Clayton, "Transcriptomics and machine learning predict diagnosis and severity of growth hormone deficiency.," *JCI insight*, vol. 3, no. 7, pp. 1–14, Apr. 2018.

[250] J. R. Bienkowska, G. S. Dalgin, F. Batliwalla, N. Allaire, R. Roubenoff, P. K. Gregersen, and J. P. Carulli, "Convergent random forest predictor: Methodology for predicting drug response from genome-scale data applied to anti-TNF response," *Genomics*, vol. 94, no. 6, pp. 423–432, Dec. 2009.

[251] J. H. Malone and B. Oliver, "Microarrays, deep sequencing and the true measure of the transcriptome.," *BMC Biol.*, vol. 9, p. 34, May 2011.

[252] J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey, "The sva package for removing batch effects and other unwanted variation in high-throughput experiments," *Bioinformatics*, vol. 28, no. 6, pp. 882–883, Mar. 2012.

[253] M. N. McCall, B. M. Bolstad, and R. a. Irizarry, "Frozen robust multiarray analysis (fRMA)," *Biostatistics*, vol. 11, no. 2, pp. 242–253, 2010.

[254] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, D. Y. W. Solis, C. Molter, R. Duque, H. Bersini, and A. Nowe, "GENESHIFT: A Nonparametric Approach for Integrating Microarray Gene Expression Data Based on the Inner Product as a Distance Measure between the Distributions of Genes," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 10, no. 2, pp. 383–392, Mar. 2013.

[255] J. Rudy and F. Valafar, "Empirical comparison of cross-platform normalization methods for gene expression data," *BMC Bioinformatics*, vol. 12, no. 1, p. 467, 2011.

[256] H.-C. Liu, P.-C. Peng, T.-C. Hsieh, T.-C. Yeh, C.-J. Lin, C.-Y. Chen, J.-Y. Hou, L.-Y. Shih, and D.-C. Liang, "Comparison of Feature Selection Methods for Cross-Laboratory Microarray Analysis," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 10, no. 3, pp. 593–604, May 2013.

[257] K. Uziela, "Making microarray and RNA-seq gene expression data comparable," Aalto University, 2012.

[258] K. Uziela and A. Honkela, "Probe Region Expression Estimation for RNA-Seq Data for Improved Microarray Comparability," *PLoS One*, vol. 10, no. 5, p. e0126545, May 2015.

[259] F. Battke and K. Nieselt, "Mayday SeaSight: Combined Analysis of Deep Sequencing and Microarray Data," *PLoS One*, vol. 6, no. 1, p. e16345, Jan. 2011.

[260] P. Soininen, A. J. Kangas, P. Würtz, T. Tukiainen, T. Tynkkynen, R. Laatikainen, M.-R. Järvelin, M. Kähönen, T. Lehtimäki, J. Viikari, O. T. Raitakari, M. J. Savolainen, and M. Ala-Korpela, "High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism," *Analyst*, vol. 134, no. 9, p. 1781, 2009.

[261] P. Würtz, A. J. Kangas, P. Soininen, D. A. Lawlor, G. Davey Smith, and M. Ala-Korpela, "Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies," *Am. J. Epidemiol.*, vol. 186, no. 9, pp. 1084–1096, Nov. 2017.

[262] J. Kettunen, S. C. Ritchie, O. Anufrieva, L.-P. Lyytikäinen, J. Hernesniemi, P. J. Karhunen, P. Kuukasjärvi, J. Laurikka, M. Kähönen, T. Lehtimäki, A. S. Havulinna, V. Salomaa, S. Männistö, M. Ala-Korpela, M. Perola, M. Inouye, and P. Würtz, "Mortality Risk in Angiography Patients," *Circ. Genomic Precis. Med.*, no. November, pp. 1–7, 2018.

[263] M. Inouye, K. Silander, E. Hamalainen, V. Salomaa, K. Harald, P. Jousilahti, S. Männistö, J. G. Eriksson, J. Saarela, S. Ripatti, M. Perola, G.-J. B. van Ommen, M.-R. Taskinen, A. Palotie, E. T. Dermitzakis, and L. Peltonen, "An Immune Response Network Associated with Blood Lipid Levels," *PLoS Genet.*, vol. 6, no. 9, p. e1001113, Sep. 2010.

[264] M. Inouye, J. Kettunen, P. Soininen, K. Silander, S. Ripatti, L. S. Kumpula, E. Hämäläinen, P. Jousilahti, A. J. Kangas, S. Männistö, M. J. Savolainen, A. Jula, J. Leiviskä, A. Palotie, V. Salomaa, M. Perola, M. Ala-Korpela, and L. Peltonen, "Metabonomic, transcriptomic, and genomic variation of a population cohort," *Mol. Syst. Biol.*, vol. 6, no. 441, p. 441, Dec. 2010.

[265] O. T. Raitakari, M. Juonala, T. Rönnemaa, L. Keltikangas-Järvinen, L. Räsänen, M. Pietikäinen, N. Hutri-Kähönen, L. Taittonen, E. Jokinen, J. Marniemi, A. Jula, R. Telama, M. Kähönen, T. Lehtimäki, H. K. Akerblom, and J. S. A. Viikari, "Cohort profile: the cardiovascular risk in Young Finns Study.," *Int. J. Epidemiol.*, vol. 37, no. 6, pp. 1220–6, Dec. 2008.

[266] J. a Miller, C. Cai, P. Langfelder, D. H. Geschwind, S. M. Kurian, D. R. Salomon, and S. Horvath, "Strategies for aggregating gene expression data: The collapseRows R function.," *BMC Bioinformatics*, vol. 12, no. 1, p. 322, 2011.

[267] N. Altman and M. Krzywinski, "The curse(s) of dimensionality.," *Nat. Methods*, vol. 15, no. 6, pp. 399–400, Jun. 2018.

[268] Z. M. Hira and D. F. Gillies, "A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data.," *Adv. Bioinformatics*, vol. 2015, no. 1, p. 198363, 2015.

[269] M. B. Kursa and W. R. Rudnicki, "Feature Selection with the Boruta Package," *J. Stat. Softw.*, vol. 36, no. 11, pp. 1–13, 2010.

[270] A. Kalousis, J. Prados, and M. Hilario, "Stability of Feature Selection Algorithms," in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, 2005, pp. 218–225.

[271] M. C. from J. W. and S. W. and A. W. and C. K. and A. E. Kuhn, "caret: Classification and Regression Training," *Journal of Statistical Software*, vol. 28, no. 5. 2008.

[272] N. De Jay, S. Papillon-Cavanagh, C. Olsen, N. El-Hachem, G. Bontempi, and B. Haibe-Kains, "mRMRe: an R package for parallelized mRMR ensemble feature selection," *Bioinformatics*, vol. 29, no. 18, pp. 2365–2368, Sep. 2013.

[273] J. Wang, D. Duncan, Z. Shi, and B. Zhang, "WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013.," *Nucleic Acids Res.*, vol. 41, no. Web Server issue, pp. W77-83, Jul. 2013.

[274] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 2016, vol. 42, no. 8, pp. 785–794.

[275] D. Venet, J. E. Dumont, and V. Detours, "Most random gene expression signatures are significantly associated with breast cancer outcome.," *PLoS Comput. Biol.*, vol. 7, no. 10, p. e1002240, Oct. 2011.

[276] E. Theodoratou, H. Campbell, N. T. Ventham, D. Kolarich, M. Pučić-Baković, V. Zoldoš, D. Fernandes, I. K. Pemberton, I. Rudan, N. A. Kennedy, M. Wuhrer, E. Nimmo, V. Annese, D. P. B. McGovern, J. Satsangi, and G. Lauc, "The role of glycosylation in IBD.," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 11, no. 10, pp. 588–600, Oct. 2014.

[277] A. O. Akinkuolie, R. J. Glynn, L. Padmanabhan, P. M. Ridker, and S. Mora, "Circulating N-Linked Glycoprotein Side-Chain Biomarker, Rosuvastatin Therapy, and Incident Cardiovascular Disease: An Analysis From the JUPITER Trial," *J. Am. Heart Assoc.*, vol. 5, no. 7, p. e003822, Jul. 2016.

[278] P. R. Lawler, A. O. Akinkuolie, P. D. Chandler, M. V. Moorthy, M. J. Vandenburgh, D. A. Schaumberg, I.-M. Lee, R. J. Glynn, P. M. Ridker, J. E. Buring, and S. Mora, "Circulating N-Linked Glycoprotein Acetyls and Longitudinal Mortality RiskNovelty and Significance," *Circ. Res.*, vol. 118, no. 7, pp. 1106–1115, Apr. 2016.

[279] A. M. Dias, J. Dourado, P. Lago, J. Cabral, R. Marcos-Pinto, P. Salgueiro, C. R. Almeida, S. Carvalho, S. Fonseca, M. Lima, M. Vilanova, M. Dinis-Ribeiro, C. A. Reis, and S. S. Pinho, "Dysregulation of T cell receptor N-glycosylation: a molecular mechanism involved in ulcerative colitis.," *Hum. Mol. Genet.*, vol. 23, no. 9, pp. 2416–27, May 2014.

[280] A. M. Dias, A. Correia, M. S. Pereira, C. R. Almeida, I. Alves, V. Pinto, T. A. Catarino, N. Mendes, M. Leander, M. T. Oliva-Teles, L. Maia, C. Delerue-Matos, N. Taniguchi, M. Lima, I. Pedroto, R. Marcos-Pinto, P. Lago, C. A. Reis, M. Vilanova, and S. S. Pinho, "Metabolic control of T cell immune response through glycans in inflammatory bowel disease.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 20, pp. E4651–E4660, May 2018.

[281] J.-F. Colombel, R. Panaccione, P. Bossuyt, M. Lukas, F. Baert, T. Vaňásek, A. Danalioglu, G. Novacek, A. Armuzzi, X. Hébuterne, S. Travis, S. Danese, W. Reinisch, W. J. Sandborn, P. Rutgeerts, D. Hommes, S. Schreiber, E. Neimark, B. Huang, Q. Zhou, P. Mendez, J. Petersson, K. Wallace, A. M. Robinson, R. B. Thakkar, and G. D'Haens, "Effect of tight control management on Crohn's disease (CALM): a multicentre, randomised, controlled phase 3 trial.," *Lancet*

(London, England), vol. 390, no. 10114, pp. 2779–2789, 2018.

[282] E. Zittan, O. B. Kelly, R. Kirsch, R. Milgrom, J. Burns, G. C. Nguyen, K. Croitoru, G. Van Assche, M. S. Silverberg, and A. H. Steinhart, "Low Fecal Calprotectin Correlates with Histological Remission and Mucosal Healing in Ulcerative Colitis and Colonic Crohn's Disease.," *Inflamm. Bowel Dis.*, vol. 22, no. 3, pp. 623–30, Mar. 2016.

[283] A. Adler, S. Geiger, A. Keil, H. Bias, P. Schatz, T. DeVos, J. Dhein, M. Zimmermann, R. Tauber, and B. Wiedenmann, "Improving compliance to colorectal cancer screening using blood and stool based tests in patients refusing screening colonoscopy in Germany.," *BMC Gastroenterol.*, vol. 14, p. 183, Oct. 2014.

[284] A. Sturm, C. Maaser, E. Calabrese, V. Annese, G. Fiorino, T. Kucharzik, S. R. Vavricka, B. Verstockt, P. van Rheenen, D. Tolan, S. A. Taylor, J. Rimola, F. Rieder, J. K. Limdi, A. Laghi, E. Krustinš, P. G. Kotze, U. Kopylov, K. Katsanos, S. Halligan, H. Gordon, Y. González Lama, P. Ellul, R. Eliakim, F. Castiglione, J. Burisch, P. Borralho Nunes, D. Bettenworth, D. C. Baumgart, J. Stoker, European Crohn's and Colitis Organisation [ECCO], and European Society of Gastrointestinal and Abdominal Radiology [ESGAR], "ECCO-ESGAR Guideline for Diagnostic Assessment in Inflammatory Bowel Disease.," *J. Crohns. Colitis*, Aug. 2018.

[285] B. Verstockt, M. Noman, I. Aerden, M. Peeters, E. Brouwers, V. Ballet, L. Vandersmissen, G. A. Van Assche, A. Gils, S. Vermeire, and M. Ferrante, "Mo1880 - Ustekinumab Induces Limited Mucosal Healing after 6 Months in a Real-Life, Prospective Cohort of Patients with Refractory Crohn's Disease," *Gastroenterology*, vol. 154, no. 6, p. S-836, May 2018.

[286] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2017.

[287] H. Wickam, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[288] C. Rungoe, S. Basit, M. F. Ranthe, J. Wohlfahrt, E. Langholz, and T. Jess, "Risk of ischaemic heart disease in patients with inflammatory bowel disease: a nationwide Danish cohort study.," *Gut*, vol. 62, no. 5, pp. 689–94, May 2013.

[289] F. Clerc, M. Novokmet, V. Dotz, K. R. Reiding, N. de Haan, G. S. M. Kammeijer, H. Dalebout, M. R. Bladergroen, F. Vukovic, E. Rapp, IBD-BIOM consortium, S. R. Targan, G. Barron, N. Manetti, A. Latiano, D. P. B. McGovern, V. Annese, G. Lauc, and M. Wuhrer, "Plasma N-Glycan Signatures Associate With Features of Inflammatory Bowel Diseases.," *Gastroenterology*, May 2018.

[290] C. Mohan and C. Putterman, "Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis," *Nat. Rev. Nephrol.*, vol. 11, no. 6, pp. 329–341, Jun. 2015.

[291] L. Chiche, N. Jourde-Chiche, E. Whalen, S. Presnell, V. Gersuk, K. Dang, E. Anguiano, C. Quinn, S. Burtey, Y. Berland, G. Kaplanski, J. R. Harle, V. Pascual, and D. Chaussabel, "Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type i and type ii interferon signatures," *Arthritis Rheumatol.*, vol. 66, no. 6, pp. 1583–1595, 2014.

[292] N. Rother, E. Pieterse, J. Lubbers, L. Hilbrands, and J. van der Vlag, "Acetylated Histones in Apoptotic Microparticles Drive the Formation of Neutrophil Extracellular Traps in Active Lupus Nephritis," *Front. Immunol.*, vol. 8, no. SEP, pp. 1–12, Sep. 2017.

[293] N. Jourde-Chiche, E. Whalen, B. Gondouin, C. Speake, V. Gersuk, B. Dussol, S. Burtey, V. Pascual, D. Chaussabel, and L. Chiche, "Modular transcriptional repertoire analyses identify a blood neutrophil signature as a candidate biomarker for lupus nephritis," *Rheumatology*, vol. 56, no. 3, p. kew439, Dec. 2016.

[294] G. K. Bertsias, M. Tektonidou, Z. Amoura, M. Aringer, I. Bajema, J. H. M. Berden, J. Boletis, R. Cervera, T. Dörner, A. Doria, F. Ferrario, J. Floege, F. A. Houssiau, J. P. A. Ioannidis, D. A. Isenberg, C. G. M. Kallenberg, L. Lightstone, S. D. Marks, A. Martini, G. Moroni, I. Neumann, M. Praga, M. Schneider, A. Starra, V. Tesar, C. Vasconcelos, R. F. van Vollenhoven, H. Zakharova, M. Haubitz, C. Gordon, D. Jayne, and D. T. Boumpas, "Joint European League Against Rheumatism and European Renal Association–European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of adult and paediatric lupus nephritis," *Ann. Rheum. Dis.*, vol. 71, no. 11, pp. 1771–1782, Nov. 2012.

[295] M. A. Connelly, J. D. Otvos, I. Shalaurova, M. P. Playford, and N. N. Mehta, "GlycA, a novel biomarker of systemic inflammation and cardiovascular disease risk," *J. Transl. Med.*, vol. 15, no. 1, p. 219, Dec. 2017.

[296] T. Dierckx, B. Verstockt, S. Vermeire, and J. van Weyenbergh, "GlycA, a nuclear magnetic resonance spectroscopy measure for protein glycosylation, is a viable biomarker for disease activity in IBD.," *J. Crohns. Colitis*, pp. 1–6, Oct. 2018.

[297] M. Petri, A.-M. Orbai, G. S. Alarcón, C. Gordon, J. T. Merrill, P. R. Fortin, I. N. Bruce, D. Isenberg, D. J. Wallace, O. Nived, G. Sturfelt, R. Ramsey-Goldman, S.-C. Bae, J. G. Hanly, J. Sánchez-Guerrero, A. Clarke, C. Aranow, S. Manzi, M. Urowitz, D. Gladman, K. Kalunian, M. Costner, V. P. Werth, A. Zoma, S. Bernatsky, G. Ruiz-Irastorza, M. A. Khamashta, S. Jacobsen, J. P. Buyon, P. Maddison, M. A. Dooley, R. F. van Vollenhoven, E. Ginzler, T. Stoll, C. Peschken, J. L. Jorizzo, J. P. Callen, S. S. Lim, B. J. Fessler, M. Inanc, D. L. Kamen, A. Rahman, K. Steinsson, A. G. Franks, L. Sigler, S. Hameed, H. Fang, N. Pham, R. Brey, M. H. Weisman, G. McGwin, and L. S. Magder, "Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus.," *Arthritis Rheum.*, vol. 64, no. 8, pp. 2677–86, Aug. 2012.

[298] A. S. Levey, J. Coresh, T. Greene, L. A. Stevens, Y. L. Zhang, S. Hendriksen, J. W. Kusek, F. Van Lente, and Chronic Kidney Disease Epidemiology Collaboration, "Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate.," *Ann. Intern. Med.*, vol. 145, no. 4, pp. 247–54, Aug. 2006.

[299] Z. Zhang, "Variable selection with stepwise and best subset approaches," *Ann. Transl. Med.*, vol. 4, no. 7, pp. 136–136, Apr. 2016.

[300] B. Carstensen, M. Plummer, E. Laara, and M. Hills, "{Epi}: A Package for Statistical Analysis in Epidemiology." 2018.

[301] P. D. Chandler, A. O. Akinkuolie, D. K. Tobias, P. R. Lawler, C. Li, M. V. Moorthy, L. Wang, D. A. Duprez, D. R. Jacobs, R. J. Glynn, J. Otvos, M. A. Connelly, W. S. Post, P. M. Ridker, J. E. Manson, J. E. Buring, I.-M. Lee, and S. Mora, "Association of N-Linked Glycoprotein Acetyls and Colorectal Cancer Incidence and Mortality," *PLoS One*, vol. 11, no. 11, p. e0165615, Nov. 2016.

[302] M. C. Harpsøe, S. Basit, M. Andersson, N. M. Nielsen, M. Frisch, J. Wohlfahrt, E. A. Nohr, A. Linneberg, and T. Jess, "Body mass index and risk of autoimmune diseases: a study within the Danish National Birth Cohort.," *Int. J. Epidemiol.*, vol. 43, no. 3, pp. 843–55, Jun. 2014.

[303] M. Versini, P.-Y. Jeandel, E. Rosenthal, and Y. Shoenfeld, "Obesity in autoimmune diseases: not a passive bystander.," *Autoimmun. Rev.*, vol. 13, no. 9, pp. 981–1000, Sep. 2014.

[304] A. Oeser, C. P. Chung, Y. Asanuma, I. Avalos, and C. M. Stein, "Obesity is an independent contributor to functional capacity and inflammation in systemic lupus erythematosus," *Arthritis Rheum.*, vol. 52, no. 11, pp. 3651–3659, 2005.

[305] A. Rizk, T. A. Gheita, S. Nassef, and A. Abdallah, "The impact of obesity in systemic lupus erythematosus on disease parameters, quality of life, functional capacity and the risk of atherosclerosis.," *Int. J. Rheum. Dis.*, vol. 15, no. 3, pp. 261–7, Jun. 2012.

[306] V. D. D'Agati, A. Chagnac, A. P. J. de Vries, M. Levi, E. Porrini, M. Herman-Edelstein, and M. Praga, "Obesity-related glomerulopathy: clinical and pathologic characteristics and pathogenesis.," *Nat. Rev. Nephrol.*, vol. 12, no. 8, pp. 453–71, 2016.

[307] C. Menni, I. Gudelj, E. MacDonald-Dunlop, M. Mangino, J. Zierer, E. Bešić, P. Joshi, I. Trbojević-Akmačić, P. J. Chowienczyk, T. D. Spector, J. F. Wilson, G. Lauc, and A. M. Valdes, "Glycosylation Profile of Immunoglobulin G Is Cross-Sectionally Associated with Cardiovascular Disease Risk Score and Subclinical Atherosclerosis in Two Independent Cohorts," *Circ. Res.*, vol. 44, no. February, p. CIRCRESAHA.117.312174, Mar. 2018.

[308] L. B. Cook, S. Fuji, O. Hermine, A. Bazarbachi, J. C. Ramos, L. Ratner, S. Horwitz, P. Fields, A. Tanase, H. Bumbea, K. Cwynarski, G. Taylor, T. A. Waldmann, A. Bittencourt, A. Marcais, F. Suarez, D. Sibon, A. Phillips, M. Lunning, R. Farid, Y. Imaizumi, I. Choi, T. Ishida, K. Ishitsuka, T. Fukushima, K. Uchimaru, A. Takaori-Kondo, Y. Tokura, A. Utsunomiya, M. Matsuoka, K. Tsukasaki, and T. Watanabe, "Revised Adult T-Cell Leukemia-Lymphoma International Consensus Meeting Report," *J. Clin. Oncol.*, vol. 37, no. 8, pp. 677–687, 2019.

[309] M. Papatriantafyllou, "Infection: the interferon paradox.," *Nat. Rev. Immunol.*, vol. 13, no. 6, p. 392, Jun. 2013.

[310] J. R. Teijaro, C. Ng, A. M. Lee, B. M. Sullivan, K. C. F. Sheehan, M. Welch, R. D. Schreiber, J. C. de la Torre, and M. B. A. Oldstone, "Persistent LCMV infection is controlled by blockade of type I interferon signaling.," *Science*, vol. 340, no. 6129, pp. 207–11, Apr. 2013.

[311] E. B. Wilson, D. H. Yamada, H. Elsaesser, J. Herskovitz, J. Deng, G. Cheng, B. J. Aronow, C. L. Karp, and D. G. Brooks, "Blockade of chronic type I interferon signaling to control persistent LCMV infection.," *Science*, vol. 340, no. 6129, pp. 202–7, Apr. 2013.

[312] P. P. Tak and G. S. Firestein, "NF-kappaB: a key role in inflammatory diseases.," *J. Clin. Invest.*, vol. 107, no. 1, pp. 7–11, Jan. 2001.

[313] T. Lawrence, D. W. Gilroy, P. R. Colville-Nash, and D. A. Willoughby, "Possible new role for NF-kappaB in the resolution of inflammation.," *Nat. Med.*, vol. 7, no. 12, pp. 1291–7, Dec. 2001.

[314] C. H. Y. Fong, M. Bebien, A. Didierlaurent, R. Nebauer, T. Hussell, D. Broide, M. Karin, and T. Lawrence, "An antiinflammatory role for IKKβ through the inhibition of 'classical' macrophage activation," *J. Exp. Med.*, vol. 205, no. 6, pp. 1269–1276, 2008.

[315] S. Gordon, "Alternative activation of macrophages.," *Nat. Rev. Immunol.*, vol. 3, no. 1, pp. 23–35, Jan. 2003.

[316] T. Lawrence, "The nuclear factor NF-kappaB pathway in inflammation.," *Cold Spring Harb. Perspect. Biol.*, vol. 1, no. 6, p. a001651, Dec. 2009.

[317] A. Waisman, J. Hauptmann, and T. Regen, "The role of IL-17 in CNS diseases.," *Acta Neuropathol.*, vol. 129, no. 5, pp. 625–37, May 2015.

[318] M. Chen, G. Chen, H. Nie, X. Zhang, X. Niu, Y. C. Q. Zang, S. M. Skinner, J. Z. Zhang, J. M. Killian, and J. Hong, "Regulatory effects of IFN-beta on production of osteopontin and IL-17 by CD4+ T Cells in MS.," *Eur. J. Immunol.*, vol. 39, no. 9, pp. 2525–36, Sep. 2009.

[319] V. S. Ramgolam, Y. Sha, J. Jin, X. Zhang, and S. Markovic-Plese, "IFN-beta inhibits human Th17 cell differentiation.," *J. Immunol.*, vol. 183, no. 8, pp. 5418–27, Oct. 2009.

[320] Y. Warabi, Y. Matsumoto, and H. Hayashi, "Interferon beta-1b exacerbates multiple sclerosis with severe optic nerve and spinal cord demyelination.," *J. Neurol. Sci.*, vol. 252, no. 1, pp. 57–61, Jan. 2007.

[321] S. M. Menezes, D. Decanine, D. Brassat, R. Khouri, S. V Schnitman, R. Kruschewsky, G. López, C. Alvarez, M. Talledo, E. Gotuzzo, A. Vandamme, B. Galvão-Castro, R. Liblau, and J. Van Weyenbergh, "CD80+ and CD86+ B cells as biomarkers and possible therapeutic targets in HTLV-1 associated myelopathy/tropical spastic paraparesis and multiple sclerosis.," *J. Neuroinflammation*, vol. 11, p. 18, Jan. 2014.

[322] F. E. Leal, L. C. Ndhlovu, A. M. Hasenkrug, F. R. Bruno, K. I. Carvalho, H. Wynn-Williams, W. K. Neto, S. S.

Sanabani, A. C. Segurado, D. F. Nixon, and E. G. Kallas, "Expansion in CD39+ CD4+ immunoregulatory t cells and rarity of Th17 cells in HTLV-1 infected patients is associated with neurological complications.," *PLoS Negl. Trop. Dis.*, vol. 7, no. 2, p. e2028, 2013.

[323] F. L. S. do Prado, R. Prado, and A. M. T. Ladeia, "Cardiovascular risk profile in patients with myelopathy associated with HTLV-1.," *Braz. J. Infect. Dis.*, vol. 21, no. 3, pp. 226–233, 2017.

[324] S. Samadi, S. Abolbashari, Z. Meshkat, A. H. Mohammadpour, T. Kelesidis, A. Gholoobi, M. Mehramiz, M. Tabadkani, F. Sadabadi, R. Dalirfardouei, G. A. Ferns, M. Ghayour-Mobarhan, and A. Avan, "Human T lymphotropic virus type 1 and risk of cardiovascular disease: High-density lipoprotein dysfunction versus serum HDL-C concentrations.," *Biofactors*, no. December 2018, pp. 1–7, Jan. 2019.

[325] M. Yamamura, Y. Yamada, S. Momita, S. Kamihira, and M. Tomonaga, "Circulating interleukin-6 levels are elevated in adult T-cell leukaemia/lymphoma patients and correlate with adverse clinical features and survival.," *Br. J. Haematol.*, vol. 100, no. 1, pp. 129–34, Jan. 1998.

[326] D. Ganguly, S. Haak, V. Sisirak, and B. Reizis, "The role of dendritic cells in autoimmunity.," *Nat. Rev. Immunol.*, vol. 13, no. 8, pp. 566–77, Aug. 2013.

[327] A. S. Elshikha, Y. Yuan, Y. Lu, M.-J. Chen, G. Abboud, M. A. Akbar, H. Plate, H. Wolney, T. Hoffmann, E. Tagari, L. Zeumer, L. Morel, and S. Song, "Alpha 1 Antitrypsin Gene Therapy Extends the Lifespan of Lupus-Prone Mice.," *Mol. Ther. Methods Clin. Dev.*, vol. 11, no. December, pp. 131–142, Dec. 2018.

[328] F. P. Siegal, N. Kadowaki, M. Shodell, P. A. Fitzgerald-Bocarsly, K. Shah, S. Ho, S. Antonenko, and Y. J. Liu, "The nature of the principal type 1 interferon-producing cells in human blood.," *Science*, vol. 284, no. 5421, pp. 1835–7, Jun. 1999.

[329] N. Shirafuji, S. Matsuda, H. Ogura, K. Tani, H. Kodo, K. Ozawa, S. Nagata, S. Asano, and F. Takaku, "Granulocyte colony-stimulating factor stimulates human mature neutrophilic granulocytes to produce interferon-alpha.," *Blood*, vol. 75, no. 1, pp. 17–9, Jan. 1990.

[330] N. Tamassia, V. Le Moigne, M. Rossato, M. Donini, S. McCartney, F. Calzetti, M. Colonna, F. Bazzoni, and M. A. Cassatella, "Activation of an immunoregulatory and antiviral gene expression program in poly(I:C)-transfected human neutrophils.," *J. Immunol.*, vol. 181, no. 9, pp. 6563–73, Nov. 2008.

[331] P. Decker, "Neutrophils and interferon-α-producing cells: who produces interferon in lupus?," *Arthritis Res. Ther.*, vol. 13, no. 4, p. 118, Jul. 2011.

[332] M. S. Pereira, L. Maia, L. F. Azevedo, S. Campos, S. Carvalho, A. M. Dias, A. Albergaria, J. Lima, R. Marcos-Pinto, P. Lago, and S. S. Pinho, "A [Glyco]biomarker that Predicts Failure to Standard Therapy in Ulcerative Colitis Patients.," *J. Crohns. Colitis*, vol. 13, no. 1, pp. 39–49, Jan. 2019.

[333] A. Dias, M. Pereira, A. Correia, I. Alves, V. Pinto, L. Azevedo, L. Maia, R. Marcos-Pinto, M. Vilanova, P. Lago, and S. Pinho, "OP04 Turning sweet in inflammatory bowel disease: glycans as novel immunomodulators of T-cell-mediated immune response," *J. Crohn's Colitis*, vol. 13, no. Supplement_1, pp. S002–S003, Jan. 2019.

[334] M. Swamy, S. Pathak, K. M. Grzes, S. Damerow, L. V. Sinclair, D. M. F. van Aalten, and D. A. Cantrell, "Glucose and glutamine fuel protein O-GlcNAcylation to control T cell self-renewal and malignancy.," *Nat. Immunol.*, vol. 17, no. 6, pp. 712–20, 2016.

[335] W. H. Yang and J. D. Marth, "Protein glycosylation energizes T cells.," *Nat. Immunol.*, vol. 17, no. 6, pp. 613–4, 2016.

[336] H. Röytiö, K. Mokkala, T. Vahlberg, and K. Laitinen, "Dietary intake of fat and fibre according to reference values relates to higher gut microbiota richness in overweight pregnant women," *Br. J. Nutr.*, vol. 118, no. 05, pp. 343–352, 2017.

[337] F. F. Anhê, T. V. Varin, J. D. Schertzer, and A. Marette, "The Gut Microbiota as a Mediator of Metabolic Benefits after Bariatric Surgery.," *Can. J. diabetes*, vol. 41, no. 4, pp. 439–447, Aug. 2017.

[338] Y. Guo, Z.-P. Huang, C.-Q. Liu, L. Qi, Y. Sheng, and D.-J. Zou, "Modulation of the gut microbiome: a systematic review of the effect of bariatric surgery.," *Eur. J. Endocrinol.*, vol. 178, no. 1, pp. 43–56, 2018.

[339] G. A. Kelley and K. S. Kelley, "Effects of aerobic exercise on C-reactive protein, body composition, and maximum oxygen consumption in adults: a meta-analysis of randomized controlled trials.," *Metabolism.*, vol. 55, no. 11, pp. 1500–7, Nov. 2006.

[340] J. M. Allen, L. J. Mailing, G. M. Niemiro, R. Moore, M. D. Cook, B. A. White, H. D. Holscher, and J. A. Woods, "Exercise Alters Gut Microbiota Composition and Function in Lean and Obese Humans.," *Med. Sci. Sports Exerc.*, vol. 50, no. 4, pp. 747–757, 2018.

[341] R. Codella, L. Luzi, and I. Terruzzi, "Exercise has the guts: How physical activity may positively modulate gut microbiota in chronic and immune-based diseases.," *Dig. Liver Dis.*, vol. 50, no. 4, pp. 331–341, Apr. 2018.

[342] C. Grimstein, Y.-K. Choi, C. H. Wasserfall, M. Satoh, M. A. Atkinson, M. L. Brantly, M. Campbell-Thompson, and S. Song, "Alpha-1 antitrypsin protein and gene therapies decrease autoimmunity and delay arthritis development in mouse model.," *J. Transl. Med.*, vol. 9, no. 1, p. 21, Feb. 2011.

[343] S. Song, K. Goudy, M. Campbell-Thompson, C. Wasserfall, M. Scott-Jorgensen, J. Wang, Q. Tang, J. M. Crawford, T. M. Ellis, M. A. Atkinson, and T. R. Flotte, "Recombinant adeno-associated virus-mediated alpha-1 antitrypsin gene therapy prevents type I diabetes in NOD mice.," *Gene Ther.*, vol. 11, no. 2, pp. 181–6, Jan. 2004.

[344] A. M. Goetze, Y. D. Liu, Z. Zhang, B. Shah, E. Lee, P. V. Bondarenko, and G. C. Flynn, "High-mannose glycans on the Fc region of therapeutic IgG antibodies increase serum clearance in humans.," *Glycobiology*, vol. 21, no. 7, pp. 949–

59, Jul. 2011.

[345] P. Zhang, S. Woen, T. Wang, B. Liau, S. Zhao, C. Chen, Y. Yang, Z. Song, M. R. Wormald, C. Yu, and P. M. Rudd, "Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs," *Drug Discov. Today*, vol. 21, no. 5, pp. 740–765, May 2016.

[346] T. M. Maguire, A. M. Gillian, D. O'Mahony, C. M. Coughlan, A. Dennihan, and K. C. Breen, "A decrease in serum sialyltransferase levels in Alzheimer's disease.," *Neurobiol. Aging*, vol. 15, no. 1, pp. 99–102, 1994.

[347] M. Taniguchi, Y. Okayama, Y. Hashimoto, M. Kitaura, D. Jimbo, Y. Wakutani, K. Wada-Isoe, K. Nakashima, H. Akatsu, K. Furukawa, H. Arai, and K. Urakami, "Sugar chains of cerebrospinal fluid transferrin as a new biological marker of Alzheimer's disease.," *Dement. Geriatr. Cogn. Disord.*, vol. 26, no. 2, pp. 117–22, 2008.

[348] S. Schedin-Weiss, B. Winblad, and L. O. Tjernberg, "The role of protein glycosylation in Alzheimer disease.," *FEBS J.*, vol. 281, no. 1, pp. 46–62, Jan. 2014.

[349] A. Epp, K. C. Sullivan, A. B. Herr, and R. T. Strait, "Immunoglobulin Glycosylation Effects in Allergy and Immunity.," *Curr. Allergy Asthma Rep.*, vol. 16, no. 11, p. 79, Nov. 2016.

[350] J. Tynkkynen, V. Chouraki, S. J. van der Lee, J. Hernesniemi, Q. Yang, S. Li, A. Beiser, M. G. Larson, K. Sääksjärvi, M. J. Shipley, A. Singh-Manoux, R. E. Gerszten, T. J. Wang, A. S. Havulinna, P. Würtz, K. Fischer, A. Demirkan, M. A. Ikram, N. Amin, T. Lehtimäki, M. Kähönen, M. Perola, A. Metspalu, A. J. Kangas, P. Soininen, M. Ala-Korpela, R. S. Vasan, M. Kivimäki, C. M. van Duijn, S. Seshadri, and V. Salomaa, "Association of branched-chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: A prospective study in eight cohorts.," *Alzheimers. Dement.*, vol. 14, no. 6, pp. 723–733, Jun. 2018.

[351] Y. Han, T. Bedarida, Y. Ding, B. K. Somba, Q. Lu, Q. Wang, P. Song, and M.-H. Zou, "β-Hydroxybutyrate Prevents Vascular Senescence through hnRNP A1-Mediated Upregulation of Oct4.," *Mol. Cell*, vol. 71, no. 6, p. 1064–1078.e5, Sep. 2018.

[352] Q. Wang, M. van Hoecke, X. N. Tang, H. Lee, Z. Zheng, R. A. Swanson, and M. A. Yenari, "Pyruvate protects against experimental stroke via an anti-inflammatory mechanism.," *Neurobiol. Dis.*, vol. 36, no. 1, pp. 223–31, Oct. 2009.

Tim Dierckx

Bio-informatician, bio-statistician, bio-organism

Heverlee, Belgium

tim.dierckx.td@gmail.com

1985-05-22

www.linkedin.com/in/tim-dierckx-210237a9

Experience

2019-01 – present
Research Associate
Clinical and Epidemiological Virology laboratory, Rega Institute, KU Leuven

2015-01 – 2018-12
IWT funded PhD student
Clinical and Epidemiological Virology laboratory, Rega Institute, KU Leuven

2014 – 2015-01
Research Assistant (student job)
Clinical and Epidemiological Virology laboratory, Rega Institute, KU Leuven

2008 – 2012
Assistant Junior Lab Technician (student job)
Quality Control Laboratory, Eval Europe

Education

2015-05 – 2019
PhD Thesis Project
'Quantifying chronic inflammatory burden from transcriptomes in viral and immune-mediated pathologies'
Clinical and Epidemiological Virology laboratory, Rega Institute, KU Leuven

2014
Master's Thesis
'Host transcriptomic response to IFN: contrasting HTLV-1 mono infection to HTLV-1/HIV-1 dual infection in vitro.'
Clinical and Epidemiological Virology laboratory, Rega Institute, KU Leuven

2012 – 2014
Master in Bio-informatics (KU Leuven, cum laude)

2011
Bachelor's Thesis
'Determining the structural causes of SERCA2b's CA2+ affinity throughout its catalytic cycle.'
Laboratory for Bio Molecular Modelling, KU Leuven

2007 – 2011
Bachelor in Biochemistry & Biotechnology (KU Leuven)

2003-2006
Bachelor in Informatics (KU Leuven)

Languages

- Dutch — Native tongue
- English — Native tongue
- French — Notes
- German — Notes
- R — Native tongue
- Python — Capable
- Shell — Capable
- Java — Notes

Expertise

- Transcriptomics
- RNA-seq
- Microarray
- Machine Learning
- Scripting
- Statistics
- Inflammatory Conditions
- Teaching

Publications

- Subramanian K.*, **Dierckx T.***, Khouri R., Menezes SM., Kagdi H., Taylor GP., Farre L., Bittencourt A., Kataoka K., Ogawa S., Van Weyenbergh J. (2019). Decreased RORC expression and downstream signaling in HTLV-1-associated adult T-cell lymphoma/leukemia uncovers an antiproliferative IL17 link: A potential target for immunotherapy?. Int J Cancer, 144 (7), 1664-1675. doi: 10.1002/ijc.31922.
- Delgobo M., Kozlova E., Rocha EL., Rodrigues-Luis G., Mendes DA G B., Mascarin L., Dias G., Patricio D., **Dierckx T.**, Bicca M., Breton G., Rovaris D., Moral JD., Mansur D., Weyenbergh JV., Bafica A. (2019). Mycobacterium tuberculosis hijacks an evolutionary recent IFN-IL-6-CEBP axis linked to monocyte development and disease severity in humans. doi: 10.1101/514943.
- **Dierckx T.**, Goletti S., Chiche L., Daniel L., Lauwerys B., Jourde-Chiche N., Van Weyenbergh J. (2018). Serum GlycA level is a candidate biomarker for disease activity in systemic lupus erythematosus and for proliferative status of lupus nephritis, independent of renal function impairment. doi: 10.1101/493809.
- Menezes SM., Leal FE., Pereira Ribeiro S., **Dierckx T.**, Roederer M., SenGupta D., Nixon D., Van Weyenbergh J. (2018). Age but not disease progression defines CD4+ and CD8+ T stem cell memory levels in human retroviral infections: contrasting effects of HTLV-1 and HIV-1. doi: 10.1101/473363.
- Fukutani KF., Nascimento-Carvalho CM., Bouzas ML., Oliveira JR., Barral A., **Dierckx T.**, Khouri R., Nakaya HI., Andrade BB., Van Weyenbergh J., De Oliveira CI. (2018). In situ Immune Signatures and Microbial Load at the Nasopharyngeal Interface in Children with Acute Respiratory Infection. FRONTIERS IN MICROBIOLOGY, 9, Art.No. ARTN 2475, doi: 10.3389/fmicb.2018.02475.
- **Dierckx T.***, Verstockt B.*, Vermeire S., Van Weyenbergh J. (2018). GlycA, a nuclear magnetic resonance spectroscopy measure for protein glycosylation, is a viable biomarker for disease activity in IBD. J Crohns Colitis doi: 10.1093/ecco-jcc/jjy162.
- Leal FE., Menezes SM., Costa EA S., Brailey PM., Gama L., Segurado AC., Kallas EG., Nixon DF., **Dierckx T.**, Khouri R., Vercauteren J., Galvao-Castro B., Saraiva Raposo RA., Van Weyenbergh J. (2018). Comprehensive Antiretroviral Restriction Factor Profiling Reveals the Evolutionary Imprint of the ex Vivo and in Vivo IFN-beta Response in HTLV-1-Associated Neuroinflammation. FRONTIERS IN MICROBIOLOGY, 9, Art.No. ARTN 985, doi: 10.3389/fmicb.2018.00985.
- Neggers J., Kwanten B., **Dierckx T.**, Noguchi H., Voet A., Bral L., Minner K., Massant B., Kint N., Delforge M., Vercruysse T., Baloglu E., Senapedis W., Jacquemyn M., Daelemans D. (2018). Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. Nature Communications, 9 (1), Art.No. 10.1038/s41467-017-02349-8.
- Dos Santos PF., Van Weyenbergh J., Delgobo M., Oliveira Patricio DD., Ferguson BJ., Guabiraba R., **Dierckx T.**, Menezes S., Báfica A., Mansur DS. (2018). ISG15-induced IL-10 is a novel anti-inflammatory myeloid axis disrupted during active tuberculosis. Journal of Immunology, 200 (4), Art.No. 10.4049/jimmunol.1701120, 1434-1442.
- Khouri R., Silva-Santos G., **Dierckx T.**, Menezes S., Decanine D., Theys K., Silva AC., Farré L., Bittencourt A., Mangino M., Roederer M., Vandamme A-M., Van Weyenbergh J. (2018). A genetic IFN/STAT1/FAS axis determines CD4 T stem cell memory levels and apoptosis in healthy controls and Adult T-cell Leukemia patients. OncoImmunology, 0 (0), Art.No. e1426423, 0-0. doi: 10.1080/2162402X.2018.1426423.
- Conceição-Neto N., Deboutte W., **Dierckx T.**, Machiels K., Wang J., Yinda KC., Maes P., Van Ranst M., Joossens M., Raes J., Vermeire S., Matthijnssens J. (2017). Low eukaryotic viral richness is associated with faecal microbiota transplantation success in patients with UC. Gut Art.No. gutjnl-2017-315281.
- Vivarini ÁD C., Calegari-Silva TC., Saliba AM., Boaventura VS., França-Costa J., Khouri R., **Dierckx T.**, Dias-Teixeira KL., Fasel N., Barral AM P., Borges VM., Van Weyenbergh J., Lopes UG. (2017). Systems approach reveals nuclear factor erythroid 2-related factor 2/protein kinase R crosstalk in human cutaneous Leishmaniasis. Frontiers in Immunology, 8, Art.No. 10.3389/fimmu.2017.01127.
- Menezes S., Leal FE., **Dierckx T.**, Khouri R., Decanine D., Silva-Santos G., Schnitman SV., Kruschewsky R., Lopez G., Alvarez C., Talledo M., Gotuzzo E., Nixon DF., Vercauteren J., Brassat D., Liblau R., Vandamme A-M., Galvao-Castro B., Van Weyenbergh J. (2017). A Fas(hi) lymphoproliferative phenotype reveals non-apoptotic Fas signaling in HTLV-1-associated neuroinflammation. Frontiers in Immunology, 8, Art.No. 10.3389/fimmu.2017.00097.
- **Dierckx T.**, Khouri R., Menezes S., Decanine D., Farre L., Bittencourt A., Vandamme A-M., Van Weyenbergh J. (2017). IFN-β induces greater antiproliferative and proapoptotic effects and increased p53 signaling compared with IFN-α in PBMCs of Adult T-cell Leukemia/Lymphoma patients. Blood Cancer Journal, 7, Art.No. e519, doi: 10.1038/bcj.2016.126.
- Put K., Vandenhaute J., Avau A., Van Nieuwenhuijze A., Brisse E., **Dierckx T.**, Rutgeerts O., Garcia-Perez J., Toelen J., Waer M., Leclercq G., Goris A., Van Weyenbergh J., Liston A., De Somer L., Wouters C., Matthys P. (2017). Inflammatory gene expression profile and defective IFN-γ and granzyme K in natural killer cells of systemic juvenile idiopathic arthritis patients. Arthritis & Rheumatology, 69, Art.No. 10.1002/art.39933, 213-224.

Marascio N., Pavia G., Strazzulla A., **Dierckx T.**, Cuypers L., Vrancken B., Barreca GS., Mirante T., Malanga D., Oliveira DM., Vandamme A-M., Torti C., Liberto MC., Focà A. (2016). Detection of natural resistance-associated substitutions by ion semiconductor technology in HCV1b positive, direct-acting antiviral agents-naïve patients. International Journal of Molecular Sciences, 17 (9), Art.No. 1416.

Fukutani KF., Nascimento-Carvalho CM., Van der Gucht W., Wollants E., Khouri R., **Dierckx T.**, Van Ranst M., Houspie L., Bouzas ML., Oliveira JR., Barral A., Van Weyenbergh J., De Oliveira CI. (2015). Pathogen transcriptional profile in nasopharyngeal aspirates of children with acute respiratory tract infection. Journal of Clinical Virology, 69, 190-196.