8-2023

Illuminating underappreciated mechanisms of receptor regulation in human lymphocytes.

Cassandra Woolley
University of Louisville

ILLUMINATING UNDERAPPRECIATED MECHANISMS OF RECEPTOR REGULATION IN HUMAN LYMPHOCYTES

By

Cassandra Renee Woolley
B.S., University of Kentucky 2018
M.S., University of Louisville 2021

A Dissertation
Submitted to the faculty of the
School of Medicine of the University of Louisville
In partial fulfillment of the requirements
for the Degree of

Doctor of Philosophy
in Microbiology and Immunology

Department of Microbiology and Immunology
University of Louisville

August 2023

ILLUMINATING UNDERAPPRECIATED MECHANISMS OF RECEPTOR REGULATION IN HUMAN LYMPHOCYTES

By

Cassandra Renee Woolley
B.S., University of Kentucky 2018
M.S., University of Louisville 2021

A Dissertation Approved on

May 24, 2023

by the following Dissertation Committee:

_____
Thomas C. Mitchell, Ph.D. (Dissertation Director)

_____
Nejat Egilmez, Ph.D.

_____
Michele Kosiewicz, Ph.D.

_____
Melissa Smith, Ph.D.

_____
Corey Watson, Ph.D.

DEDICATION

This dissertation is dedicated to my parents,

Mrs. Mechelle Massie and Mr. Jeffrey Massie,

Dr. Steven Woolley and Mrs. Lori Woolley,

And every other member of my family and friends who support me always

and have helped me to achieve all thus far in my career.

I would not be the woman I am today without you who have shaped me.

ACKNOWLEDGEMENTS

I would first like to thank my mentor, **Tom Mitchell**. Had I not met you on the day I interviewed, I am not sure I would have decided to make UofL my home. Your excitement for research combined with your dedication to maintaining a life outside of science has always inspired me. Throughout my training, you have treated me with respect and as an equal while still challenging me to help me grow. You have always shown you truly care for your students as people as well as scientists. I appreciate it endlessly. I would next like to thank **Carolyn Cassella**. You have been a fantastic role model, teacher, and mentor. Your input on my projects has been invaluable, and your advice is always on point. I would also like to thank my lab mates and friends, **Caleb Whitley and Zach VanWinkle**. Whether it be listening to my rants, leaving funny notes at opportune times, or DMing a weekly DND/TTRPG game, you have both helped me maintain my sanity through this process.

Special thanks to my friends and collaborators in the **Smith and Watson Labs,** including **Easton Ford, Kamille Rasche, and Elizabeth Hudson,** as well as both **Eric Rouchka** and **Julia Chariker**. You all welcomed me in my learning of RNA sequencing techniques and bioinformatics, and without your mentorship and patience, my project would not have come to fruition. **Melissa Smith**, you took me under your wing as a budding scientist and have taught and inspired me so much throughout my training. Your criticisms were always constructive and as a mentor, you knew exactly what I needed to hear whenever I would come to you. Thank you.

Endless thanks as well to my other Ph.D. committee members, **Nejat Egilmez, Michele Kosiewicz,** and **Corey Watson.** I appreciate your understanding as I navigated my desires for the trajectory of my career. Throughout this process, your valuable feedback has shaped not only my project but also me as a scientist. I would also like to thank our collaborators, **Swapna Chandran** and **Elizabeth Cash**, who secured tonsil tissue and helped me navigate the clinical side of research.

Last, I would like to thank my family; you are everything to me. I owe a piece of myself to each of you. To my mom **Mechelle Massie** and my dad **Steve Woolley,** you bought me a mini microscope for Christmas in middle school, you put up with my endless questioning in youth, you drove me to academic meets and cheered me on in my achievements, you have always supported me and helped develop my passions. I cannot thank you enough; you are both of whom I strive to be. To my stepdad **Jeff Massie**: you bought me flowers and balloons when I won but also gave me honest truths when needed. You have always been available to fix things or help with a move or a house project. Thank you for constantly being there for me. To my stepmom **Lori Woolley**, my **Granny Thompson**, my **Mamaw Woolley,** my **Aunt Nae Nae,** my **Mamaw Lucy**, my **Mimi Jerry Lowe**, my **Aunt Lynnette**, my sister **Olivia,** my cousin **Keegan**, and all the other incredible women whom I have had in my life to inspire me: thank you for all being amazing role models. To my brothers, **Alex, Chris,** and **Cory**, you showed me there is more to life than academics (like DND and video games). Thank you for helping to keep me grounded. To my partner, **Nick Brinkman**, you have loved and supported me through this all, kept me smiling, and always listened: you are my person, thank you.

ABSTRACT

ILLUMINATING UNDERAPPRECIATED MECHANISMS OF RECEPTOR REGULATION IN HUMAN LYMPHOCYTES

Cassandra R. Woolley

May 24, 2023

A thorough understanding of receptor regulation is imperative to predict expression in varying contexts of disease or treatment. Lymphocyte surface receptors are often used as biomarkers and drug targets, making them particularly important for study. For receptors of debated functionality, such as the Fc receptor for IgM (FcMR), understanding regulation can also help to predict expression *in vivo* to supplement hypotheses of biological roles. Various mechanisms exist for altering receptor surface expression, including direct feedback mechanisms such as ligand-induced endocytosis and broader mechanisms such as transcriptional and translational control. In this dissertation, we explore selected underappreciated mechanisms of lymphocyte receptor regulation. Specifically, we investigate the effects of cell culture conditions on FcMR availability and the potential for global regulation of lymphocyte receptors via isoform variation. FcMR is a constitutively expressed Fc receptor on human T cells, though its function there remains debated. It was previously thought that FcMR was kept low in circulation by FcMR-IgM complex internalization. However, we found that FcMR expression was independent of IgM levels in culture and was higher on direct *ex vivo* stained lymphocytes than in processed PBMC. Instead, increasing cell culture

density inhibited FcMR expression in an apparent cell-contact mediated mechanism, suggesting higher circulating expression of FcMR than previously appreciated and a primary role for FcMR in cell-scarce environments. When next attempting to investigate the potential for isoform-based regulation of FcMR in lymphocytes, we found no applicable isoform-level references. We thus decided to fill this gap using Pacific Biosciences Isoform Sequencing (Iso-Seq) and developed the first Iso-Seq reference transcriptomes of human lymphocytes and activated CD4 T cells. In these references, we discovered many potentially novel transcripts, including end-variant transcripts that only differed from annotated counterparts on their 5' or 3' end. Using plasmids designed to express novel *CXCR5* end-variant isoforms in a HEK293T cell system, we further validated the potential for novel 5' end-variants to affect both mRNA stability and protein expression. The studies presented here provide valuable contributions to the understanding of lymphocyte receptor regulation by positing novel regulatory mechanisms that lay the groundwork for many future studies.

TABLE OF CONTENTS

LIST OF TABLES

PAGE

CHAPTER II

CHAPTER IV

CHAPTER V

LIST OF FIGURES

PAGE

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

CHAPTER I

INTRODUCTION

The human immune system is made up of a diverse set of cells that work in concert to protect our bodies from pathogens. To properly perform this function, these cells must have an array of surface receptors that allow them to sense environmental cues and potential targets. The receptors possessed must also be in constant flux as changes arise that necessitate varying responses. For example, naïve cell receptors that direct lymphocytes to sites of antigen encounter are generally replaced with receptors homing to a site of functional significance once a cell is made aware of a threat (1-7). The exchange of receptors tells a story about a cell on its journey of immune functionality, and to fully understand that story we must also understand the causality and mechanisms of altering receptor cell-surface expression.

Categorization and characterization of immune surface receptors have been long-time goals for scientists, hence the creation of the cluster of differentiation (CD) system and continued efforts to elucidate cell-specific receptor expression and global receptor interplay (8-10). Such an understanding is vital in predicting biomarkers and targets for pharmacotherapies, as it is estimated that over 60% of existing drugs target cell-surface molecules (11). This is particularly important in the case of lymphocytes, which are the most common target of immunomodulatory treatments (12).

Manipulating relevant lymphocyte receptor expression is an attractive target for immune-responsive cancers and immune hyperactivity disorders, such as transplant rejection and autoimmunity (12). However, only a thorough grasp of receptor behavior can serve to predict response and functionality in differing backgrounds of disease or treatment. In the studies presented here, we illuminate selected underappreciated mechanisms of receptor regulation in human lymphocytes with the hope that our findings will add to the knowledge of factors affecting receptor availability in varying immune contexts.

MORE THAN A MARKER: RECEPTORS AND IMMUNE CONTEXT

In the field of immunology, surface receptors are commonly seen as markers used to identify cell types or subsets, as in traditional T helper (Th) or B cell subclassifications (8, 13-16). However, it has recently been realized that phenotypic and functional plasticity is common among cell subsets, and the lines differentiating these subsets are often blurred (17-19).

For example, in CD4 T cell biology, CXCR5 is a marker of T follicular helper (Tfh) cells while CXCR3 is a marker of T helper 1 (Th1) cells (14-16, 20). CXCR5 will guide cells to the lymphoid follicle due to the local expression of its ligand CXCL13 (21, 22), whereas CXCR3 typically directs traffic to inflamed tissue by responding to gradients of CXCL9, 10, and 11 (4, 23). CXCR5 and CXCR3 are commonly considered markers for unique Th cell lineages with differing functions, yet, differentiated $CXCR3^{+}CXCR5^{+}$ cells also exist, called “Tfh1” as a nod to their mixture of Tfh/Th1 phenotypes and functional roles (24-26). Further, in early CD4 T cell activation and

before true differentiation, CXCR5 and CXCR3 are transiently co-expressed in what is hypothesized to be some form of a transitional state, though the true reasoning and mechanism for this expression have not yet been determined (27, 28). Such challenges to the subset dogma highlight the value of separating receptors from their "marker" designation and independently considering what receptor expression and regulation suggest of a cell in varying contexts.

Changes to receptor expression through a cell's life often mirror temporal changes in functionality, matching the fluid need for variation in response to differing ligands. In T cell activation, lymph node homing receptors like CCR7 (CD197) are quickly downregulated while receptors promoting expansion and differentiation such as IL-2Rα (CD25) are upregulated (1, 3, 7, 29). During this time, T-cell co-stimulators such as ICOS and CD40L are also tightly regulated, with the general trend being co-stimulator expression peaking early in activation and fading later (7, 30-33). An opposite trend occurs for co-inhibitors, like PD1 and CTLA-4, following the tidal model of co-signaling that suggests an early need for enhancement of activation but a later need for suppression to prevent overactivation (7, 30-33).

Recognition and characterization of patterns like these give clues for better prediction of novel expression and functionality for other receptors in related situations. Additional clues can come from a deeper understanding of the mechanisms altering receptor display, which vary even between similarly behaving receptors in the same context (30, 31). Particularly for receptors of unknown or debated function, clues like this are important in generating, supporting, or challenging hypotheses that predict context-dependent modifications to receptor expression during the immune response.

Direct feedback: Ligand-directed internalization

One of the most direct methods of receptor regulation is that of ligand-directed internalization. This form of negative feedback is particularly important for receptors where overstimulation in the presence of high levels of ligand would be detrimental to cellular function. Immune receptors whose expression is at least partially regulated by internalization include IL-2Rα, IL-2Rβ, and the common gamma chain (34-36); B and T cell antigen receptors (BCR and TCR, respectively) and partners such as CD3 (37-41); as well as the T cell co-stimulator CD28 (42, 43). In these cases, ligand binding triggers endocytosis of the receptor and a transient decrease in surface levels. This mechanism allows swift changes to modulate responsiveness and maintain receptor fluidity based on cellular needs.

It is ultimately the balance of internalization, degradation, receptor recycling, and new protein synthesis that determine a receptor's overall surface expression levels. For TCR and its CD3 co-receptors, constant internalization and recycling keep surface levels relatively stable in resting T cells (38, 41). Upon stimulation by cognate antigen, internalization and degradation begin to outbalance recycling and protein synthesis, and only then does surface expression decrease (38-41, 44).

Pharmacotherapeutics can take advantage of this process to target downregulation and inhibition of receptors mediating disease, as is one proposed mechanism of the recently approved drug Teplizumab which targets CD3 to induce tolerance in type I diabetes (45-47). Therapies may also take advantage of activation-induced internalization of receptors to target medications to the intracellular compartment of cells. Zynlonta® is

one such antibody-drug conjugate that targets CD19 in B cell lymphoma to deliver cytotoxic therapy (48, 49). In developing such therapeutics, it is important to understand how specific interactions might affect the continued expression of a target receptor over the course of treatment.

Interestingly, ligand-mediated internalization does not always force a decrease in overall surface levels for a receptor. This seems to be the case in the regulation of the Fc receptor for IgM (FcMR). FcMR is a potential drug target expressed on chronic lymphocytic leukemia B cells that is also constitutively expressed on human B, T, and NK cells (50-55). FcMR is known to internalize upon binding to multimeric human IgM in these cells (50-53). Yet, we have found that physiologic levels of IgM do not modulate FcMR expression on lymphocytes in cell culture (Chapter II) (56). This lack of modulation is important to consider when anticipating the levels of FcMR on both healthy and target cells throughout FcMR-targeted therapeutics. Additionally, in the normal immune response of healthy cells, additional mechanisms of FcMR regulation may be more pertinent for the appropriate prediction of situational expression and function.

Even when relevant, direct ligand-mediated receptor modulation often works in concert with other mechanisms of regulation to ensure a desired outcome. In the case of CD28, which is transiently downregulated by ligand binding, both direct internalization and signaling-directed transcriptional changes contribute to the observed decrease in surface expression levels (42). Unlike receptor internalization, which is primarily mediated by ligand presence, transcriptional regulation can simultaneously account for many input signals that independently affect messenger RNA (mRNA) and therefore

protein expression. Thus, in response to a combination of environmental signals, a specific level of mRNA is made available for protein synthesis, contributing to highly specific and context-dependent levels of receptor availability.

Transcriptional regulation

Deep sequencing technologies used to sequence mRNA transcripts have illuminated many examples of differential mRNA expression between immune cell states and during disease pathology (57-60), revealing dynamic regulation of gene expression in the immune response. Transcriptional regulation may occur at any one of the many steps required for successful mRNA synthesis.

First, for initiation of transcription, the target DNA must be liberated from dense chromatin to allow appropriate access by RNA polymerase and other transcriptional proteins. The varying needs of differential gene expression can force changes in this DNA accessibility through epigenetic alteration. These alterations occur "above" the level of the nucleotide sequence of the genome and include histone modifications, which influence accessibility by altering how tightly an area of chromatin is wound (61, 62), as well as DNA methylation, which physically inhibits the binding of transcription-factors (62, 63). Epigenetic modifications permit cells to either restrict the expression of undesired genes or to rapidly express functionally dynamic proteins, and epigenetic changes may affect receptor expression and subsequently receptor surface levels. In non-follicular T cells, methylation in cysteine-guanine dinucleotide (CpG) motifs near the transcription start site of *CXCR5* normally inhibit transcription at this locus (64-67). Inappropriate demethylation or hypermethylation leads to aberrant transcription in

diseases with over- or under-expression of cell-surface CXCR5, respectively (64-67), highlighting the importance of DNA accessibility to allow adequate transcriptional control of receptor expression.

An accessible DNA profile not only allows for RNA polymerases to bind the promoter region and begin transcription but also allows effective binding of associated transcription factors that independently serve to enhance or inhibit mRNA expression. The activity and availability of transcription factors coordinating gene expression are tightly regulated; most transcription factors are not specific to a single gene but are instead broad regulators of a group of genes needed in a specific state of cellular function. Further, there are often multiple transcription factors per locus that differentially interact to regulate a gene's expression (68).

In the type I interferon response, for instance, a transcriptional regulatory complex of interferon regulatory factor 9 (IRF9) and STAT1/STAT2 is an important mediator in the upregulation of multiple genes whose loci contain an interferon-stimulated response element (69-73). However, in the type II interferon response, STAT1 predominantly forms a homodimer that allows it to preferentially bind and upregulate genes containing gamma-activated sequences (70-73). Differential modes of interaction among transcription factors allow cells to tailor their transcriptional responses to the specific extrinsic signals received.

There are also cases of single transcription factors that are sufficient to mediate multi-gene responses. When these single transcription factors coordinate a response required for a specific differentiation program, the factor might be considered a "master regulatory transcription factor". For example, Pax5 is considered a B cell-specific master

regulator as it is required for lineage commitment and subsequent function in B cells (74-76). Each subset of T helper cell also expresses a lineage-specific master regulator, such as Bcl6 (Tfh), T-bet (Th1), GATA3 (Th2), and RORγT (Th17) (20, 68, 77-80). Each of these transcription factors is tightly associated with the expression of surface receptors that define each T helper subset. Most of these transcription factors regulate subset-specific receptor expression through either direct binding in or near the transcription start site, as in T-bet binding to the promoter of *CXCR3* in Th1 cells (81, 82), or through indirect mechanisms of enhancement, as in Bcl6 regulation of *CXCR5* in Tfh through disinhibition of E2A binding to enhancer elements (77-79, 83, 84).

However, the dogma of "master transcription regulators" has also come under scrutiny, particularly for T cell subset differentiation (85-87). While "master regulators" hold importance in key aspects of receptor expression during cellular differentiation, they may not be as singularly vital as initially thought. For example, stable expression of CXCR5 on differentiated Tfh and Tfh1 requires Bcl6 (77-79), but transient expression of CXCR5 in early T cell activation is thought to be independent of this master transcription regulator and may be dependent on a second transcription factor, Ascl2 (88, 89).

Translational regulation

While transcription factors often predict the expression profiles of cell-surface and other proteins, they are not always defining determinants. Similarly, directly measuring the transcript abundance of genes of interest gives an incomplete picture of their actual protein expression patterns. It is estimated that just 40% of the variation in protein expression can be explained by a corresponding variation in transcript expression

(90), suggesting that the majority of protein expression is regulated after mRNA synthesis. Thus, we must also consider regulation at post-transcriptional levels in seeking to understand changes to receptor expression during the immune response and how to best develop therapeutics targeting them.

A primary cause of the imperfect correlation between mRNA levels and protein expression is the presence of intrinsic sequences within mRNA transcripts that can affect the amount of protein synthesized (90-93). These sequences may regulate protein expression through alterations in translational efficiency or through alterations in mRNA stability that determine the amount of time a transcript is available for translation. For any given gene, multiple forms of its mRNA may possess different kinds or amounts of these regulatory elements. Distinct forms of protein-invariant mRNA under differential regulation would be classified as types of mRNA isoforms (94-96).

Isoforms are variants of mRNA transcribed at the same gene locus, commonly generated by alternative splicing or through alteration in transcription start or termination sites (94-96). Though isoforms are most studied in contexts where differential protein variants are expressed, mRNA isoforms may also harbor differences only in noncoding regulatory regions of the mRNA. For a mature mRNA, these regulatory regions are the 5' and 3' untranslated regions (UTR), which flank the coding sequence on either end.

Within UTRs, elements such as differential secondary structure, sequence motifs bound by RNA binding proteins (RBPs) or micro-RNAs (miRNAs), or sequences triggering ribosomal stall, will ultimately serve to adjust the amount and efficiency of translation for a specific transcript (92, 96-98, 99, comprehensive review in Chapter III). The influence of these regulatory elements may be changed based on the state-dependent

presence of their interacting factors, such as miRNA and RBP, as well as altered availability of ribosomal and initiation factors (100, 101). Small differences within these elements can lead to large differences in protein expression and may contribute to clinically relevant differences in responses, such as in the NLRP3 29940 G to C polymorphism in the 3' UTR which is associated with resistance to septic shock (102). This is thought to occur because the G to C polymorphic form of the mRNA contains a binding site for a miRNA that increases degradation and leads to inhibition of expression of NLRP3, resulting in less downstream inflammation during sepsis (102).

Translation-altering variation in mRNA could play an important role in receptor regulation. However, compared to our knowledge of changes to mRNA abundance, the range and impact of differential expression of mRNA regulatory isoforms in human immune cells remain relatively understudied (90-93). Though methods are improving, this is in part due to the limitations of popularized RNA-Seq analyses, which generally lack the capacity to accurately annotate isoforms because of inherent algorithmic challenges of reconstructing *de novo* isoforms from only short sequencing reads (103-105). Therefore, to gain a full picture of receptor regulation, we must achieve an "isoform-aware" understanding of the human immune transcriptome that relies on long, contiguous reads of intact full-length mRNA transcripts.

DISSERTATION GOALS AND OVERVIEW

In this dissertation, we explore underappreciated mechanisms of receptor regulation in human lymphocytes with the goal of using knowledge gained for better prediction of receptor expression and functionality.

We first investigate the role of the ligand IgM in regulating cell-surface expression of the only constitutively expressed Fc receptor on human T cells, FcMR, whose function remains relatively unknown in these cells (106, 107) (Chapter II). Elucidating FcMR expression patterns is important to lend clues to physiologic contexts where this receptor may be most functional and predict responses to pharmacotherapeutics targeting the receptor.

Next, we describe a new and developing sequencing technology, Iso-Seq, that relies on single-molecule real-time sequencing for the unambiguous determination of intact mRNA transcript sequences. We report the first isoform-aware transcriptomes, using Iso-Seq, of human circulating lymphocytes and activated CD4 T cells, laying the groundwork for additional studies of isoforms in these cells (Chapters III-V). We further report the discovery of novel 5' end-variant isoforms in transcripts of immune-important receptors expressed only during CD4 T cell activation, indicating the existence of an underappreciated layer of transcriptional and translational regulation (Chapter V). Finally, we suggest that an isoform-aware understanding of transcriptomic changes is vital to gaining a more complete understanding of the regulation of receptor protein expression during the immune response.

The studies reported here provide novel insights into lymphocyte surface receptor biology, cataloging knowledge and variation important for better predicting the context-dependent behavior of lymphocyte receptors for their use as biomarkers and as drug targets (11, 12).

CHAPTER II

AN UNEXPECTED ROLE FOR CELL DENSITY RATHER THAN IGM IN CELL-SURFACE DISPLAY OF THE FC RECEPTOR FOR IGM (FCMR)[1]

INTRODUCTION

The Fc receptor for IgM (FcμR or FcMR) was identified in 2009 and remains the only known IgM-exclusive Fc receptor (106, 107). Cell-surface FcMR has been confirmed on human B, T, and NK cells, and on mouse B cells (106-110) with debated expression in myeloid lineages of both humans and mice (108, 111-116). Although FcMR is the only FcR constitutively expressed by human T cells (108), its role in these cells remains unclear. Indeed, many aspects of FcMR function remain to be elucidated. Because regulation of expression often hints at where and when a protein may be functionally relevant, we decided to characterize patterns of FcMR expression in human lymphocyte populations at the level of cell surface display.

FcMR surface abundance is regulated by cellular activation status and composition of the surrounding tissue milieu (106, 108, 110). At baseline, lymphocyte cell-surface FcMR is reported to be low, but detectable, in blood and peripheral lymphoid organs (53, 106, 108). One hypothesized contributor to this low baseline surface level is the presence of the FcMR ligand, IgM (106, 108, 110), supported by findings that FcMR

[1] Woolley, C. R., N. C. Brinkman, E. D. Cash, S. K. Chandran, and T. C. Mitchell. 2022. An Unexpected Role for Cell Density Rather than IgM in Cell-Surface Display of the Fc Receptor for IgM on Human Lymphocytes. *ImmunoHorizons* 6: 47-63. doi: 10.4049/immunohorizons.2100094.

internalizes after binding to multimeric IgM (50-53). Indeed, pre-incubation of cells in IgM-deficient media has been recommended to raise surface FcMR to levels adequate for study, particularly for human T cells (53, 106, 110). Studies of primary lymphocytes allowed to recover FcMR in this manner have led to important insights whose recurring theme seems to be that FcMR plays markedly different roles in B versus T cells. For example, B cell activation increases surface FcMR, whereas T cell activation decreases it (53, 106). Further, FcMR limits the tonic BCR signaling by reducing transport of IgM-BCR to the surface of mouse B cells *in vivo* (50), whereas in cultures of primary human T cells, FcMR engagement by exogenous IgM resulted in increased transport of the TCR signaling complex to the cell surface (53).

Hypotheses regarding the function of FcMR are more advanced for B cells relative to T cells, presumably because its expression in T cells is species-specific and thus cannot be evaluated in *Fcmr*$^{-/-}$ mice. From mouse models, a consensus appears to be forming around a B cell 'rheostat' hypothesis (117) in which FcMR-mediated effects on BCR signal strength enhance detection of self-antigens by immature B cells during development as well as of foreign antigens by mature B cells in secondary lymphoid organs. By contrast, hypotheses to explain why FcMR is expressed by human T cells are far less comprehensive and currently center around the tonic effect, noted above, that FcMR internalization was observed to have on the TCR complex and co-stimulatory molecules. In this model, naïve T cells are envisioned as encountering abundant IgM upon entry into lymph nodes or spleen that will internalize as complexes with FcMR, resulting in enhanced surface expression of the TCR complex and co-stimulatory molecules, thus preparing the cells for cognate interactions (53). Less consideration has

been given to potential functions of FcMR in T cell populations outside of lymphoid organs, perhaps because the increase in surface display of FcMR observed after culturing PBMC in IgM-deficient media implies low abundance prior to harvesting the cells from whole blood, i.e., while in circulation.

Changes in FcMR expression by circulating lymphocytes observed in some disease states such as chronic lymphocytic leukemia (CLL) may provide clues about its function. In CLL patients, FcMR is elevated compared to healthy counterparts not only on leukemic B cells, but also on non-leukemic B and T cells (108, 118). Abnormally elevated FcMR expression by leukemic cells is thought to be due to antigen-independent cross-linking of BCR *in cis*, which is ultimately mitogenic (106, 108). In the same patients, the mechanism by which FcMR expression on non-leukemic B and T cells is unclear, but may reflect the fact that CLL patients commonly experience a global deficiency in serum IgM such that comparatively less IgM is available to drive internalization of FcMR (108, 118). However, in patients with selective IgM deficiency, surface display of FcMR is unchanged on most lymphocyte subsets (and is actually decreased on circulating naïve marginal zone B cells) relative to healthy controls (119). The lack of a consistent relationship between IgM abundance and FcMR surface expression in these two disease states suggests a need for further characterization of the regulatory mechanisms that determine when and where FcMR is available for functional interactions.

In our study, we confirmed that FcMR surface levels are upregulated on peripheral blood lymphocytes cultured in IgM-deficient media but also found that, surprisingly, FcMR upregulation was equally robust after culture in the presence of

human serum, which contains an average of 1.5 mg/mL IgM (120), indicating IgM exposure has little effect on steady-state surface FcMR expression. Cell-surface FcMR was instead strongly affected by culture at higher cell densities. Downregulation of FcMR independent of IgM abundance occurred through a mechanism requiring close cell-cell proximity that does not appear to require the presence of a particular cell type or soluble factor and thus remains undefined. We also found that *ex vivo* processing of whole blood decreases surface expression of FcMR, implying circulating lymphocytes express it at significantly higher levels than previously believed despite continuous exposure to IgM. Collectively, our findings suggest that the physiological environments in which FcMR is available for functional interactions, especially for T cells, are different than previously thought, which has implications for the role FcMR may play in the human immune response.

MATERIALS AND METHODS

Blood Sampling and Peripheral Blood Mononuclear Cell Preparation

Between 15 and 300 mL of venous blood was collected by standard phlebotomy techniques from consented donors who reported themselves to be healthy at the time of the blood draw. Blood collection was approved by the University of Louisville Institutional Review Board under expedited review. Donors were between 20 and 60 years of age, 25% female and 75% male, with some donors repeated across different experiments such that 28% of experiments were performed with female cells.

Blood was collected in a total of 6 mM K_3EDTA as an anticoagulant. In some experiments, blood cells were stained under conditions of minimal manipulation in which

flow cytometric antibodies were added immediately before or after lysing red blood cells (RBCs) prior to flow cytometry, as described below. In all other experiments, peripheral blood mononuclear cells (PBMCs) were isolated using Sepmate™ PBMC Isolation Tubes (StemCell Technologies, 85450) as directed by the manufacturer. Briefly, fresh anticoagulated blood was diluted 1:1 with PBS lacking calcium and magnesium ($PBS^{-/-}$, Thermo Fisher Scientific, 10010-023) and then layered into Sepmate™ tubes preloaded with 15mL of Lymphoprep™ Density Gradient Medium (StemCell Technologies, 07801). Tubes were centrifuged at 1200 x g for 10 minutes at room temperature (RT). After centrifugation, PBMCs were poured off into fresh 50 mL tubes and washed twice with $PBS^{-/-}$ by centrifugation first at 600 x g for 10 minutes then at 300 x g for 10 minutes. Total cell yield was determined using the count per μL feature of a Cytek® Northern Lights flow cytometer. Counts were performed by generating a 1:10 dilution of the original cell suspension in 200 μL $PBS^{-/-}$, running the diluted suspension on the flow cytometer, and using the count per μL of events in a cell gate generated on an FSC/SSC plot in SpectroFlo® for subsequent calculations. PBMCs were used immediately for flow cytometric staining or culture procedures or were further manipulated to purify or deplete blood cell subsets.

Cell Type Depletions and Isolations

Cell depletions were performed via immunomagnetic selection using EasySep™ kits with the EasyEights™ Magnet (StemCell Technologies, 18103) and the manufacturer's recommended medium ($PBS^{-/-}$ supplemented with 2% FBS and 1 mM K_3EDTA). For depletion of NK, B, or CD8 T cells, EasySep™ Biotin Positive Selection

Kit II (StemCell Technologies, 17683) was used in combination with anti-CD56 biotinylated antibody, anti-CD19 plus anti-CD20 biotinylated antibodies, or anti-CD8 biotinylated antibody, respectively (**Table 2.1)**. For $CD14^+$ monocyte depletion, EasySep™ Human CD14 Positive Selection Kit II (StemCell Technologies, 17858) was used following a modified protocol provided by the manufacturer (121). Briefly, after traditional addition of reagents per manufacturer's instructions, dwell time in the EasyEights™ magnet was doubled and then unbound cells were transferred to fresh tubes for a second round of depletion before final collection. Mock depletions were performed in parallel by adding isotype pre-matched control antibodies (NK, B, and CD8 T depletions) or without addition of isolation cocktail (CD14+ monocyte depletion) as the cocktail composition is proprietary.

For platelet depletion, PBMCs were first isolated using a modified Sepmate™ protocol in which the top platelet enriched fraction was pipetted off and discarded after density gradient separation before pouring the PBMC fraction into a new tube. Two wash steps were then performed with centrifugation at reduced speed to preferentially pellet nucleated cells (120 x g 10 minutes at RT). Platelets were further depleted using magnetic separation in the EasyEights™ Magnet with EasySep™ Human Platelet Removal Cocktail (StemCell Technologies, 19369C component of 19359). Mock platelet depletion was performed in parallel by processing PBMC without removal of the top fraction after density gradient separation, without low-speed wash steps, as well as without addition of the platelet depletion cocktail in subsequent steps.

Table 2.1. Staining and Depletion Antibodies

Target Specificity	Fluorochrome/ Conjugate	Clone	Company	Catalog #
STAINING ANTIBODIES				
FcMR	BV421	HM14	BD Biosciences	564714
KLH (Isotype ctrl for clone HM14)	BV421	X40	BD Biosciences	562438
FcMR	APC	HM7	Biolegend	398104
Unknown (Isotype ctrl for clone HM7)	APC	MPC-11	Biolegend	400322
FcMR	Unlabeled (primary)	Rabbit Polyclonal	Sigma-Aldrich	HPA003910
Rabbit IgG Fc	FcMR (secondary)	Goat Polyclonal	BD Biosciences	565014
CD4	BV570	RPA-T4	Biolegend	300534
CD8a	Super Bright 645	OK-T8	Thermo Fisher Scientific	64-0086-42
CD8a	Super Bright 645	RPA-T8	Thermo Fisher Scientific	64-0088-42
CD8a	BB700	RPA-T8	BD Biosciences	566452
CD3	Alexa Fluor 532	UCHT1	Thermo Fisher Scientific	58-0038-42
CD56	BV786	NCAM16.2	BD Biosciences	564058
CD19	eFluor 506	HB19	Thermo Fisher Scientific	69-0199-42
CD20	Alex Fluor 700	2H7	Thermo Fisher Scientific	56-0209-42
CD185 (CXCR5)	PerCP-eFluor 710	MU5UBEE	Thermo Fisher Scientific	46-9185-42
CD279 (PD-1)	APC	J105	Thermo Fisher Scientific	17-2799-42
CD41a (ITGA2B)	BB515	HIP8	BD Biosciences	565938
Fixable Viability Dye	APC-eFluor 780	-	Thermo Fisher Scientific	65-0865-14
DEPLETION ANTIBODIES				
MOPC (Isotype ctrl for biotinylated abs)	Biotin	MOPC-21	StemCell	60070BT
CD56	Biotin	NCAM	Thermo Fisher Scientific	13-0567-82
CD19	Biotin	HIB-19	StemCell	60005BT
CD20	Biotin	2H7	StemCell	60008BT
CD8a	Biotin	RPA-T8	StemCell	60022BT

Tonsil Collection and Tonsil Mononuclear Cell Preparation

Tonsil tissue was collected from pediatric patients at Norton Children's Hospital undergoing tonsillectomy performed by Swapna Chandran, M.D. Prior to tonsillectomy, the patient's legal guardian signed an informed consent following IRB ethical guidelines. All patients between ages 2-18 presenting for tonsillectomy with or without adenoidectomy were eligible. Consecutive patients were invited into the study. Of the four donors used for this study, three were male and one was female. All patients were in the 4-11-year age range and undergoing tonsillectomy for management of sleep-disordered breathing with no other immune system-impacting comorbidities. Immediately following tonsillectomy, one half of each of the right and left tonsil were placed into a cold solution of 35 mL tonsil buffer made with 1mM K_3EDTA, 100 U/mL penicillin with 100 μg/mL streptomycin (Thermo Fisher Scientific, 15140-122), 5 μg/mL gentamicin (Thermo Fisher Scientific, 15710-064), and 0.5 μg/mL amphotericin B (Sigma-Aldrich, A2942) in $PBS^{-/-}$ for storage on ice up to 4 hours prior to processing.

Tonsils were processed to obtain tonsil mononuclear cells (TMCs) using an optimized version of a previously published protocol (122). Tonsil tissue was minced in a sterile 100 mL petri dish while being kept wet with tonsil buffer solution. Once 1-3 mm fragments were obtained, the tissue was then transferred to a metal cell strainer sitting in a fresh 100 mL petri dish with additional tonsil buffer solution. Tissue was gently pushed through the strainer using the plunger of a 5 mL syringe. The tissue in the cell strainer was periodically washed with tonsil buffer to ensure it stayed wet and to encourage release of cells. The resulting cell suspension was then passed through 40 μm plastic cell strainers into fresh 50 mL tubes and pelleted by centrifugation at 600 x g for 10 minutes

at RT. The pellet was resuspended in 20 mL fresh tonsil buffer and split such that each of two 10 mL cell suspensions was layered on top of 25 mL Lymphoprep™ Density Gradient Medium (StemCell Technologies, 07801) in a 50 mL tube. This density gradient suspension was centrifuged at 800 x g for 20 minutes at RT with the brake off. Mononuclear cell layers at the resulting interfaces were pipetted into new 50 mL tubes and washed twice using tonsil buffer solution by RT centrifugation at 600 x g for 10 minutes. Cells were pooled by donor and mononuclear cell yield was determined using the count per μL feature of a Cytek® Northern Lights flow cytometer as described above. TMCs were used immediately for flow cytometric staining or for culture procedures.

Fresh Serum Collection

For experiments requiring the use of fresh serum, venous blood (5 to 15 mL) was collected in BD Vacutainer® SST™ serum separation tubes (BD Biosciences, 368013) at the time of blood collection for PBMCs from either the same blood donor (autologous) or a separate donor (non-autologous). Serum was isolated according to manufacturer's protocol in which collected blood was mixed with clotting agents by inversion of the tubes and incubation at RT for 30 minutes before centrifugation at 1200 x g for 10 minutes at RT. Serum was collected from above the polymer gel plug. An additional centrifugation at 1200 x g for 10 minutes at RT was used as necessary to remove any remaining RBC.

Cell Culture

For all experiments, unless otherwise indicated, PBMC or TMC were cultured in suspension of 2.5 million cells/mL plated at 200 μL/well in 96-well Falcon® U-bottom tissue culture-treated plates (Corning, 353072). In most experiments, PBMCs or derivatives were cultured in complete RPMI generated using RPMI 1640 media (Thermo Fisher Scientific, 21870-076) with the addition of 1X Glutamax (Thermo Fisher Scientific, 35050-061), 100 U/mL penicillin and 100 μg/mL streptomycin (Thermo Fisher Scientific, 15140-122), and 10% sterile-filtered, heat-inactivated male AB serum (Sigma-Aldrich, H3667). TMCs were cultured in complete RPMI supplemented with 5 μg/mL gentamicin (Thermo Fisher Scientific, 15710-064) and 0.5 μg/mL amphotericin B (Sigma-Aldrich, A2942).

For serum dilution experiments, complete RPMI was supplemented with 0, 10, 20, 40, or 70% human serum. For 100% human serum cultures, only antibiotics and Glutamax were added at the same concentrations as in complete RPMI. Human serum in these experiments was either fresh and autologous/non-autologous, collected as described above, or sterile-filtered, heat-inactivated male AB serum (Sigma-Aldrich, H3667). IgM concentrations in these sera ranged from 0.36 to 2.3 mg/mL, as determined by ELISA. For density dilution experiments, cells were cultured at densities of 0.2, 0.5, 1, 1.5, 2, or 4 million cells/well in 200 μL of complete RPMI. In some experiments, Falcon® flat-bottom tissue culture-treated plates (Corning, 353077) were used alongside U-bottom plates for these cultures.

For transwell culture experiments, HTS Transwell ® (Corning, 3388) were used with a total of 200 μL per well. The bottom "receiver" wells were first loaded with 100

μL of either complete RPMI alone or containing 0.2 or 2 million cells After loading the bottom wells, the top "insert" wells were loaded with 100 μL of complete RPMI containing either 0.2 or 2 million cells to yield a total of 200 μL/well with top and bottom cell amounts as specified in **Figure 2.10b**.

Transwell plate wells are flat-bottomed, so a muted density-dependent effect in bottom wells would be expected based on our results. However, in some experiments cell density effects for the cells in the transwell bottom wells were muted compared even to what we saw in Falcon® flat-bottom plates. As controls, FcMR expression on cells plated in the bottom wells were assessed to determine if the expected density-driven effect had occurred, and thus if the results could be meaningfully compared to those observed when using other plate formats. These controls were assessed by calculating the ratio between FcMR ΔMFI measured for bottom well 0.2 million and 2 million cell densities and comparing this ratio to the range of ratios observed in four experiments performed in traditional Falcon® flat-bottom tissue culture-treated plates (Corning, 353077). In two experiments with transwell plates we found that ratios were well outside this range. Failure of these positive controls led to the exclusion of one experiment for all subsets, and exclusion in one other experiment for the B and NK cell subsets only.

To test for the presence of secreted factors, cell-free culture supernatants (SUPs) were collected after 24-hour (24hr) cultures. SUPs were generated by centrifuging culture plates at 860 x g for 3 minutes at RT**,** transferring media supernatants to 1.5 mL Eppendorf tubes, and storing at -80^{o} C until use in subsequent experiments. Immediately prior to use, SUP tubes were centrifuged 10,000 x g for 3 minutes at RT to remove cell debris. Cells tested for responses to potential soluble factors were either cultured in 200

μL fresh complete RPMI with 10% human serum alone (No SUP) or with 100 μL complete RPMI mixed with 100 μL of culture supernatants (50% SUP, collected from cultures of 0.2 or 2 million cells/well).

All cells were analyzed after 24hr culture or, for time course experiments, after 0.5, 1, 2, 4, 6, 12, and 24hr culture in a standard 5% CO_2 humidified incubator (Thermo Fisher Scientific, HERAcell 150i CO_2 incubator). Tests of the effects of oxygen tension were performed by culturing cells in a standard incubator (18-20% O_2) in parallel to cells cultured at approximate physioxia (5% O2; 123, 124) in a humidified, 5% CO_2, nitrogen-controlled incubator (Sanyo O_2/CO_2 incubator, MCO-5M).

PBMC Flow Cytometric Staining and Analysis

Cells were stained for flow cytometric analysis in the same 96-well U or flat-bottom plates in which they had been cultured. 0hr measurements were performed by mock plating cells in 96-well U-bottom plates. Cells in these 96-well plates were pelleted and then washed twice with $PBS^{-/-}$ prior to viability staining. All pelleting and wash steps were performed by 4ºC centrifugation at 860 x g for 3 minutes. To assess viability, cells were resuspended in 100 μL of a viability stain containing eBioscience™ Fixable Viability Dye (Thermo Fisher Scientific, 65-0865-14) in $PBS^{-/-}$ lacking added serum or protein. Cells plated at 0.2 or 0.5 million cells per well were resuspended directly in 100 μL of viability stain while cells plated at higher densities were first split to achieve a uniform density of 0.5 million cells per 100 μL stain. After addition of viability stain, cells were incubated 20-30 minutes at 4º C, washed once with $PBS^{-/-}$, and washed again with stain buffer ($PBS^{-/-}$ containing 0.09% sodium azide and 2% human serum to prevent

binding of fluorescent antibodies to FcR (125)). After the second wash, cells were resuspended in 100 μL of stain master mix. Stain master mix was generated by first mixing staining antibodies (**Table 2.1**) in BD Horizon^{TM} Brilliant Stain Buffer (BD Biosciences, 566385 or 563794) according to manufacturer's instructions, then adding standard stain buffer to reach a cumulative 100 μL. Cells were incubated in this stain master mix for 20-30 minutes at 4^{o} C, washed twice with standard stain buffer, and resuspended in cold stain buffer containing 1% formaldehyde for fixation prior to transfer to 12 x 75 mm polystyrene tubes for flow cytometric analysis.

For flow cytometry with rabbit polyclonal anti-FcMR (RpAb), cells were stained in a buffer containing TruStain FcX (Biolegend, 422302) diluted in PBS$^{-/-}$ with 2% FBS and 0.09% sodium azide according to the manufacturer's instructions. After primary incubation with RpAb, two washes with stain buffer were performed prior to secondary incubation with an antibody master mix supplemented with BV421 labelled goat anti-rabbit staining antibody, as described above.

Flow cytometry was performed with a Cytek® Northern Lights 3-laser flow cytometer. Spectral profiles were unmixed accounting for autofluorescence using SpectroFlo® software and appropriate single-stain and unstained controls. Processed data files were then analyzed in FlowJo (BD Biosciences), with additional fluorescence compensation performed using FlowJo compensation matrices if needed. Gating strategies used to identify cell populations are shown in **Figure 2.1.** FcMR stain mean fluorescence intensity (MFI) of gated cell populations were determined using FlowJo and corrected for background to calculate ΔMFI as FcMR MFI of each technical replicate minus the average MFI of all isotype control replicates in the same experiment. For the

rabbit polyclonal anti-FcMR stains, ΔMFI was calculated as FcMR MFI of each experimental replicate minus the average MFI of all fluorescence-minus-one control replicates in the same experiment. If the calculated ΔMFI was negative, indicating a replicate MFI was below that of the average isotype for the experiment, its ΔMFI was recorded as zero.

Fresh Blood RBC Lysis and Staining

For experiments involving direct staining of leukocytes without using density gradients for isolation of PBMC, BD Pharm Lyse™ (BD Biosciences, 555899) solution at 1X concentration was used to lyse red blood cells (RBCs). Three variations of the manufacturer's protocol were performed, all starting with whole blood supplemented with 6 mM K_3EDTA as anticoagulant and performed at room temperature unless otherwise indicated:

1) "Stain then lyse": 100 μL freshly drawn whole blood collected was added to 12 x 75 mm polystyrene flow cytometry tubes. Staining was performed by adding 50 μL of a stain master mix containing staining antibodies (**Table 2.1**) in BD Horizon™ Brilliant Stain Buffer (BD Biosciences, 563794). Cells were incubated for 15-30 minutes at 4ºC in the dark. After incubation, 2mL of 1X BD Pharm Lyse™ was added, the tubes were gently vortexed, and then incubated for 10 minutes in the dark at RT for RBC lysis. Tubes were centrifuged at 300 x g for 5 minutes at RT, washed once with 3mL of stain buffer, and then resuspended in $PBS^{-/-}$ containing 1% formaldehyde for fixation. Cells were kept on ice prior to flow cytometric analysis.

2) “Lyse 1x then stain”: 1 mL whole blood was added to 15mL polypropylene tubes. RBC lysis was performed by adding 10mL 1X BD Pharm Lyse™, gently vortexing the mixture, and incubating for 15 minutes. After incubation, tubes were centrifuged at 300 x g for 5 minutes. The cell pellet was resuspended in 2 mL staining buffer, split into a 96-well U-bottom plate at 200 μL/well, and centrifuged at 860 x g for 3 minutes at 4° C. Cells were stained in 96-well plates as described above for PBMC but without viability staining. After staining, cells were resuspended in a 1% formaldehyde solution for fixation prior to transfer into 12 x 75 mm polystyrene flow tubes. Cells were kept on ice prior to flow cytometric analysis.

3) “Lyse 2x then stain”: Initial RBC lysis and cell centrifugation was performed as described above. The cell pellet after first lysis was resuspended in 1 mL stain buffer and RBC lysis was repeated by adding 10 mL 1X BD Pharm Lyse™, gently vortexing the mixture, and incubating for 15 minutes. After incubation, tubes were centrifuged at 300 x g for 5 minutes. This pellet was resuspended in 2 mL staining buffer, split into a 96 well U-bottom plate at 200 μL/well, and stained then fixed for flow cytometric analysis as described for the “lyse 1x then stain” procedure.

IgM ELISAs

Serum IgM concentrations were measured using the HRP/TMB based Human IgM ELISA Kit (Thermo Fisher Scientific, 88-50620) following the manufacturer’s protocol. All sera were tested in triplicate at dilutions of 1:4,000, 1:8,000, and 1:16,000. Measurements were taken using “Emax Precision Microplate Reader” (Molecular Devices) at 450nM wavelength. Triplicate values from 1:4,000 dilutions, which

consistently had the lowest coefficient of variability compared to other dilutions, were averaged, and used to calculate serum IgM content in mg/mL. Values plotted in **Fig. 2.4C** show final IgM concentrations in culture media containing 10% human serum.

Statistics

Statistical significance was assessed using GraphPad Prism software version 9.2.0 with the test specified in each figure legend. Normality was tested with the same software by assessing linear fit of the data set to a Normal Q-Q plot generated using GraphPad Prism. For tonsil data, values for technical replicates outside of 3 standard deviations of the data set mean were deemed outliers and excluded from figures and calculations. This led to the exclusion of one technical replicate in one experiment.

RESULTS

FcMR levels on the surface of blood lymphocytes increase after 24-hour culture.

Previous studies reported that overnight culture in IgM-deficient media increased FcMR surface levels on human B, T, and NK cells (53, 106). We confirmed this pattern using flow cytometry to measure surface FcMR on peripheral blood B, T, and NK cells that were stained after processing whole blood into PBMC and then again after culture. Using a BV421-conjugated anti-FcMR mAb (clone HM14), surface FcMR was found to be increased after 24-hour culture in serum-free media for CD4 T cells, CD8 T cells, B cells, and NK cells (**Figure 2.2a-d**). No sex-specific differences in FcMR surface expression were observed, either at 0hr or after 24hr culture (not shown). To validate the specificity of FcMR staining in controlling for epitope- or fluorochrome-specific effects,

staining was also performed with a second anti-FcMR mAb (clone HM7) conjugated to APC. In addition, FcMR expression was measured with rabbit anti-FcMR polyclonal antibodies (RpAb) as a primary stain and BV421-conjugated anti-rabbit IgG Fc goat polyclonal antibody as a secondary. HM7 and RpAb staining of FcMR showed identical patterns of expression for CD4 T cells and CD8 T cells (**Figure 2.2a,b**) and NK cells (**Figure 2.2d**). B cells cultured in serum-free media displayed more variable amounts of surface FcMR such that only mAb clone HM14 registered a statistically significant increase when comparing "no serum" culture to 0hr samples (**Figure 2.2c**). However, all anti-FcMR antibody stains showed roughly the same magnitude of increases over the 24hr culture period, including for these B cell populations. Our observations of similar staining patterns with three anti-FcMR antibodies and two different fluorochromes suggest the differences we observe are not artifacts of antibody affinity, epitope specificity, or fluorochrome.

To assess whether the presence of human serum, and thus IgM, would suppress or reduce cell-surface display of FcMR, PBMC were cultured with media containing 10% human serum in parallel to cultures with no serum. Unexpectedly, on all four cell types, FcMR upregulation from 0hr baseline occurred in the presence of human serum to the same extent when measured with all three antibody stain formats (**Figure 2.2a-d,** red overlays and bars). Further, for all cell types and all stains there was no difference in the observed FcMR level after culture in serum-free media compared to culture in the presence of 10% human serum (**Figure 2.2a-d**, red compared to blue overlays and bars). These results show that the use of IgM-deficient culture media is not needed to detect

FcMR expression and suggest that IgM may not determine the steady state amount of FcMR on the cell surface to the extent reported elsewhere.

Surface FcMR levels are reduced by cell processing.

Baseline surface FcMR on peripheral blood lymphocytes has been reported previously to be low (53, 106, 108). However, most measurements were taken using PBMC prepared by processing whole blood to remove red blood cells and granulocytes through density gradient separation and multiple wash steps. To test if manipulation of blood lymphocytes affects surface FcMR levels, three separate direct-from-blood staining procedures were performed to determine which allowed reliable FcMR detection with as little cell handling as possible (**Figure 2.3**). The average amounts of FcMR on the surfaces of each lymphocyte cell type were the highest on the least manipulated cells (**Figure 2.3** "stain then lyse", green bars and overlays) and decreased with increasing manipulation. These decreases did not reach statistical significance but were consistent across all cell types. Hence, the comparatively intense manipulation needed to process whole blood into PBMC with a density gradient is likely to underestimate the amount of FcMR displayed by blood lymphocytes *in vivo*. Moreover, the average FcMR ΔMFI obtained by the "stain then lyse" method likely gives a more accurate estimate of FcMR present on the surfaces of circulating lymphocytes *in vivo* than values obtained after preparation of PBMC fractions by standard methods.

Serum or IgM content in culture does not affect surface FcMR display.

To further test for any linkage between IgM content and surface display of FcMR, PBMC were cultured in media supplemented with human sera from multiple donors which contain naturally variable amounts of IgM (**Figure 2.4a**). For these experiments, PBMC from the same donor were tested so that IgM content was the primary variable. Five separate serum sources were used including two separate lots of filtered and heat-inactivated sera, one non-autologous serum, and two separate collections of autologous sera. The IgM content in these sera ranged from 0.36 to 2.3 mg/mL as measured by ELISA, and therefore the dose range in 10% culture was 0.036 to 0.23 mg/mL. For all subsets of lymphocytes, no correlation was found between the amount of IgM in culture media and surface FcMR after 24hr culture (see **Figure 2.4a** for R- and *p*-values), indicating no dose-responsive relationship exists, at least within this dose range.

We next asked whether regulation of surface FcMR requires more IgM than is present in culture media containing 10% human serum. This was done by culturing PBMC with increasing amounts of human serum and assessing FcMR display after culture. For all lymphocyte subsets tested, no consistent or significant differences in surface levels of FcMR were observed when cells were cultured for 24 hrs with 0, 10, 40, 70, or even 100% human serum (**Figure 2.4b**). This indicated that even after culture in physiologic amounts of IgM, FcMR surface levels could be maintained to a similar extent as when IgM was absent. Although not seen consistently, some experiments did show a trend toward decreasing FcMR with increasing serum content, especially in cultures with 70% or 100% human serum. To determine if a culture medium with low (30%), or no, RPMI present had limited buffering capacity we measured pH after 24hr use in culture

and found both 70% and 100% sera media were acidified well outside of physiologic pH range (data not shown). We speculate that this is because RPMI is buffered specifically for tissue culture in incubators with 5% CO_2 such that human serum becomes acidified without added buffers. For this reason, 40% serum was the maximum amount used in subsequent experiments.

To investigate the kinetics of FcMR surface display and determine if increasing serum amounts influenced surface FcMR at time points earlier than at 24 hrs, as tested thus far, time course measurements were performed with PBMC cultures containing 0, 10, 20, or 40% human serum. For B and T cells it appeared that the change in FcMR surface levels were biphasic, with an initial short-term increase that reached a limited plateau as early as 1hr after culture initiation before a secondary increase that began sometime after 6 hrs (**Figure 2.4c**). This biphasic pattern was observed in B and T cell populations regardless of the amount of serum added. When serum was absent, FcMR surface display on CD4 and CD8 T cells lagged that of the serum-replete cultures, indicating that there is no early time point at which FcMR is of greater abundance with IgM absent. Interestingly, the initial plateau reached for both CD4 and CD8 T cells in 10-40% serum culture was similar to the ΔMFI measured for minimally manipulated blood cells (e.g., **Fig 2.3,** "Stain then Lyse" samples), which may provide a more accurate estimate of surface FcMR display by T cells while in circulation. These patterns collectively suggest that T cells cultured in the presence of 10-40% human serum quickly recover surface FcMR that was lost during cell processing, while in the absence of serum the recovery is comparatively slow rather than accelerated. Ultimately, these data further support a limited or absent role for IgM in regulating FcMR surface display of circulating

lymphocytes, which appear to have more FcMR available for functional interactions than previously appreciated.

Higher cell densities during culture impede FcMR display.

During pilot experiments to optimize culture conditions, we observed surprisingly low surface display of FcMR when cells were plated at higher densities. To rigorously test how high-density culture might affect FcMR surface display, PBMC were cultured at cell densities ranging from 0.2 to 1.5 million cells per well and surface FcMR levels were measured after 24 hrs. As in the pilot experiments, when PBMC were cultured at higher densities, FcMR levels on all lymphocyte cell types did not reach the same levels as when PBMC were cultured at lower densities (**Figure 2.5a**). For all cell types, the relationship between FcMR display and cell density in 24hr culture fit with a regression line that had a significantly non-zero negative slope, strongly indicative of a cell density-dependent mechanism.

To determine if higher cell density in culture influenced the kinetics of FcMR expression, PBMC were cultured at cell densities ranging from 0.2 to 2 million cells per well and tested at time points ranging from 0.5 to 24 hrs. As in the previous time courses, B and T cells cultured at low densities of 0.2 or 0.5 million cells per well had a biphasic increase in FcMR display with an initial plateau and subsequent increase between 6 and 24 hrs of culture (**Figure 2.5b**). The initial plateau in surface FcMR expression reached by T cells again approximated the amount estimated to be present on circulating cells *in vivo*, whereas T cells cultured at higher densities of 1 or 2 million cells per well never

recovered to reach this level and, further, remained low at times when the second upregulation of FcMR expression was evident in lower cell density cultures.

FcMR expression by B cells was not as clearly influenced by cell density with significant differences between the 0.2 and 2 million cell groups observed only at one early time point, 2 hrs, which became amplified at the later time points of 12 hrs and 24 hrs. NK cells did not show a discernable pattern, with no significant differences found between the two high and low cell density groups at any time point. Collectively, the findings depicted in **Fig. 2.5** indicate a density-driven suppressive effect on surface FcMR for B and T cells occurring early in culture, most notably for T cells at early time points when it is plausible that increases in surface FcMR are due primarily to recycling of FcMR from internalized pools (53).

One possible explanation for cell density effects is that oxygen becomes limiting when greater numbers of cells are cultured together. Standard incubators keep cultures at close to room oxygen (around 20% O_2), which is hyper-oxic compared to approximate physiologic oxygen tension of 5% O_2 (123, 124). To assess whether oxygen tension might regulate surface FcMR display, cells were cultured at different densities in either normoxia (room oxygen) or physioxia (5% O_2). However, the same patterns of lower surface FcMR after culture at higher cell densities were observed for all lymphocyte cell types regardless of oxygen tension (**Figure 2.6**). Statistically only one difference was observed between the surface FcMR of cells cultured in normoxia compared to physioxia, in NK cell cultures containing 1 million cells per well. All other comparisons showed no differences as a function of oxygen tension, indicating that the density-

dependent effect on surface FcMR expression is probably not a reflection of differences in oxygen availability.

Tonsil cells in culture are not affected by serum IgM or cell density.

Peripheral blood lymphocyte subsets have a different composition than those of lymphoid organs, with the latter possessing high cell densities as well as activated and specialized subsets of cells. To determine if the lymphocyte subsets in one such lymphoid organ, the tonsil, exhibit similar patterns of surface FcMR display in culture to those of PBMC, tonsil mononuclear cells (TMC) were cultured for 24 hrs with varying amounts of human serum (**Figure 2.7a**). Interestingly, the maximum surface FcMR on tonsil lymphocytes was lower than that of corresponding PBMC subsets, whether cultured for 24 hrs with no serum or with 10% human serum. Analyses of NK cells were not included due to low counts in the TMC preparations. For all tonsil lymphocyte subsets observed after 24hr TMC culture without serum, there were no significant increases in surface FcMR compared to those observed at 0hr. In cultures with 10% human serum, a significant increase in surface FcMR from 0hr to 24hr was only observed for the Tfh subset. However, for all subsets there were significant increases in surface FcMR from 0hr for 40% serum cultures. The relative difficulty in observing increases in surface FcMR in culture with lower serum levels could reflect the more extensive manipulation needed to prepare tonsil cells for culture, as evidenced by the higher serum amounts required to recover and FcMR and sustain cell health in our experiments. Regardless, the lack of expected surface FcMR differences after culture with increasing serum amounts

and failure to increase FcMR in cultures without serum suggest that IgM does not affect FcMR display on tonsil lymphocyte subsets.

Tonsil-derived lymphocytes were also tested for the same density-dependent regulation of surface FcMR in culture as previously observed with PBMC. Unlike their PBMC counterparts, a linear regression showed no significant relationship between the cell density and surface FcMR when TMC were cultured for 24 hrs (see **Figure 2.7b** for R- and *p*-values). These data suggest 24hr culture is insufficient for FcMR to recover from high cell density *in vivo* or from the extensive manipulation needed to isolate TMC from intact tonsils. Alternatively, the regulation of surface FcMR in TMC culture may differ from that of PBMC.

The density-dependent regulation of FcMR may not depend on a specific PBMC cell type.

We next asked if a particular cell type present in PBMC might be responsible for cell density-dependent suppression of FcMR surface display. For these experiments we chose to focus on FcMR display by CD4 T cells as these cells consistently exhibited more pronounced responsiveness than CD8 or B cells. To assess whether a specific cell type in PBMC cultures was necessary for the density-driven regulation of surface FcMR on CD4 T cells, five different cell types were independently depleted from PBMC prior to culture at varying cell densities **(Figure 2.8).** All targeted cell types were depleted by at least ten-fold compared to mock depleted controls (platelets, $CD14^+$ monocytes, B, CD8 T, or NK cells; see **Figure 2.9**), but in no depleted culture did surface FcMR fail to be downregulated on CD4 T cells under conditions of high culture density. Although these findings were not definitive, they suggested that density-dependent regulation of

CD4 T cell surface FcMR is not likely to reflect the activity of any specific cell type but may instead reflect bulk cell density.

FcMR display is regulated by cell-cell proximity rather than a soluble mediator.

As no specific cell type was readily identified as responsible for the effect of cell density on FcMR display, we turned to evaluations of cell proximity and secreted factors as explanations. We first tested if the spatial relationships of cells cultured together played a discernible role by comparing FcMR expression after culture over a range of cell densities in U- versus flat-bottom wells (**Figure 2.10a)**. The magnitude of differences attributable to well shape seemed greater for T and B cells than for NK cells, in which FcMR surface expression is generally low under any conditions. FcMR abundance on cells cultured in flat-bottom microplates trended higher than after culture in U-bottom microplates, in which cells settle into multilayer cell clusters at comparatively lower cell densities. Statistically, differences were significant at intermediate PBMC densities; e.g., FcMR was increased more on CD4 T cells in flat-bottom wells containing 1 or 2 million cells per well, but not 0.2, 0.5 or 4 million. This pattern was suggestive of cell-to-cell contact playing a larger role than secreted factors but was not definitive.

To test directly for a soluble factor produced in high-density cultures we first used a transwell culture system that allows exchange of soluble factors between separated populations of cells. PBMC were plated at high- (2 million cells per well) and low- (0.2 million cells per well) density in opposite transwell chambers and surface FcMR was measured after 24hr culture. No significant differences were observed in FcMR surface display on cells in low-density cultures that had been continuously exposed to factors

secreted by high-density cultures (**Figure 2.10b**) suggesting no soluble factors played a role. However, the failure of positive controls for expected density-dependent effects on CD8 T and NK cells indicated that this conclusion was more justified when considering soluble factors that could affect FcMR display on CD4 T and B cells. Specifically, compared to previously significant differences in FcMR display for low- vs high-density cultures when using traditional U-bottom wells, flat-bottom transwell control cultures showed no significant differences in FcMR display between high- and low- density culture for NK or CD8 T cells (though the difference was trending for CD8, $p=0.057$).

To further test for the presence of a soluble factor in high-density culture affecting FcMR display, PBMC were cultured at 0.2 million cells per well in media with 1:1 addition of cell-free supernatant (SUP) collected previously from high- (2 million cells per well) or low- (0.2 million cells per well) density cultures (**Figure 2.10c**). In these experiments, control cultures plated at high-density, without SUP added, exhibited the expected downregulation of surface FcMR. Surface expression of FcMR by all cell types tested was not affected by the addition of SUP from a high-density culture compared to those cultured with SUP from a low-density culture. These results support the transwell data in finding no evidence that a soluble factor produced during high-density culture regulates surface FcMR display. In combination with data demonstrating that the density-driven inhibition of FcMR display is muted in flat- compared to U-bottom wells, these findings suggest the density-driven effect on surface FcMR in culture is dependent on cell proximity rather than a soluble factor.

DISCUSSION

We report here that FcMR expressed on the surfaces of human lymphocytes is not decreased in the presence of IgM, contrary to the prevailing hypothesis of a role for ligand-dependent downregulation of FcMR display (52, 53, 106, 110). We further observed higher surface FcMR on blood lymphocytes when stained immediately *ex vivo*, without prior RBC lysis or processing to isolate PBMC fractions, suggesting FcMR is expressed on the surfaces of circulating lymphocytes at higher levels than previously appreciated (53, 106, 108). We also report a novel cell density effect that strongly restricts FcMR surface display in culture, which we believe to be mediated by a yet unknown cell-proximity-dependent mechanism. Our findings alter the current understanding of factors influencing FcMR display and suggest more FcMR is available at the surface of circulating lymphocytes for functional interactions than had been considered.

The idea that IgM abundance is not correlated with cell-surface FcMR is supported by our finding that FcMR levels were as high after 24-hour culture with IgM-containing human serum present as in serum- and IgM-free media. We additionally found no evidence of a relationship between FcMR display and IgM content when accounting for natural variation amongst different sources of human sera. In time course studies, we saw no time point at which surface FcMR levels were reduced as a function of increasing serum supplementation. In fact, for T cells, the only cultures in which FcMR display lagged were those lacking serum entirely where, based on previous reports (53, 106, 110), the absence of IgM should have increased surface FcMR the most.

As we found no evidence that serum IgM influenced FcMR display, we investigated cell manipulation-related decreases as an alternative explanation for the low amounts of surface FcMR observed on lymphocytes in PBMC fractions of whole blood. To this end, we tested for and found higher FcMR levels on cells after direct-from-blood antibody staining, suggesting circulating lymphocytes have higher surface FcMR display than previously reported by investigators who may have inadvertently stressed cells when using density gradient centrifugation and multiple washes immediately prior to flow cytometric analysis (53, 106, 108). This conclusion seems strongest in the case of peripheral T cells, whose surface FcMR rebounded quickly upon culture within 1 hr and then stabilized at levels approximating those measured by direct-from-blood antibody staining.

Though we did not test for it here, FcMR receptor recycling on T cells has been previously reported to take place in culture (53). We speculate that cell processing to isolate PBMC triggers the internalization of FcMR, causing the apparent amount of FcMR on circulating B, T, and NK cells to seem artificially low, and further speculate that display returns to true circulating levels due to receptor recycling in unperturbed PBMC cultures. Sometime after 6 hrs of culture, surface expression of FcMR began to increase again, which is suggestive of *de novo* synthesis. Hence, we propose that tests of FcMR function in human T cells may best be performed with PBMC that have been rested in complete culture medium for 1-6 hrs, diverging from prior recommendations of overnight pre-culture in serum free media (53, 106, 110), which were based on studies performed before monoclonal antibodies for the receptor became available. Fluorescently labeled IgM was therefore used to stain what would come to be known as FcMR (106,

126), and because serum IgM can block binding by labeled IgM, serum-free media was thought to be necessary. Our results using monoclonal antibodies to stain the receptor reveal that the presence of serum IgM does not alter true receptor display.

In contrast to circulating cells, lymphocytes in secondary lymphoid organs may truly have very low surface FcMR *in vivo*. T cells have been shown to have low surface FcMR not only in our and others' tests of tonsil T cells from TMC, where more extensive cell processing may admittedly be a confounding factor, but also in tonsil thin section microscopy (53, 106). This idea is further supported by our evaluation of high-density PBMC cultures, which are believed to approximate the lymphoid environment in part because they yield better T cell activation outcomes in studies of the optimal conditions needed to prepare T cells for adoptive transfer immunotherapy (127-130). We found that high-density culture markedly downregulated surface expression of FcMR by T cells, especially in U-bottom microplates in which cells cluster at the bottoms of the wells. At no time point tested did T cells in high-density cultures (1-2 million cells/well) reach FcMR amounts we estimate to be present on the surfaces of circulating lymphocytes, a stark comparison to their low-density (0.2-0.5 million cells/well) counterparts which quickly reached this level. Assuming high-density PBMC cultures favor at least some cell-to-cell interactions more typical of lymphoid organs, an intriguing implication is that FcMR may be more functionally relevant in populations of circulating T cells.

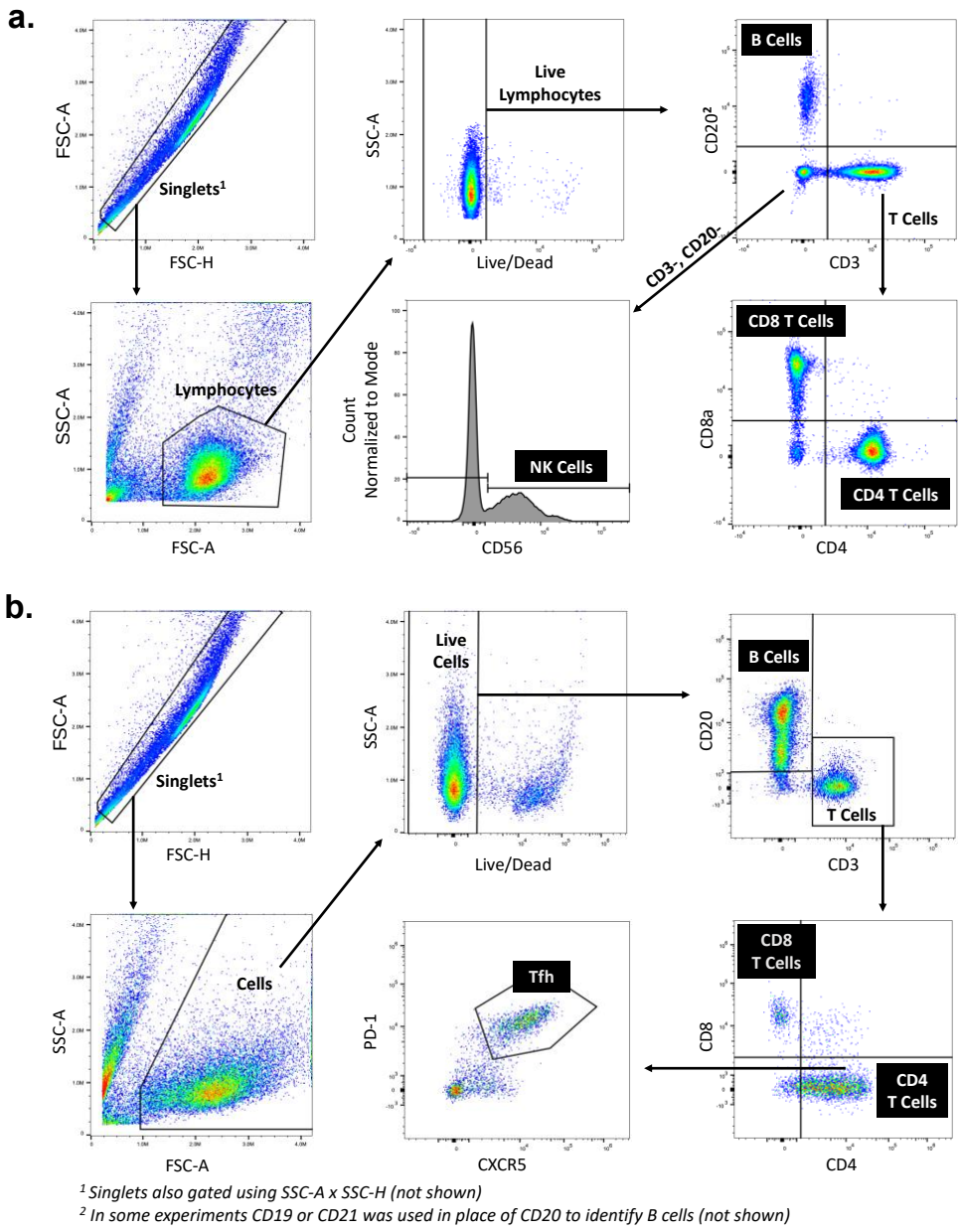
Interestingly, TMC T cells did not display cell density-dependent regulation of FcMR and consistently showed limited increases from a comparatively low starting point for all varieties of cultures tested. This pattern may indicate that mechanisms acting on T cells in high-density environments *in vivo* can continue to suppress surface FcMR display

for as long as 24 hrs in culture. Thus, a crowded cellular environment may not only be a novel contributing factor to low surface FcMR on T cells within lymphoid organs, but exposure to such environments could have lingering effects.

Understanding the mechanism by which high cell density downregulates surface display of FcMR remains an elusive goal. We found no evidence that oxygen tension or secreted factors, including IgM, played a role, and further tested multiple cell types in systematic depletion experiments and found none that could explain the effect. We cannot rule out cell types present in PBMC fractions that were not tested. However, based on the existing data, we speculate that high cell density downregulates FcMR by a mechanism that is not cell-type specific but is driven by close cell proximity, likely requiring either cell-to-cell contact or production of a short-range factor, or both. Future studies are needed to decipher the mechanism(s) involved, as well as to evaluate the significance of the effect for human immune responses, especially those involving T cells.

In summary, our findings reveal a novel cell density-dependent effect on FcMR surface display in culture experiments and support the idea that FcMR expression, and therefore functional relevance, is likely to be greater on circulating lymphocytes than has been previously appreciated. A new understanding of FcMR regulation may in turn contribute to generation of new hypotheses that advance efforts to decipher when, where, and why this unusual receptor is expressed by human T cells.

CHAPTER II FIGURES



[1] Singlets also gated using SSC-A x SSC-H (not shown)
[2] In some experiments CD19 or CD21 was used in place of CD20 to identify B cells (not shown)

Figure 2.1. Gating strategy to identify lymphocyte subsets in PBMC and TMC. Representative flow cytometry dot plots for **(a)** fresh PBMC or **(b)** tonsil mononuclear cells (TMC) showing the gating used to identify B cells, CD4 T cells, CD8 T cells, NK cells and T follicular helper cells (Tfh). Some early PBMC stains did not have viability dye so live cell gating was omitted for those experiments.

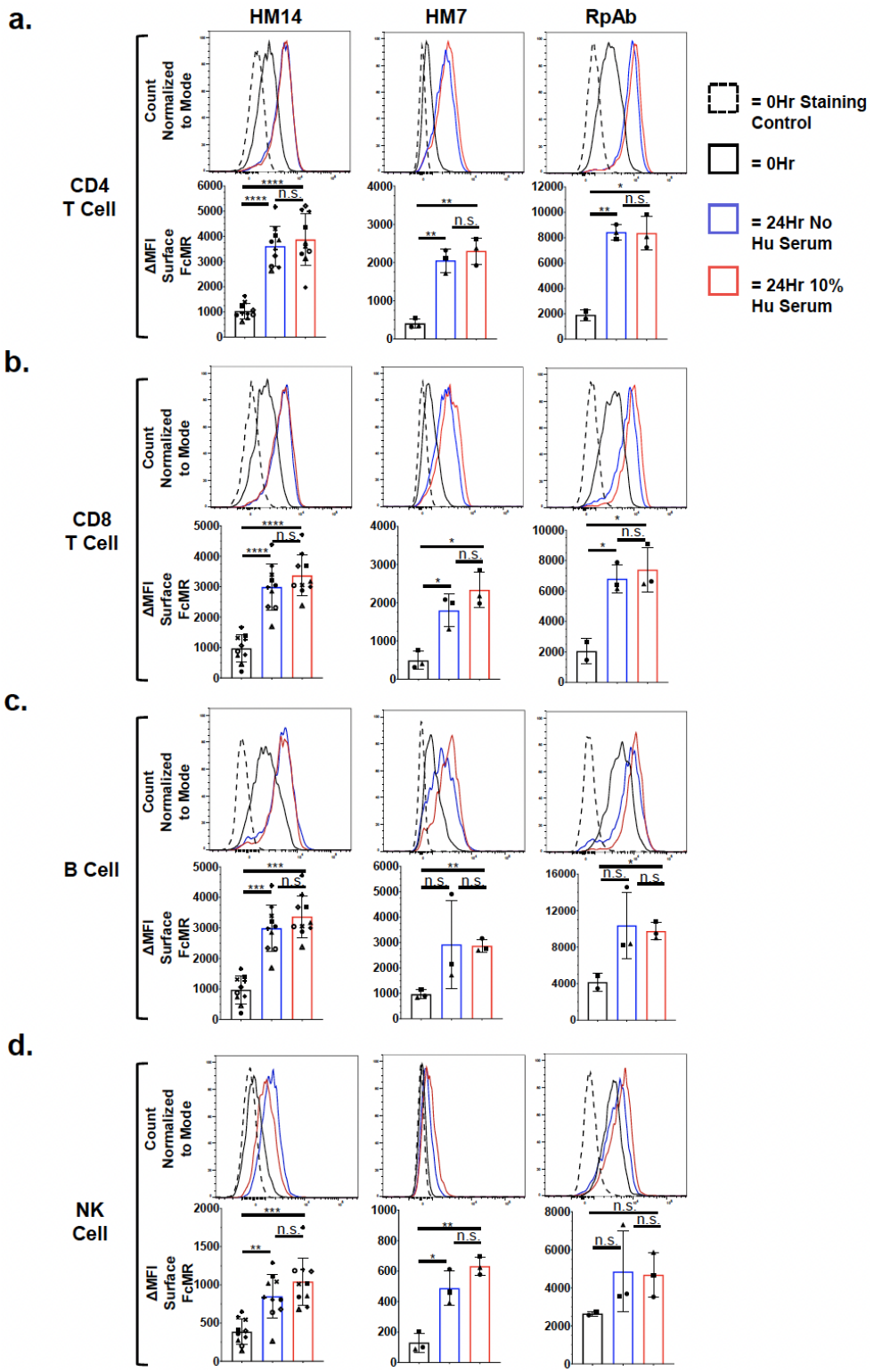


Figure 2.2. Flow cytometric detection of FcMR after culture with or without human serum. PBMC processed from whole blood were stained for FcMR before and after culture using multiple antibodies. **(a-d)** Histogram overlays show FcMR fluorescence

intensity after staining with mAb clones HM14 or HM7, or polyclonal rabbit anti-FcMR (RpAb). Bar graphs show mean fluorescence intensity (MFI) for each stain corrected to ΔMFI by subtracting fluorescence observed with isotype-matched controls for HM14 and HM7, or fluorescence-minus-one for RpAb. Bar heights indicate average values measured from 4 donors in 10 experiments for HM14, 3 donors in 3 experiments for HM7, and 3 donors in 3 experiments for RpAb; error bars denote standard deviations. Independent experiments are represented with different symbol shapes. Welch and Brown-Forsyth one-way ANOVA with Dunnett's T3 post-hoc analyses were used to obtain adjusted *p*-values; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.001$, n.s. not significant.

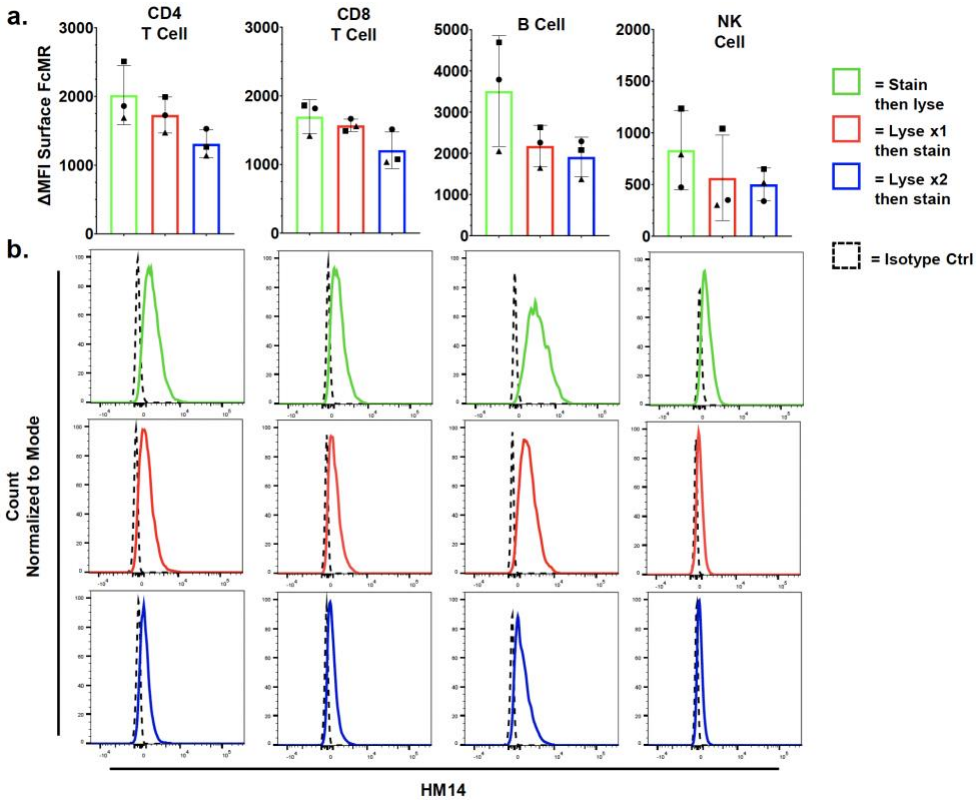


Figure 2.3. Surface FcMR is reduced by processing of blood cells ex vivo. Staining procedures of either minimal necessary manipulation ("stain then lyse") or slightly increased manipulations ("lyse 1x then stain" or "lyse 2x then stain", respectively) were performed in parallel directly after peripheral blood collection to assess how cell processing affected FcMR abundance. **(a)** Bar graphs show FcMR mean fluorescence intensity (MFI, HM14 mAb) corrected to ΔMFI by subtracting fluorescence observed with isotype-matched controls. Bar heights indicate average values measured from 3 donors in 3 experiments; error bars denote standard deviations. Independent experiments/donors are represented with different symbol shapes. Welch and Brown-Forsyth one-way ANOVA with Dunnett's T3 post-hoc analyses were used to obtain adjusted *p*-values; no significant differences were observed. **(b)** Representative histogram overlays showing FcMR fluorescence intensity for each cell type after staining with HM14 mAb versus isotype controls, which were repeated for each processing procedure.

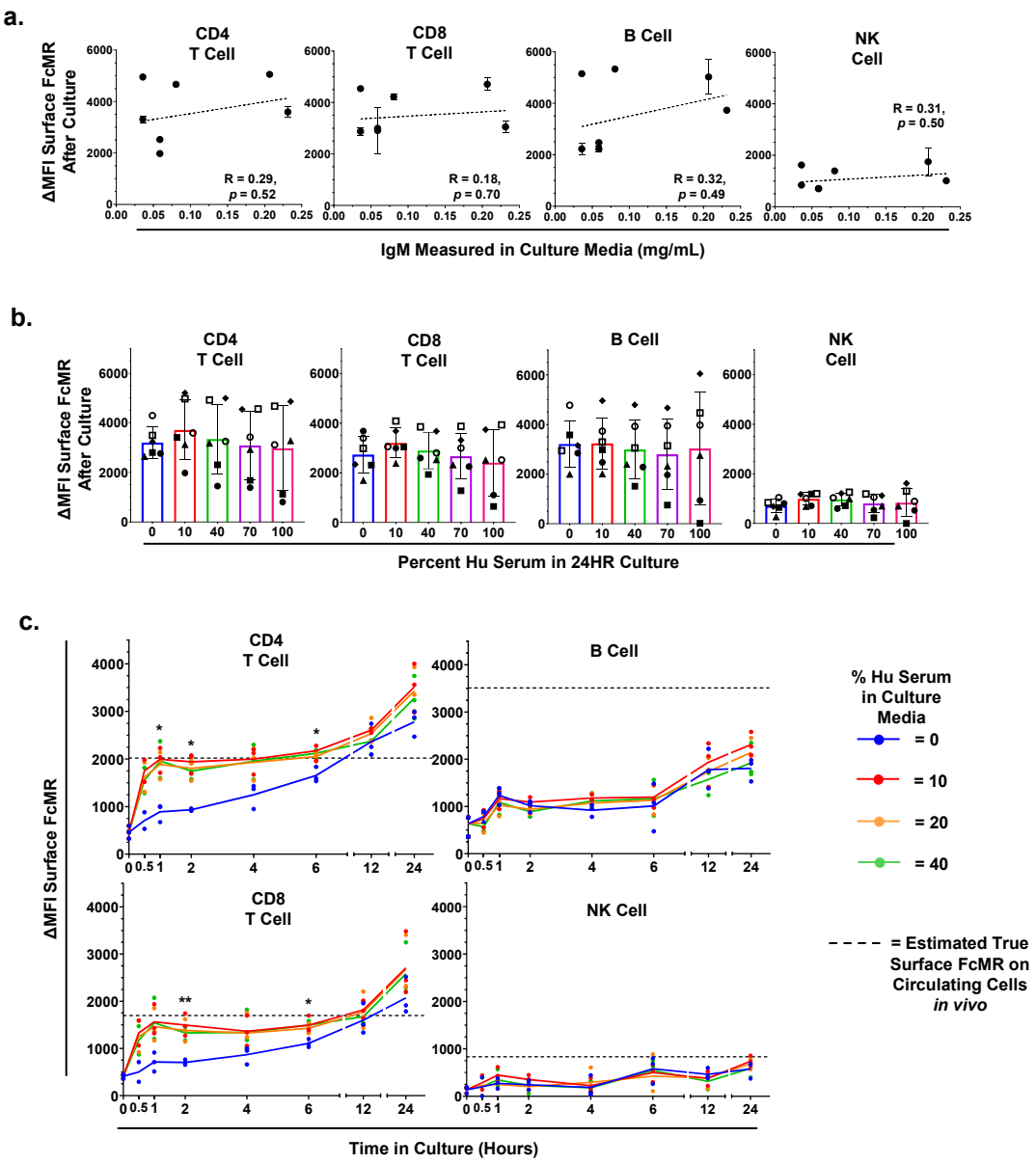


Figure 2.4. Surface FcMR increases to similar levels after 24hr culture regardless of serum or IgM content. PBMC were processed from whole blood and cultured in medium supplemented with varying amounts of human serum prior to flow cytometric measurement of FcMR. **(a)** Cell surface FcMR as a function of serum content after 24hr culture. FcMR mean fluorescence intensity (MFI, HM14 mAb) was corrected to ΔMFI by subtracting fluorescence observed with isotype-matched controls. Bar heights indicate average values measured from 4 donors in 6 experiments; error bars denote standard

deviations. Independent experiments are represented with different symbol shapes. Welch and Brown-Forsyth one-way ANOVA with Dunnett's T3 post-hoc analyses were used to obtain adjusted *p*-values. No significant differences were observed. **(b)** Time course of surface FcMR on B, NK, and T cells cultured with varying serum amounts in media. Solid lines run through average values measured at each time point for 3 donors in 3 independent experiments (in one experiment, the 0.5hr timepoint was not collected). Dotted lines show ΔMFI surface FcMR for each cell type when whole blood was stained directly to minimize cell manipulation (see Fig. 2 "stain then lyse"). Two-tailed Welch's t-tests were performed to compare differences between the 0% and 40% serum cultures at each time point; $*p < 0.05$, $**p < 0.01$, unlabeled is not significant. **(c)** No correlation of IgM abundance in culture media with cell surface FcMR. Cell surface FcMR was measured after 24hr culture in 10% human serum media containing varying amounts of IgM, quantified by ELISA. PBMCs, from the same donor, were tested in five independent experiments. Human serum was derived from one of five sources with varying IgM amounts; two sources were repeated across experiments. Dotted lines show lines of best fit for each cell type; Pearson two-tailed correlation analyses were performed to generate the *p*- and R-values depicted.

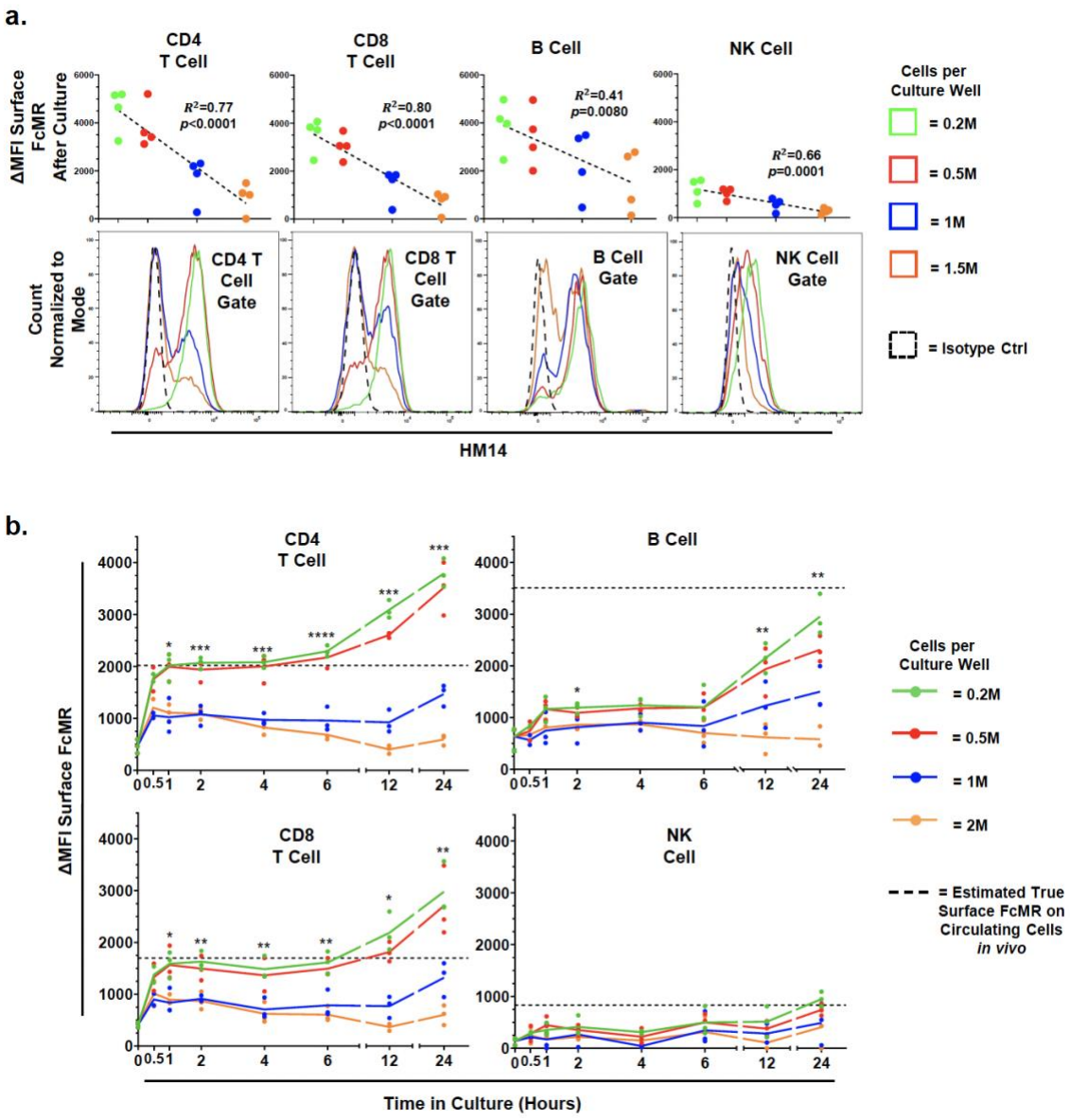


Figure 2.5. High cell density during culture suppresses surface FcMR expression. PBMC were processed from whole blood and cultured at varying cell densities prior to flow cytometric measurement of FcMR. **(a)** Effect of cell density during culture on FcMR expression. Scatterplots for different PBMC cell types show 4 donors in 4 experiments with FcMR mean fluorescence intensity (MFI, HM14 mAb) corrected to ΔMFI by subtracting fluorescence observed with isotype-matched controls. Regression line, R^2, and *p*-values from linear regression analyses are depicted. Histogram overlays

show representative FcMR fluorescence intensity after staining with HM14 mAb or isotype control. **(b)** Time course of surface FcMR on B, NK, and T cells cultured at varying cell densities. Solid lines run through average values measured at each time point for 3 donors in 3 independent experiments (in one experiment, the 0.5hr timepoint was not collected). Dotted lines show ΔMFI surface FcMR for each cell type when whole blood was stained directly to minimize cell manipulations (see Fig. 2 "stain then lyse"). Two-tailed Welch's t-tests were performed to compare differences between the 0.2M and 2M cells/well cultures at each time point; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p <$ 0.001, unlabeled is not significant.

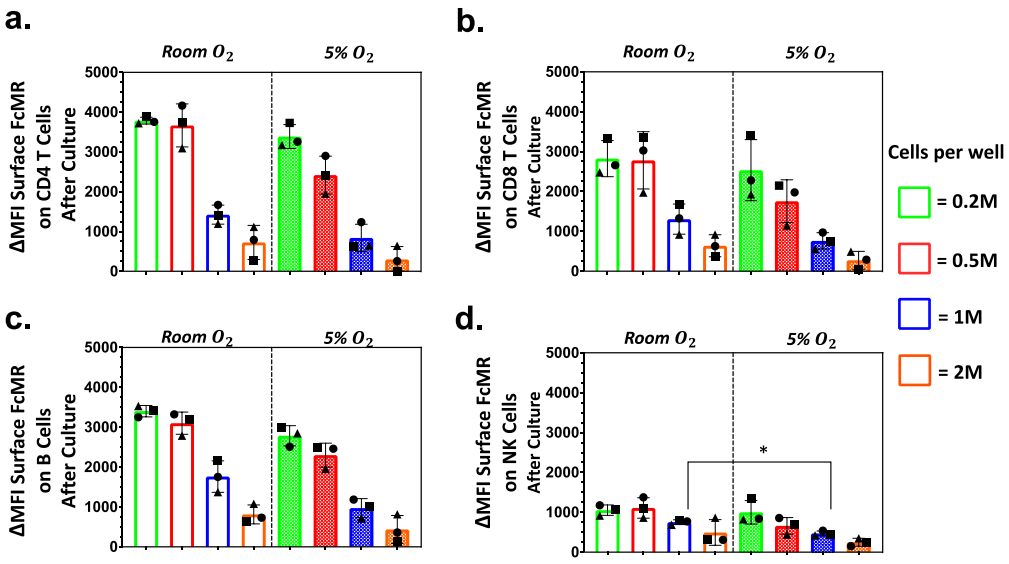


Figure 2.6. Surface FcMR is modulated by cell culture density independent of oxygen tension. PBMC were processed from whole blood and cultured for 24hr at varying cell culture densities at either normoxia (room O_2, 18-20%) or physioxia (5% O_2) prior to flow cytometric measurement of FcMR. **(a-d)** Oxygen tension does not affect density-driven regulation of surface FcMR in culture. FcMR mean fluorescence intensity (MFI, HM14 mAb) corrected to ΔMFI by subtracting fluorescence observed with isotype-matched controls. Bar heights indicate average values measured from 3 donors in 3 experiments; error bars denote standard deviations. Independent experiments/donors are represented with different symbol shapes. Welch and Brown-Forsyth one-way ANOVA with Dunnett's T3 post-hoc analyses were used to obtain adjusted *p*-values. Only comparisons between each respective cell density at room O_2 and 5% O_2 are shown. *p < 0.05, unlabeled is not significant.

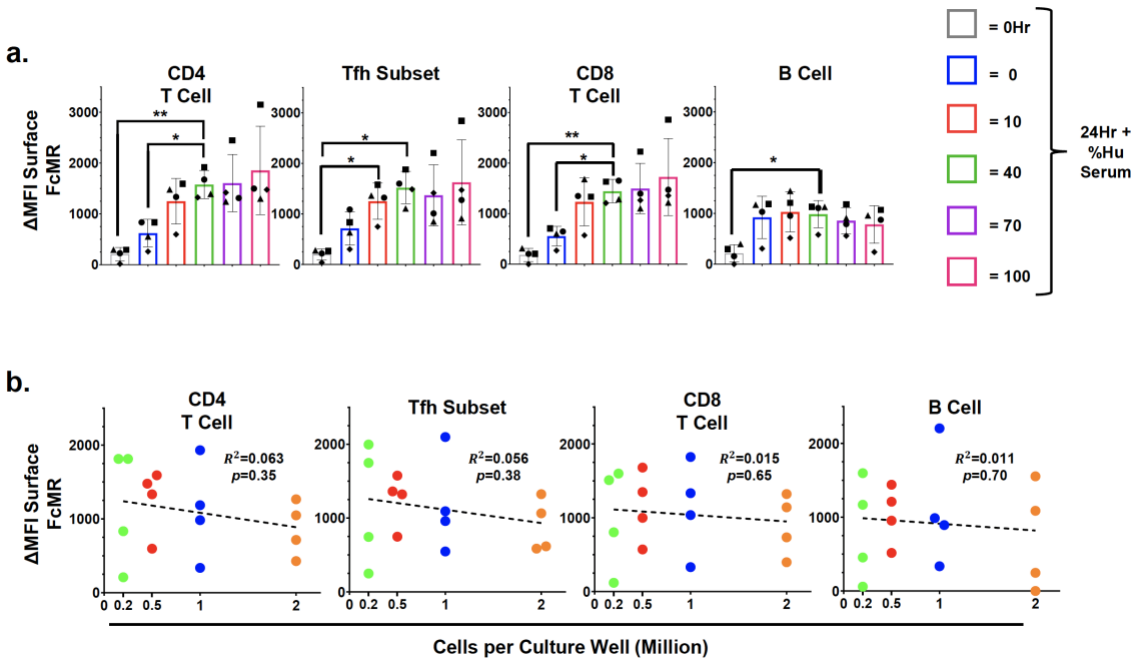


Figure 2.7. Surface FcMR changes after tonsil cell culture. Whole tonsils from obstructive sleep apnea patients were processed to prepare and stain tonsil mononuclear cells (TMC) to assess FcMR surface levels. **(a)** TMC stained at 0hr or after culture for 24hr with varying amounts of human serum. FcMR mean fluorescence intensity (MFI, HM14 mAb) corrected to ΔMFI by subtracting fluorescence observed with isotype-matched controls. Bar heights indicate average values measured from 4 donors in 2 experiments; error bars denote standard deviations. Independent donors are represented with different symbol shapes. Welch and Brown-Forsyth one-way ANOVA with Dunnett's T3 post-hoc analyses were used to obtain adjusted *p*-values. *$p < 0.05$, **$p < 0.01$, unlabeled is not significant. **(b)** The relationship of TMC cell surface FcMR after culture to cell culture density. Scatterplots for different TMC cell types show 2 donors in 4 experiments. Regression line, R^2, and *p*-values from linear regression analyses are shown.

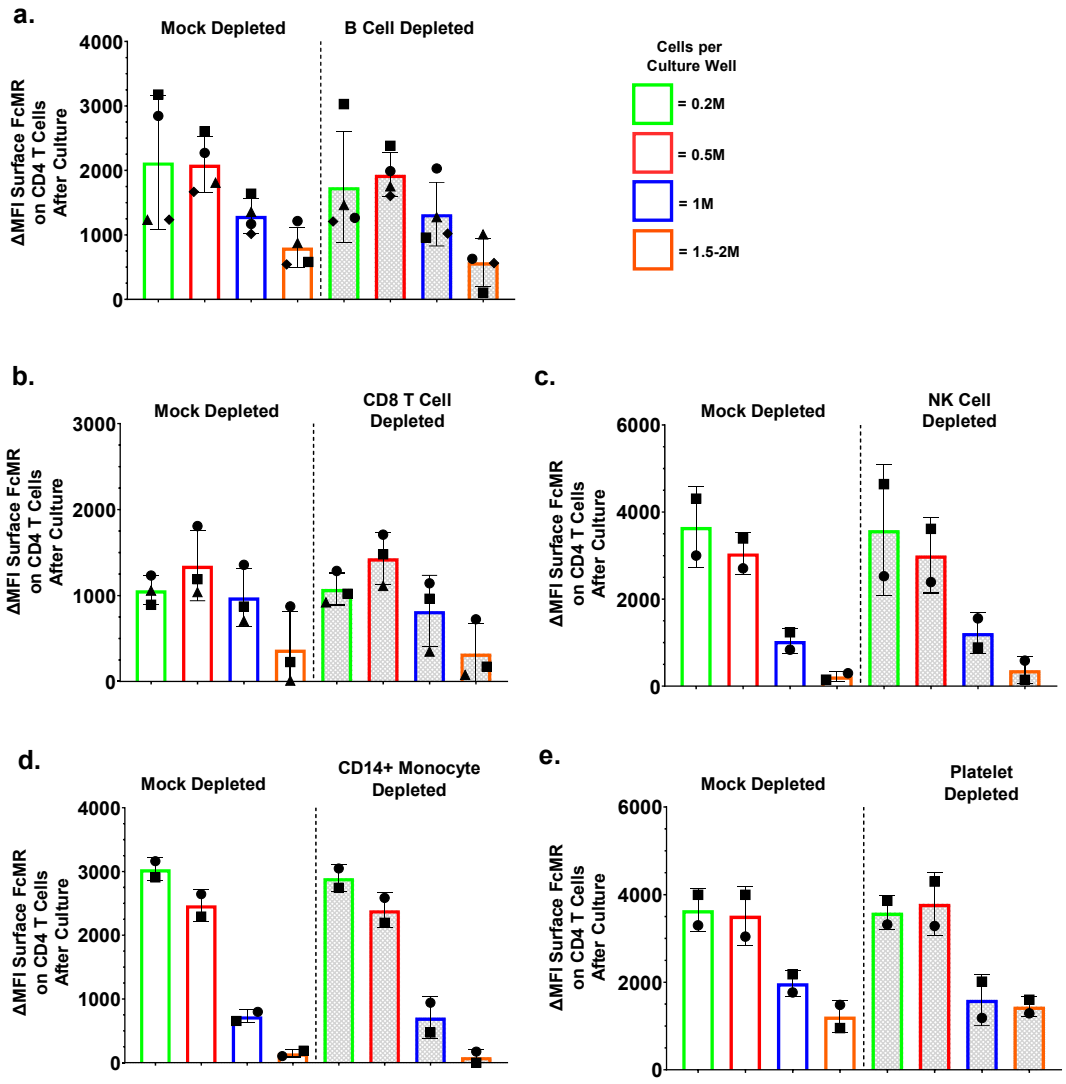


Figure 2.8. The density-dependent regulation of surface FcMR does not depend on a specific cell-type. PBMC processed from whole blood were depleted of various cell types, cultured at increasing cell densities, and stained for FcMR. **(a-e)** The effect of cell density on CD4 T cell surface FcMR is not attributable to B cells, CD8 cells, NK cells, CD14+ monocytes, or platelets. FcMR mean fluorescence intensity (MFI, HM14 mAb) was corrected to ΔMFI by subtracting fluorescence observed with isotype-matched controls. Bar heights indicate average values measured from **(a)** 4 donors in 4 experiments, **(b)** 3 donors in 3 experiments, or **(c-e)** 2 donors in 2 experiments; error bars

denote standard deviations. Independent experiments/donors are represented with different symbol shapes. Welch and Brown-Forsyth one-way ANOVA with Dunnett's T3 post-hoc analyses were used to obtain adjusted *p*-values for comparisons of similar cell densities between mock versus depleted groups; no significant differences were observed.

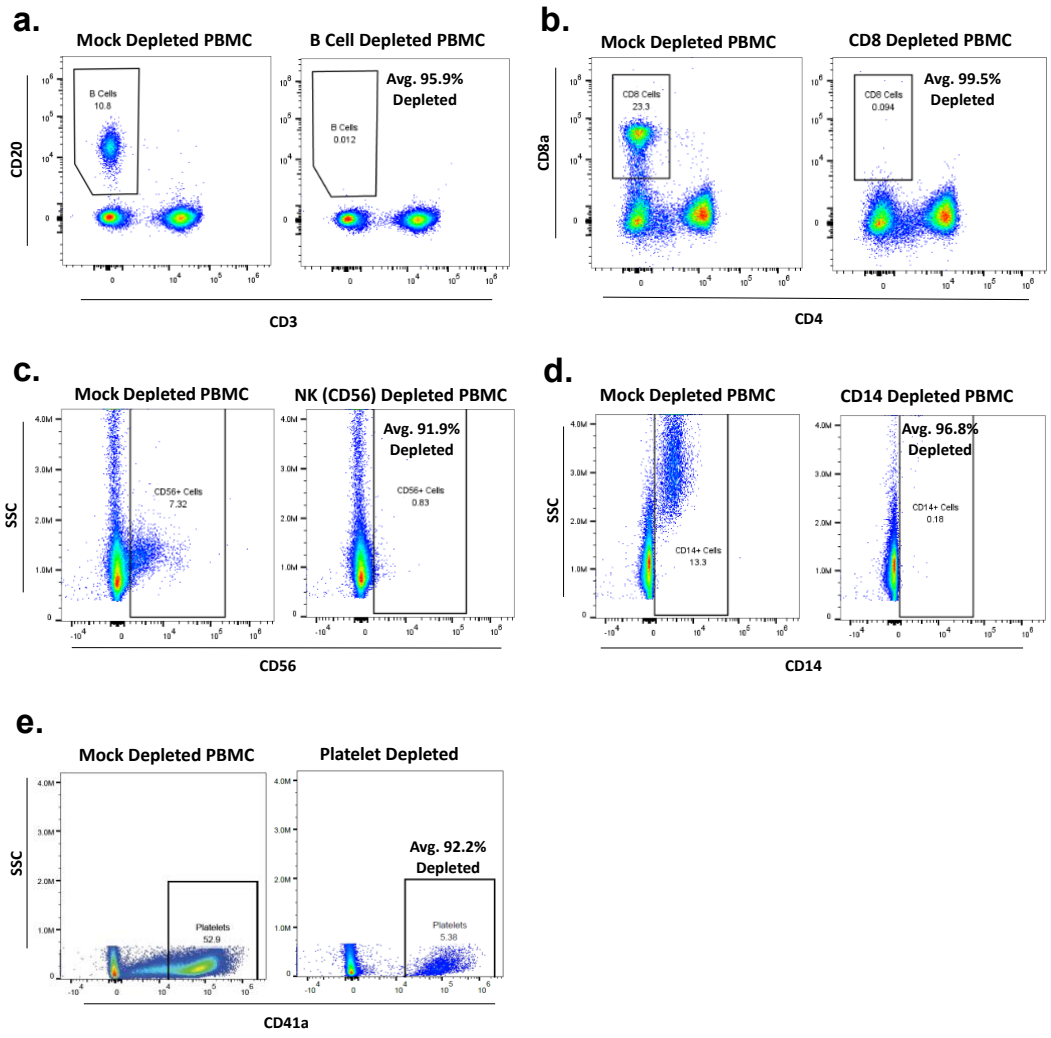


Figure 2.9. Confirmation of cell-type specific depletions. Flow cytometric stains to confirm depletion of **(a)** B cells, **(b)** CD8 cells**, (c)** NK cells, (**d)** CD14+ monocytes, and **(e)** platelets. Plots shown had been gated on live, singlet cells (**a-d**) or on all events less than 0.75 M FSC and 0.75 M SSC (**e**). Average depletion success for each cell type was determined by first finding the average percent depleted compared to mock depleted at 0hr for each experiment and subsequently averaging these percent depleted for all experiments depleting a specific cell type.

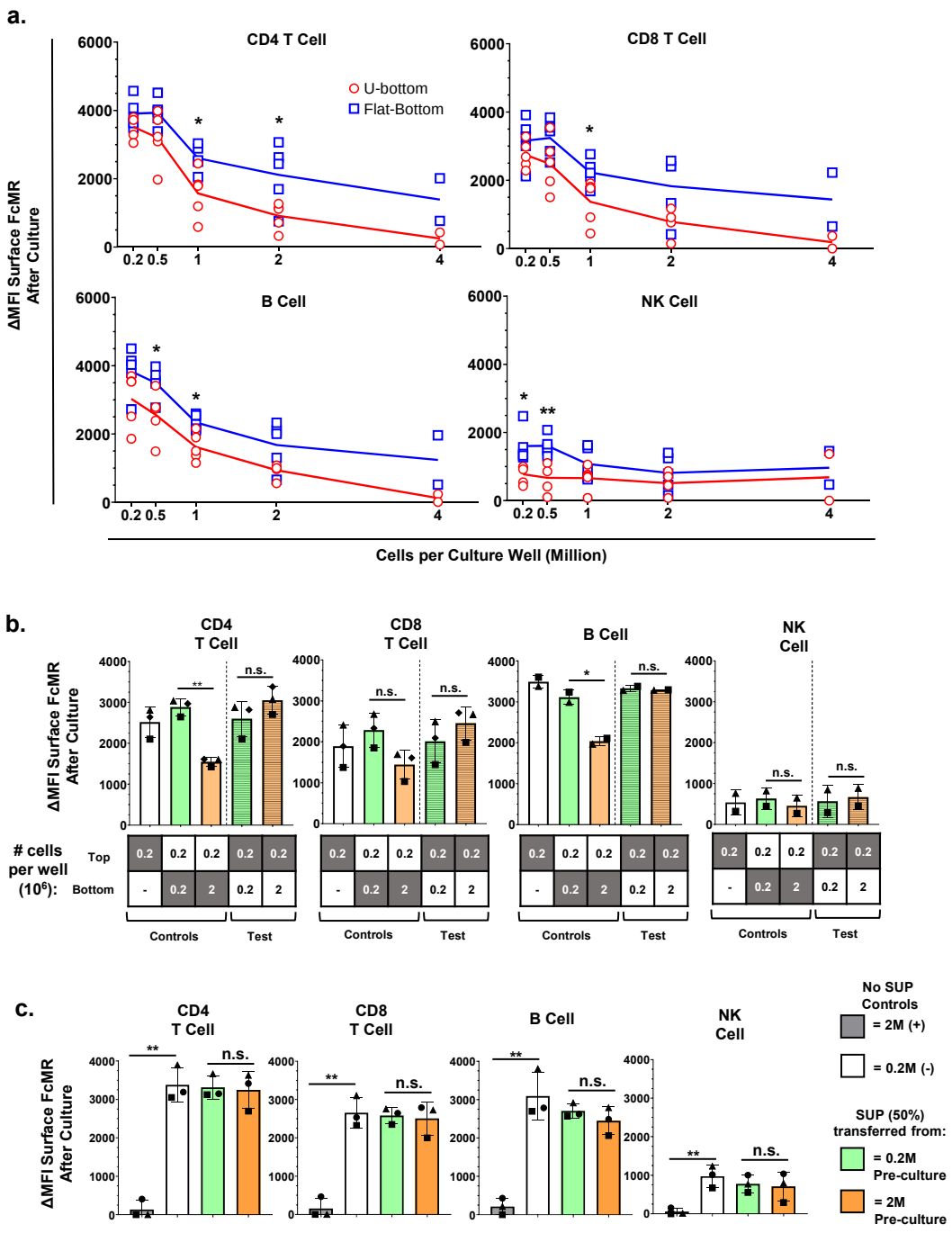


Figure 2.10. Surface FcMR is regulated at high densities in culture by cell-cell proximity but not a soluble factor. PBMC were cultured in varying formats to test for soluble factors versus cell-to-cell proximity effects. **(a)** PBMC cultured at varying cell densities in flat- or U-bottom wells prior to FcMR staining. Solid lines run through average values at each cell density for flat- (blue) or U-bottom (red) wells for 5 donors in

5 experiments (0.2 – 2M cells/well) or 2 donors in 2 experiments (4M cells/well). Two-tailed Welch's t-tests were performed to compare differences between the flat- and U-bottom cultures at each cell density; *$p < 0.05$, **$p < 0.01$, unlabeled is not significant. **(b)** Cells were cultured in the tops or bottoms of transwell plates, allowing media but not cell communication, prior to staining for FcMR. Shaded cells in the table indicate cell populations whose FcMR expression is reported in the bars above. FcMR mean fluorescence intensity (MFI, HM14 mAb) was corrected to ΔMFI by subtracting fluorescence observed with isotype-matched controls. Bar heights indicate average values measured from 3 donors in 3 experiments (for CD4 and CD8 T cells) or 2 donors in 2 experiments (for B and NK cells); error bars denote standard deviations. Independent experiments/donors are represented by different symbol shapes. **(c)** Cells were cultured at 0.2M cells/well in media with 50% cell-free supernatant (SUP) from prior culture of 0.2M or 2M cells/well for 24hr. Positive effect control was PBMC cultured at 2M cells/well with no SUP; negative control was PBMC cultured at 0.2M cells/well with no SUP. Bar heights indicate average values measured from 3 donors in 3 experiments. For **(b)** and **(c)** independent experiments/donors are represented with different symbol shapes. Two-tailed Welch's T test was used to compare means to first assess success of controls and separately assess significance between test groups. *$p < 0.05$, **$p < 0.01$, n.s. not significant.

CHAPTER III

INTRODUCTION TO ISO-SEQ

PRELUDE

In seeking to continue expanding our characterization of the Fc receptor for IgM (FcMR) in human lymphocytes, we considered that isoform-specific variation in this receptor could serve as a means of regulation during the immune response. FcMR is known to possess both a canonical membrane-bound isoform as well as a secreted form that retains the ability to bind IgM but lacks a membrane-spanning domain (118, 131). However, few regulatory mRNA variants of FcMR have been identified (132, 133), and the specific isoforms of FcMR expressed in each lymphocyte cell type have yet to be fully characterized (131). We thus sought to visualize isoforms that were differentially present across lymphocyte subsets and activation states, particularly within T cells where FcMR functional significance and expression patterns remain understudied. In attempting to find reference isoform-aware transcriptomes in human lymphocytes to use as benchmarks, we were surprised to find none had been published. Thus, we sought to bridge this gap by using Pacific Biosciences Isoform Sequencing (Iso-Seq) to generate the first isoform-aware reference transcriptomes for circulating human lymphocytes (Chapter IV) and activated CD4 T cells (Chapter V). These reference transcriptomes are intended to serve as benchmarks for future studies of isoform-specific regulation across a wide variety of lymphocyte receptors.

WHAT IS ISO-SEQ?

Iso-Seq is a long-read sequencing method that makes use of single-molecule real-time (SMRT) sequencing technology to allow visualization of full-length mRNA molecules (94, 134). This is achieved by allowing the entirety of a captured mRNA to enter cDNA synthesis, with subsequent circularization of cDNA permitting a sequencing polymerase to make multiple read passes that can then be collapsed into a highly accurate consensus read for each molecule (**Figure 3.1**, image synthesized using figures from *Corney and Basturea, 2016 (135),* and https://www.pacb.com/blog/understanding-accuracy-in-dna-sequencing/ with permission from Pacific Biosciences' Nancy Francoeur). After initial cleanup and calling of the most accurate reads known as HiFi reads, this technology achieves > 99.9% sequencing accuracy of intact, full-length mRNA transcripts (127, 134, 136).

Importantly, the majority of isoforms that can be visualized using Iso-Seq methods cannot be easily detected using traditional RNA-Seq (134). Although both methods start with polyA capture, in RNA-Seq mRNA is fragmented before cDNA synthesis and generation of "short-read" sequences (**Figure 3.1**). The sequenced fragments must be individually aligned to the genome after which a sequence for the initial, full-length mRNA is inferred from the positions and abundance of the short reads. RNA-Seq is frequently used to quantify gene expression, a purpose for which it is well-suited as the primary analytical output needed is based on the relative read coverage of a particular genetic locus (137). However, when attempting to use traditional RNA-Seq to analyze isoform-aware expression, it is algorithmically challenging to reconstruct and deconvolute isoforms based on fragmented reads (103-105, 137) (**Figure 3.2**, from

https://www.pacb.com/wp-content/uploads/short-reads-vs-hifi-reads-1.svg). This is particularly true when RNA-Seq technology is used in isolation (103-105). Though isoform-aware analysis methods using traditional RNA-Seq are ever-evolving, it is currently beneficial to use both short- and long-read sequencing technologies when investigating differential isoform expression. The former approach provides superior information about transcript abundance, while the latter provides full-length sequences of intact mRNA.

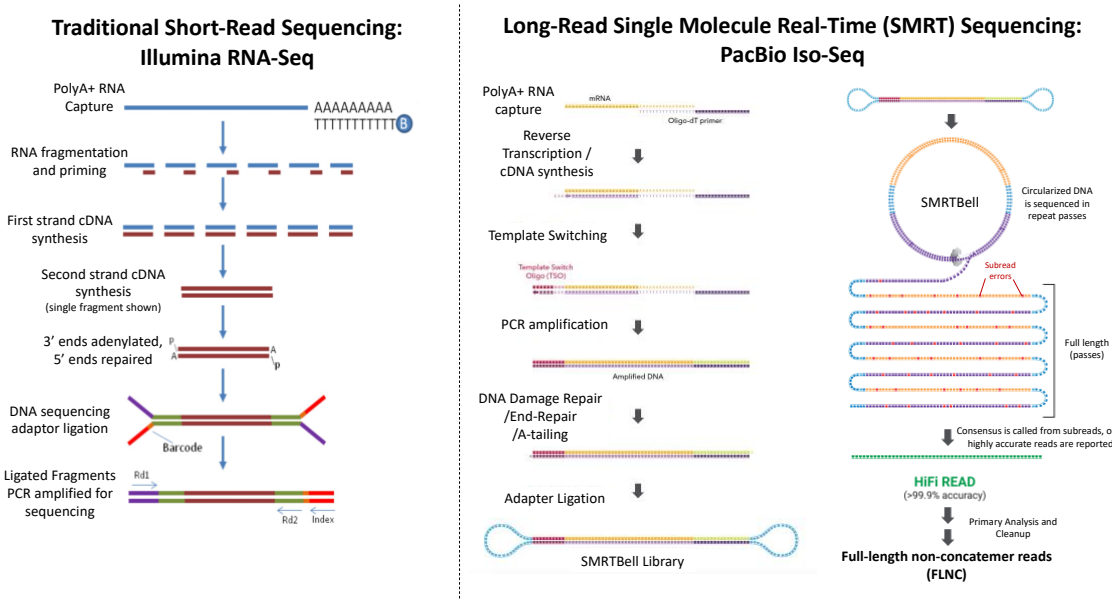


Figure 3.1. Illumina RNA-Seq vs. Pacific Biosciences Iso-Seq. (Left) Image taken from *Corney and Basturea, 2016.* (Right) Image from https://www.pacb.com/blog/understanding-accuracy-in-dna-sequencing/ altered with permission from Pacific Biosciences' Nancy Francoeur.

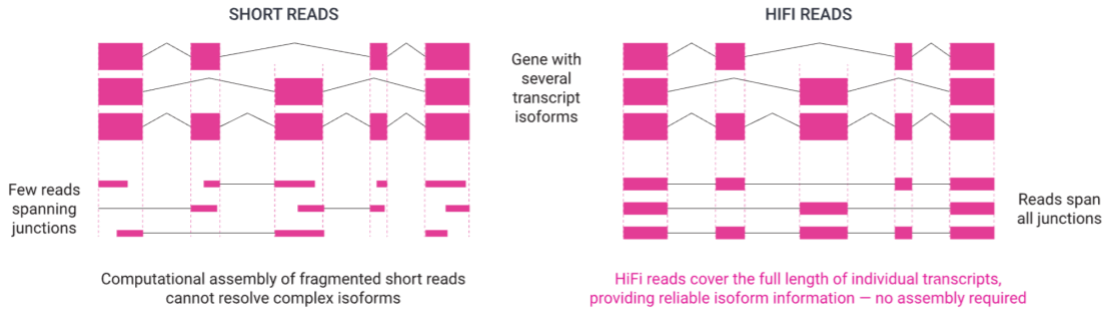


Figure 3.2. Mapping and assembly of RNA-Seq short reads compared to direct mapping of full-length HiFi reads with Iso-Seq. Image taken from https://www.pacb.com/wp-content/uploads/short-reads-vs-hifi-reads-1.svg.

MECHANISMS OF RNA-INTRINSIC REGULATION

Alternative isoform use has become increasingly acknowledged for its role in altering the expression and function of important immune response mediators. The most frequently recognized mechanisms are those involving splice variant isoforms that alter protein coding sequences (138-147). However, there is also a range of immune-significant isoforms that differ only in the in the noncoding regions of mature mRNA, the 5’ and 3’ untranslated regions (UTR) (91, 96, 148-150).

Thanks to the common use of 3’ polyA capture methods in sequencing, 3’-UTR associated regulation of gene expression is relatively well characterized (151). In fact, one of the first mRNA-intrinsic mechanisms of regulation discovered was that of the adenyl-uridine-rich elements (ARE) in the 3’ UTR (151, 152). These regions have been found to be binding sites of RNA binding proteins (RBPs) that typically lead to decreased

expression by destabilization of a transcript, though certain RBP presence and activity may result in the opposite effect as well (151, 153-156).

Exemplifying this, the RBP Tristetraprolin, also known as TTP or ZFP36, mediates the degradation of various inflammatory mediators in resting immune cells by binding to the ARE of their 3' UTR (153, 157-159). Upon exposure to inflammatory agents such as LPS, phosphorylation of TTP inhibits the recruitment of degradation-associated proteins such that the transcripts of desired inflammatory mediators become protected and the proteins they encode are expressed at higher levels (153, 157-159). In a similar situation, during the interferon response, TTP-mediated degradation is antagonized by the simultaneous stabilization of transcripts via additional RBPs, such as EVAL1 (155).

Outside of RBP binding sites, the 3' UTR may also contain binding sites for microRNA (miRNA) that further regulate expression and stability. MiRNAs are small non-coding RNA molecules that bind to messenger RNA and trigger degradation through the induction of the RNA-induced silencing complex (RISC) (160, 161). Alterations in miRNAs and their binding sites can serve as important mechanisms of gene regulation in immune function, as in the hypothesized regulation of peripheral tissue antigen (PTA) expression during thymic immune tolerance (150). Medullary thymic epithelial cells regulate PTA expression by altering levels of both available miRNA and corresponding mRNA isoforms containing complementary miRNA binding sites (150). The concerted and situation-specific effects of miRNA and RBPs binding to 3' UTR can contribute to a variety of RNA-intrinsic regulation outcomes.

Though arguably less studied, the variation of components within the 5' UTR is another important mediator of translational regulation. In traditional cap-dependent translation, the 5' UTR is the first encountered portion of the mRNA during ribosomal scanning, and thus the elements present within or binding to this region may affect the efficiency of translation initiation (97, 98, 162, 163). Like the 3' UTR, the 5' UTR may regulate the expression of a transcript through differential binding sites of RBPs. However, RBPs interacting within the 5' UTR more often directly modulate translation as compared to indirect modulation via changes in RNA stability. Most 5'-binding RBPs will inhibit translation by either blocking ribosomal scanning or interfering with initiation factors. This is the case for the protein bound by iron-response elements; it blocks the association of the 43S ribosomal subunit with the initiation complex (163-165). However, other RBPs promote translation through 5' UTR interaction, such as EVAL1 protein which enhances translation of hypoxia-inducible factor 1α (156, 163, 166).

The 5' UTR also may confer regulatory effects through differences in mRNA secondary structures (97, 98, 163). Highly structured regions upstream of the mRNA initiation sequence can stall a ribosome in its scanning and thus decrease translation (163, 167). Specific secondary structures may also carry out additional gene-specific regulatory purposes, like that of pseudoknot regulation of interferon-gamma (*IFNG*) mRNA (163, 168). In this case, the pseudoknot structure of the 5' UTR activates interferon-inducible kinase PKR which subsequently acts to inhibit translation initiation of *IFNG* mRNA. This allows autoregulatory feedback through PKR because its expression is increased by interferon activation (168).

Aside from secondary structure and altered RBP binding sites, the 5' UTR may also contain upstream start codons or upstream open reading frames (uORFs) that regulate translation outcomes, either directly or by altering transcript stability (163, 169-173). Upstream start codons, both alone and as a part of uORFs, will be recognized as a protein-coding start and thus can stall ribosomes during scanning and force re-initiation prior to translation of the main ORF (163, 172, 174). Because of their short length, the uORF stop codon may also be recognized as premature by RNA surveillance mechanisms and trigger nonsense-mediated decay of a transcript (163, 175, 176). Outside of these *cis*-regulatory mechanisms, a uORF may further encode for a small peptide that can independently interact and regulate the expression or activity of its associated gene (171, 176, 177). In total, due to a combination of these mechanisms, it is estimated that uORFs can reduce the translation of a transcript by 30 to 80% (173). Thus, upstream start codons and, especially, uORFs can be a powerful mechanism of regulation, simultaneously conferring both differential translational efficiency and differential transcript stability.

ISOFORM-SPECIFIC EFFECTS ON THE IMMUNE RESPONSE

Changes in transcript isoforms for a given gene allow immune cells to tailor the availability and functionality of important immune mediators by altering the levels or structure of an expressed protein. Transcripts with alternatively spliced coding exons most often mediate the latter, resulting in mRNA isoforms with corresponding protein variants. Alternative splicing is a functionally significant mechanism of regulation for many immune-important genes, including mediators of T cell apoptosis (144), interleukins and their receptors (146), and HLA family members (147), among others

(141-143, 178). The importance of these alternatively spliced transcripts is reflected in the global changes to the spliced isoform landscape that occur during an immune response to infection or to a vaccine (138-140, 179), and the many dysfunctional outcomes associated with dysregulated RNA splicing (138, 180, 181).

Compared to protein-variant isoforms, transcript isoforms differing only in the UTR, or end-variants, are understudied. Using the mechanisms outlined previously, these end-variant isoforms can confer large variations in protein expression without altering the sequence or structure of the resulting protein. Furthermore, 5' or 3' UTR differences reflect the use of different transcription start or termination sites, respectively, suggesting corresponding differential transcriptional regulation. A recently discovered 5' end-variant isoform of AIM2, for instance, not only possesses novel intrinsic regulation through the inclusion of an iron response sequence element but also maps to a previously unannotated promoter (95).

A shift in end-variant use can mediate a quick change in protein synthesis rates which may play important roles both in normal immune response and in disease states such as cancer (91, 96, 148-150, 182-184). Alternatively spliced protein-variant isoforms are also thought to be important in these processes, mediating both physiologic and disease-associated changes to protein function (142-144, 146, 147, 178, 181). Still, many of the nuances surrounding isoform-specific regulation remain under-characterized, particularly in cells where isoform-aware insight is lacking.

A NEED FOR ISOFORM-AWARE REFERENCE TRANSCRIPTOMES

As most protein expression variation cannot be explained by fluctuations in mRNA levels alone (90-93), established differential expression analyses are not enough to gain a full understanding of the changes observed in disease and during altered immune states. We must also understand the landscape of isoforms present within expressed mRNA. While traditional RNA-seq analyses rely on algorithmic prediction to infer full-length sequences (103-105, 137), the Iso-Seq method allows direct visualization of the true sequence of novel isoforms, with the potential to identify differences between cell types and cellular environments.

Before we can begin investigating highly specific aspects of isoform variation, it is best to have a well-annotated roadmap of what to broadly expect in the form of an isoform-aware reference transcriptome. Isoform-aware references are particularly important for studies using single-cell RNA-Seq isoform prediction where fragmented sequences are mapped to known isoforms to estimate the abundance of transcript variants (185). Here, having the most complete isoform reference possible is essential. Even with the advent of newer single-cell isoform sequencing technology, bulk sequencing remains unmatched in depth and breadth of coverage for novel transcript discovery (186-189). Bulk Iso-Seq can identify lower-expressed novel isoforms of likely functional significance such that validation and consequences of these isoforms in specific cell types can be followed up with less expensive and more targeted studies. Iso-Seq also serves as a particularly powerful tool for identifying novel regulatory elements, as full-length sequencing is championed for identifying and visualizing end-variant isoforms present in a sample (190-192).

Due to a current gap in reference isoform-aware transcriptomes, databases of transcript variants (94, 132, 193) and UTRs (97, 98) are likely incomplete for human immune cells. There is a need for isoform-aware references to comprehensively cover isoform presence in these samples and provide a baseline for future targeted studies. We present here the first reference Iso-Seq transcriptomes of four circulating human lymphocyte subsets (Chapter IV) and early activated human CD4 T cells (Chapter V), which provide insight into the isoform landscape of these cellular states. We further identify novel end-variant transcripts from these references that warrant further study for their potential role in the regulation of immune-important genes.

CHAPTER IV

REFERENCE LONG-READ ISOFORM-AWARE TRANSCRIPTOMES OF FOUR HUMAN PERIPHERAL BLOOD LYMPHOCYTE SUBSETS[2]

INTRODUCTION

Structural details of transcript isoforms, also called transcript variants, are important to catalog to advance understanding of the roles they play in cellular fitness. With the isoform sequencing (Iso-Seq) method, individual mRNA molecules are first converted into cDNA and then into circular templates which undergo multiple rounds of iterative sequencing using Single Molecule Real-Time (SMRT) technology (Pacific Biosciences). Ultimately, this generates a highly accurate (>99.9%) intramolecular consensus read of each full-length mRNA transcript (127, 136). Iso-Seq has the potential to identify important changes in isoform structure, such as those occurring in activated and/or memory human lymphocytes participating in the immune response. However, to robustly characterize functional changes in RNA biology in the context of innate and adaptive immunity, a reference transcriptome is needed to serve as a benchmark. Thus far, most human Iso-Seq transcriptomes currently published have included stem cells, cancer lines, or are pooled broadly from multiple tissues (94, 136, 194, 195). To our

[2] Woolley, C. R., J. H. Chariker, E. C. Rouchka, E. E. Ford, E. A. Hudson, S. J. Waigel, M. L. Smith, and T. C. Mitchell. 2022. Reference long-read isoform-aware transcriptomes of 4 human peripheral blood lymphocyte subsets. *G3 Genes|Genomes|Genetics* 12. doi: 10.1093/g3journal/jkac253.

knowledge, no Iso-Seq transcriptome had been published for primary human B, T, or NK cells, even though isoform-specific expression patterns are known to play a role in defining lymphocyte development and functions (178, 181). To begin to fill this gap, we purified these four lymphocyte subsets from the peripheral blood of a healthy donor to obtain high-quality RNA (RIN>8) for SMRT sequencing and Iso-Seq analysis. Full-length non-concatemer reads (FLNC) as well as processed Iso-Seq data files have been deposited to the Gene Expression Omnibus (GEO) database (196) alongside matched data from Illumina short-read RNA-Seq that was performed in parallel (GSE202329 Super Series). Many novel transcript isoforms supported by both Iso-Seq and RNA-Seq data were identified. Each cell-type sample and its corresponding sequence data set met several metrics of quality as assessed by flow cytometric and SQANTI3 (192) analysis, respectively, indicating they will be valuable as benchmarks for future studies.

RESULTS AND DISCUSSION

Quality of purified cell samples

Purified cells used to extract RNA for sequencing were sampled and found to be highly viable (viability $\geq$ 91%, **Table 4.1**). $CD4^+$ T, NK, and Pan B cells showed excellent purity (purity > 95%, **Table 4.1**), suggesting RNA extracted from these samples was also reasonably exclusive of other cell types. $CD8^+$ T cells, on the other hand, were less pure (91.7%), with the major contaminating cell types being $CD4^-CD8^-CD3^+$ cells (6.8%, **Table 4.1**) and Pan B cells (1.4%, **Table 4.1**). Thus, sequencing data obtained for the $CD8^+$ T cell sample, though still highly reflective of the cell type, should be considered less specific.

The lymphocytes used in this study were purified by negative selection to avoid transcriptional artifacts caused by antibody binding. Because our goal was to generate reference transcriptomes that were broadly representative of subsets within each circulating lymphocyte population, purification kits with fewer exclusions were chosen whenever possible. For example, the Pan B cell kit was selected because it did not exclude $CD43^+$ cells, a marker of plasma B cells (**Table 4.2**, $CD19^+$ and/or $CD43^+$ B cell frequencies), and the $CD4^+$ and $CD8^+$ T cell purification kits did not exclude $CD4^+CD56^+$ or $CD8^+CD56^+$ NK T cells, respectively (**Table 4.1**, NK T cell frequencies reported among $CD45^+$ cells alongside purity). Further, based on flow cytometric data from two previous experiments, the peripheral blood donor possessed a range of circulating naïve and memory T cell subtypes which we viewed as desirable for these reference data sets (**Table 4.3**). Ultimately, all four isolated populations were high-quality and representative samples of their respective lymphocyte subset.

RNA-Seq reads and genomic mapping

RNA sequencing generated ~102-124 million short reads per lymphocyte population, and when allowing the reads to map to multiple loci the alignment rates were above 98% for all samples (**Table 4.4**). To parallel the parameters used for Iso-Seq analyses, we also mapped RNA-seq reads to hg38 in a manner that restricted their alignments to a single locus. Alignment rates for these uniquely mapped reads were between 85 and 89 percent (**Table 4.4**). The high alignment rates achieved with both mapping strategies indicated the RNA-Seq datasets were of high quality.

Iso-Seq sequencing and initial data metrics

To generate Iso-Seq transcriptomes, pairs of barcoded samples were loaded on each of two 8M SMRTcells and sequenced on the Sequel IIe system, which produced ~1-2 million full-length non-concatemer reads (FLNCs) per lymphocyte population. SMRTcell loading, read length and yield metrics fell within expected ranges for both SMRTcells (**Table 4.5**). For data derived from both SMRTcells, 50% of bases were in reads > 110,000 (read length N50), allowing > 30 passes of the average 3 kb mRNA, which resulted in highly accurate HiFi data (Q38, 99.98%) as the input into Iso-Seq analysis. These metrics were consistent with the generation of highly accurate and extensive sequencing data for all four lymphocyte populations.

Iso-Seq transcript annotation

After processing HiFi reads into FLNCs the mRNA transcripts they represented were categorized and summarized using SQANTI3, provided any given transcript had been observed as at least two unique molecules. The four lymphocyte populations expressed 11,542 to 14,487 genes with 34,211 to 59,845 distinct isoforms, as determined by the locations of splice junctions and transcription start and stop boundaries (**Table 4.6**). The number of annotated genes ranged from 10,402 ($CD8^+$ T cells) to 12,162 (B cells) across the individual samples. The 961 to 2,325 novel genes (**Table 4.6**) were categorized as such because they had no annotated counterparts in GENCODE v39 and may represent Ig or TCR gene rearrangements, although several hundred mono-exon transcripts mapped to novel locations in the genome. The 15,433 to 33,918 novel

transcripts identified (**Table 4.6**) were categorized as such because they did not match any known Gencode v39 annotated transcript.

Iso-Seq results after filtering for short-read coverage

To reduce the risk of artifacts, the genes and isoforms detected by Iso-Seq alone (**Table 4.6**) were filtered to include only those with at least one RNA-Seq short-read for every internal splice junction present in a transcript. As shown in **Table 4.7**, the exclusion of long-read transcript sequences lacking short-read support reduced the number of unique isoforms detected to 60% of the value observed before filtering. This reduction occurred more frequently with annotated genes than novel genes because 92% of the latter remained after filtering (an average of 1255 novel genes detected after, as compared to 1359 before, filtering). However, only 55% of novel transcripts remained after filtering (an average of 13,563 novel transcripts detected after, as compared to 24,680 before, filtering), suggesting filtering effectively refined the dataset, as desired, while retaining most novel isoforms.

Classification of transcript isoform structures

Filtered transcript isoforms were classified by SQANTI3 into several structural categories, based on alignment to previously reported reference transcripts, and the usage of known donor and acceptor splice sites. These were either full splice match (FSM), incomplete splice match (ISM), novel in catalog (NIC), novel not in catalog (NNC), intergenic (between annotated genes), antisense (anti-sense to an annotated gene), fusion (fusion of two annotated genes), genic genomic, or genic intron (**Figure 4.1**). Most

isoform structures matched their reference transcripts completely (FSM), but many novel transcript structures were also identified within each sample (e.g., NIC and NNC, **Table 4.8**).

Across all samples, the median values and distribution of lengths in novel (NIC and NNC) isoforms were similar to those of corresponding transcripts with complete or partial reference matches (FSM and ISM) (**Figure 4.2a**). Further, most novel transcripts identified were predicted to correspond to protein-coding transcripts (**Figure 4.2b**). Though these transcripts remain to be fully validated, these attributes support the notion that many novel transcripts identified in our sequence datasets are likely real variants of known transcripts with potential functional relevance in protein expression.

Novel transcripts (NIC and NNC) were primarily enriched within immune-important signaling pathways (**Figure 4.3**), further supporting the potential for relevance to the field. However, enrichment here does not directly mean a pathway was active, as many cell-type specific pathways share common genes, such as T and B cell receptor signaling sharing GRB2. It is also important to consider that isoform functionalities can be vastly different, even for the same gene (197), so the roles of novel isoforms may not align with their assigned gene-level functional classifications. Still, pathways enriched for novel isoforms, and particularly those differentially enriched between samples such as IL-2 signaling, could have yet unexplored layers of cell-specific isoform-level regulation. Further functional validation of novel isoforms is necessary before more accurate annotation and enrichment analyses can be performed.

Quality metrics of filtered isoforms

Using the top four isoform categories (FSM, ISM, NIC, NNC) to assess metrics of quality for each sample dataset via SQANTI3, we observed relatively high percentages of transcripts within each category demonstrating good quality attributes and low percentages of transcripts with bad quality attributes (**Figure 4.4a-b**, respectively). For example, over 90% of transcripts in all categories mapped to known annotated genes, and the majority had cap analysis gene expression (CAGE) peak and polyadenylation (polyA) motif support (**Figure 4.4a**). On the other hand, low fractions of transcripts were predicted to have undergone nonsense-mediated decay (NMD) or reverse transcriptase (RT) switching, both common causes of artifactual novelty (**Figure 4.4b**).

Compared to the unfiltered transcriptomes, filtering for short-read coverage of internal splice junctions increased the fraction of transcripts for which all splice junctions were supported by short-read data (**Figure 4.4a**, unfiltered data not shown). Yet, for all categories within our filtered dataset, there were a fraction of transcripts with no short-read coverage for at least one splice junction. This is likely due to our filtering based on only internal splice junctions rather than all splice junctions, as was considered for this metric. This fraction was particularly high for the novel, not-in-catalog (NNC) transcripts where 73.8% to 84.2% of transcripts contained at least one splice junction without short-read coverage (**Figure 4.4b**). A high number of our NNC transcripts also contained at least one noncanonical splice junction (35.7% to 40.8%, **Figure 4.4b**). However, when considering the overall number of noncanonical splice junctions out of all junctions in these transcripts, the percentage is much lower (all samples < 10% noncanonical junctions with the rest being novel canonical or known canonical, **Figure 4.4c**),

suggesting this characteristic may be reflective of the NNC classification itself rather than the transcript quality.

Taken together with the high quality of each contributing cell sample, these metrics support the notion that the filtered datasets provided here consist of representative and high quality polyadenylated transcripts, some entirely novel and previously unannotated. These will add to the known catalog of transcript isoforms present in healthy circulating B, T, and NK cells, and will be valuable for reference in future lymphocyte transcriptomic analyses.

MATERIALS AND METHODS

Human blood collection and PBMC isolation

Venous blood, 175 mL, from a healthy consented 57-year-old male donor was collected in and adjusted to 6mM K_3EDTA using standard phlebotomy. Collection was approved by the University of Louisville Institutional Review Board under expedited review (IRB 14.0661).

PBMCs were isolated using Sepmate™ PBMC Isolation Tubes (StemCell Technologies, cat no. 85450) as previously described (56), with an additional granulocyte depletion step (RosetteSep™ Human Granulocyte Depletion Cocktail, StemCell Technologies, cat no.15624) as directed by the manufacturer. Cell yield was determined using the count per μL feature of a Cytek® Northern Lights flow cytometer as previously described (56), immediately after which the PBMCs were further processed to purify blood lymphocyte subsets.

Human lymphocyte subset enrichment

Lymphocyte subsets were purified from freshly prepared PBMC using EasySep™ (StemCell Technologies) negative magnetic selection kits: Human $CD4^+$ T Cell Isolation Kit (StemCell Technologies, cat no. 17952), Human $CD8^+$ T Cell Enrichment Kit (StemCell Technologies, cat no. 19053), Human NK Cell Enrichment Kit (StemCell Technologies, cat no. 19055), and Human Pan-B Cell Enrichment Kit (StemCell Technologies, cat no. 19554) per manufacturer's instructions. After purification, two million (Pan-B) or three million ($CD4^+$ T, $CD8^+$ T, and NK) isolated cells were lysed in Buffer RLT Plus (RNeasy® Plus Mini Kit, Qiagen, cat no. 74134), per the manufacturer's protocol and lysates were kept at 4^o C until RNA extraction. The remaining isolated cells were used for flow cytometric staining of markers to assess cell viability and purity as described below.

Flow cytometric staining and analysis

Isolated cells were distributed at 0.2 million cells per well in a 96-well plate for flow cytometric staining (56) with some modifications. For viability staining, cells were first washed twice with PBS lacking calcium and magnesium ($PBS^{-/-}$) and then resuspended in 100 µL of $PBS^{-/-}$ containing eBioscience™ Fixable Viability Dye eFluor™ 780 (Thermo Fisher, cat no. 65-0865-14). Cells were incubated for 30 minutes prior to wash and resuspension in $PBS^{-/-}$ followed by fixation with 1% formaldehyde and transfer to 12x75 mm flow cytometry tubes. Fixed cells were kept on ice until flow cytometric analysis.

To evaluate the purity of the lymphocyte subset samples, cells were washed twice with standard stain buffer ($PBS^{-/-}$ with 0.09% NaN_3 and 2% human serum) prior to resuspension in a cocktail of antibodies specific for markers of myeloid and lymphoid lineages. The antibody cocktail was generated by first adding appropriate amounts of fluorescent-conjugated monoclonal antibodies (**Table 4.9**) to Brilliant Stain Buffer Plus (BD Biosciences, cat no. 566385), and subsequently adding standard stain buffer to reach a cumulative 100 µL per test. After resuspension in this antibody cocktail, cells were incubated for 30 minutes prior to wash and resuspension in stain buffer. Cells were then fixed with 1% formaldehyde and transferred to flow cytometry tubes which were kept on ice until flow cytometric analysis.

Flow cytometry was performed with a Cytek® Northern Lights 3-laser flow cytometer, and spectral profiles of each fluorophore were unmixed using SpectroFlo® software (Cytek Biosciences) and appropriate single-stain and unstained controls to account for autofluorescence. Processed data files were analyzed in FlowJo™ (BD Biosciences). Purity and viability were determined using the gating strategies shown in **Figure 4.5**. Purity is reported as the average of two technical replicates, and viability as a single replicate value, of a fraction of the same cells used for RNA isolation.

RNA extraction and purification

Total cellular RNA was extracted and purified from 2-3 x 10^6 cells per sample using RNeasy® Plus Mini Kit (Qiagen, cat no. 74134) following the manufacturer's protocol. RNA from each lymphocyte subset sample was eluted in 30 µL of RNA-ase

free water and kept in a cold block on ice until Iso-Seq library preparation the same day. Remaining RNA was frozen and kept at -80^{o}C for RNA-Seq library preparation.

Iso-Seq library preparation, sequencing, and initial data analyses

To generate Iso-Seq libraries, 500 ng of high-quality RNA (RIN > 8) was used as initial input into oligo-dT primed cDNA synthesis using commercially available NEB Next reagents (New England Biolabs, cat no. E6421L). Barcoded primers were incorporated into the cDNA during second-strand synthesis. Following double-stranded cDNA amplification, transcripts were equimolar pooled to include two samples per SMRTbell library preparation ($CD4^{+}$ and $CD8^{+}$ T; NK and Pan B). SMRTbell libraries were generated from the pooled cDNA as recommended by the manufacturer. Briefly, the pooled cDNA underwent enzymatic DNA damage and end repair prior to ligation with SMRTbell hairpin adaptors. Final libraries were purified with magnetic beads prior to annealing to sequencing primer (v4) and binding to polymerase (v2.1). Sequencing was performed using one SMRTcell 8M per pair of barcoded samples on a Sequel IIe system in the UofL Sequencing Technology Center. Following data generation, multiple, iterative sequences covering a single SMRTbell molecule were collapsed to generate highly accurate circular consensus sequence (CCS) reads, followed by analysis using the IsoSeq 3 tool in the SMRTLink software suite (v10.1). CCS reads were further filtered on those CCS reads with quality > 99%, producing "HiFi" reads. The IsoSeq 3 pipeline demultiplexed HiFi reads per individual samples from the pooled sequencing data, filters out amplification artifacts, trims primers and polyA tails, and produces *de novo* FLNC transcripts for downstream mapping and annotation.

Tertiary Iso-Seq data analysis and isoform characterization

A custom pipeline was developed to integrate steps in the cDNA Cupcake protocol for post-processing of Iso-Seq v3 clustered FLNC reads (https://github.com/Magdoll/cDNA_Cupcake). The alignment files for the FLNC reads were converted to fastq (198) and fasta format. Fasta files were aligned to the *Homo sapiens* reference genome assembly (hg38) using minimap2 v2.18-r1015 (199). The resulting alignment files were collapsed into isoforms based on sequence similarity using the cDNA Cupcake Python script collapse_isoforms_by_sam.py. Isoform abundance was calculated using get_abundance_post_collapse.py, and isoforms were filtered to include those with at least two supporting reads with filter_by_count.py. Isoforms were also filtered to remove possible non-full length reads by removing those with indications of a degraded 5' prime region, i.e., apparent 5' shortened isoforms with otherwise equivalent long reads (filter_away_subset.py). SQANTI3 (192) was used to summarize the results for each of the individual samples.

RNA-Seq library preparation and sequencing

RNA-Seq libraries were prepared using the Universal Plus mRNA-Seq with NuQuant (NuGEN, cat no. MO1485). For each sample, 100ng of RNA (in a volume of 50ul) was used for poly A enrichment. First and second strand cDNA was synthesized followed by adapter and unique index ligation. Samples were barcoded using the Universal Plus (UDI) 96-Plex Adaptor Plate (NuGEN, cat no. S02480). The concentration of each library was measured using a Qubit dsDNA HS kit (Thermo Fisher Cat# Q32854). The correct size of each library was confirmed by Agilent Bioanalyzer

analysis using the DNA High Sensitivity Kit (Agilent Technologies, cat no. 5067-4626). In addition, correct adapter and index ligation, as well as the library concentration, was validated by sequencing all libraries on a MiSeq Nano Kit V2 300 cycles (Illumina, Cat. No. MS-103-1001).1.5 pM of barcoded library was denatured, and sequencing was performed on an Illumina NextSeq 500 using the NextSeq 500 75 cycles High Output Kit v2.5 (Illumina, cat no. 20024906). One single-end 75 bp read was performed for each sample. The MiSeq quality control run, as well as the data run on the Illumina NextSeq 500 yielded reads with a data quality of 94.9% of reads at or above Q30. The read alignment rate to the human reference genome Hg38 was 98.6% or higher for each sample.

RNA-Seq data analysis

The quality of the sequenced reads was assessed using FastQC (v.0.10.1, 200), which indicated high-quality reads such that no sequence trimming was necessary. The sequences were directly aligned to the *Homo sapiens* reference genome assembly (hg38) using the STAR aligner (v2.6, 201) two-pass method with Gencode (v39) annotations (133). When used to filter Iso-Seq long reads, RNA-seq short reads were restricted to uniquely mapped genomic positions by setting the STAR aligner option outFilterMultimapNmax to 1.

Integration of Iso-Seq and RNA-Seq data

To assess short-read coverage across the splice junctions identified in the long-read data, short reads were realigned using the STAR two-pass method with the reference

GTF produced by the cDNA Cupcake analysis of the Iso-Seq results. The resulting SJ.out.tab file was provided as input to SQANTI3 (v4.0) which produced a file with short read coverage across each splice junction in the long reads (junctions.txt). Additional SQANTI3 input included the cDNA Cupcake GTF, Gencode (v39) annotations, the hg38 assembly, long read transcript abundance, a polyA motif list, and a CAGE peak (TSS sites) BED file for human and mouse. A custom script was written to identify isoforms with at least one short read covering all internal splice junctions (not including the first and last junction) using the junction file produced by SQANTI3. Coverage of only internal junctions was used for filtering since we saw a pattern of noticeably less short read coverage of long-read transcript ends, particularly 5' ends, similar to as previously observed by others (192).

Transcripts not meeting the filtering criteria were removed from the GTF, resulting in a filtered GTF with high-confidence transcripts based on short read coverage. The filtered GTF was then annotated with SQANTI3. To make visual comparisons across the two methods, UCSC Genome Browser tracks were created to explore expression across the genome (202). The custom tracks for Iso-Seq and RNA-Seq data were created using guidelines and utilities available on the UCSC Genome Browser website. Custom tracks are available to view at http://genome.ucsc.edu/cgi-bin/hgHubConnect?hgHub_do_redirect=on&hgHubConnect.remakeTrackHub=on&hgHub_do_firstDb=on&position=chr1:206,903,317-206,921,941&hubUrl=http://162.215.210.70/~tracks/Mitchell_Lymphocyte_Reference_Transcriptomes/hub.txt.

Functional annotation for novel isoforms

Ensembl gene IDs were extracted for all isoforms classified as novel in catalog (NIC) and novel not in catalog (NNC) from the SQANTI3 classification file for each cell type. The IDs were used in functional annotation analysis, performed separately for individual cell types, with gProfiler2 (203). WikiPathways was selected as the primary annotation source. Disease-related pathways were excluded on account of samples being from a healthy donor, and only pathways which returned an adjusted p-value of less than 0.01 in at least one of the cell-type samples were reported. Expression levels were not considered for this analysis.

Data Availability Statement

All Sequencing data sets were deposited to the Gene Expression Omnibus (GEO). For each Iso-Seq human lymphocyte dataset, raw FLNC files were deposited in BAM format alongside processed data files (Iso-Seq only: GSE202328, SuperSeries: GSE202329). The processed files consist of raw gene isoform counts, annotation (GTF) files, and UCSC Genome Browser tracks. The annotation files were obtained prior to and after filtering for RNA-Seq coverage. Raw Illumina RNA-Seq data for each dataset were deposited as fastq.gz files (RNA-Seq only: GSE202327, SuperSeries: GSE202329). Processed RNA-Seq files deposited consist of raw gene counts and UCSC Genome Browser tracks.

CHAPTER IV TABLES

Table 4.1. Viability and purity of samples. Four lymphocyte populations were purified by negative selection and a fraction was set aside for flow cytometric analysis. % Live Cells, percent of events in a live gate amongst non-debris singlets. % Cell Type, percent of each cell type amongst all $CD45^+$ cells; *, below 0.1%. See Table 4.1 for cell surface markers used. Bold text headings and highlighted entries indicate the cell types targeted for purification and sequencing.

Purified Sample Name:	**% Live Cells**	**% Cell Type**						
		CD4⁺ T *(CD4⁺NK T)*	***CD8⁺ T*** *(CD8⁺NK T)*	*CD4⁻ CD8⁻ CD3⁺*	***NK***	***Pan B***	*CD14⁺ CD16⁻ Monocyte*	*CD14⁺ CD16⁺ Monocyte*
CD4⁺ T	**99.1**	**98.3** (0.4)	*	0.9	*	0.6	*	*
CD8⁺ T	**99.2**	*	**91.7** (3.9)	6.8	0.3	1.4	0.1	*
NK	**98.3**	*	*	*	**97.6**	0.8	*	*
Pan B	**91.0**	*	*	0.3	1.5	**95.8**	0.1	*

% Cell Type, percent amongst $CD45^+$ cells; *<0.1%.

Table 4.2. Frequencies of B cell subtypes in each sample. Calculated by flow cytometric analysis of a fraction of cells in each of the purified samples that were sequenced, reported both out of all $CD45^+$ cells within a sample as well as out of the Pan B subset in the sample. * = below cutoff value of 0.1%.

Purified Sample:	**% of $CD45^+$ Cells**			**% of Pan B Cells**		
	CD19⁺ CD43⁻	***CD19⁺ CD43⁺***	***CD19⁻ CD43⁺***	***CD19⁺ CD43⁻***	***CD19⁺ CD43⁺***	***CD19⁻ CD43⁺***
CD4⁺ T	0.3	0.1	0.2	61.6	9.1	29.3
CD8⁺ T	*	0.2	1.2	0.51	11.7	87.8
NK	*	*	0.8	0.76	2.3	96.9
Pan B	88.1	4.4	3.1	92.2	4.6	3.3

*<0.1%.

Table 4.3. Frequencies of T cell subsets within the peripheral blood of the donor selected. Reported as the average percent ± standard error of the mean based on flow cytometric analysis in two prior experiments, three technical replicates per experiment, using the same male donor. Subset percentages are reported out of the total $CD4^+$ or $CD8^+$ T cell parent population within peripheral blood mononuclear cells. Tcm = T central memory; Tem = T effector memory.

T Cell Type:	% Subset			
	***Naïve* ($CD62L^+$ $CD45R0^-$)**	***Tcm* ($CD62L^+$ $CD45R0^+$)**	***Tem* ($CD62L^-$ $CD45R0^+$)**	**$CD62L^-$ $CD45R0^-$**
$CD4^+$	**56.1** ± 3.6	**34.2** ± 3.7	**9.1** ± 0.1	**0.6** ± 0.3
$CD8^+$	**36.8** ± 0.8	**21.3** ± 2	**34.1** ± 0.5	**7.8** ± 1.7

% Subset, percent amongst $CD4^+$ or $CD8^+$ T cell parent population; mean ± std error

Table 4.4. Total RNA-Seq Reads and Alignment. Multi-mapped parameters allow reads to map to one or multiple places across the genome. Uniquely mapped parameters allow only mapping to a single locus.

Sample	***Total Reads***	***Multi-Mapped***		***Uniquely Mapped***	
		Aligned Reads	**Alignment Rate**	**Aligned Reads**	**Alignment Rate**
$CD4^+$ T cells	102,630,137	101,328,169	98.73%	86,939,812	84.71%
$CD8^+$ T cells	124,630,896	123,279,274	98.91%	108,717,331	87.23%
NK cells	114,296,338	113,001,142	98.86%	101,944,728	89.19%
Pan B cells	112,101,017	110,586,693	98.64%	95,317,079	85.02%

Table 4.5. Number of FLNCs, clusters, and non-zero transcripts for Iso-Seq.

Sample	*SMRTcell Pool*	*Quality Loading Efficiency (P1)*	*Read Length N50 (bp)*	*Number of HiFi Reads*	*Mean HiFi Read Quality*	*Number of FLNC*
CD4$^+$ T cells	**1**	65.7%	120,750	2,805,160	Q38 (99.98%)	1,495,624
CD8$^+$ T cells						976,691
NK cells	**2**	93.1%	113,750	4,325,837	Q38 (99.98%)	2,179,981
Pan B cells						2,091,657

Table 4.6. Number of genes and isoforms detected by Iso-Seq.

Sample	*Unique Genes*	*Unique Isoforms*	*Annotated Genes*	*Novel Genes*	*Novel Transcripts*
CD4$^+$ T cells	11,973	44,159	10,962	1,011	22,248
CD8$^+$ T cells	11,542	34,211	10,402	1,140	15,433
NK cells	12,176	50,905	11,215	961	27,119
Pan B cells	14,487	59,845	12,162	2,325	33,918

Table 4.7. Number of genes and isoforms detected by Iso-Seq after RNA-Seq filtering.

Sample	*Unique Genes*	*Unique Isoforms*	*Annotated Genes*	*Novel Genes*	*Novel Transcripts*
CD4$^+$ T cells	7,797	23,524	6,904	893	10,647
CD8$^+$ T cells	8,417	22,304	7,326	1,091	9,380
NK cells	9,052	32,555	8,178	874	15,730
Pan B cells	10,649	34,858	8,488	2,161	18,496

Table 4.8. Number of transcripts identified in each structural category after filtering by short-read coverage. FSM, full splice match; ISM, incomplete splice match; NIC, novel in catalog; NNC, novel, not in catalog.

Sample	*FSM*	*ISM*	*NIC*	*NNC*	*Genic Genomic*	*Intergenic*	*Antisense*	*Fusion*	*Genic Intron*	*Intron Retention**
CD4$^+$ T cells	10,998	1,879	4,524	4,406	603	544	393	177	0	2103
CD8$^+$ T cells	11,066	1,858	4,296	3,290	494	720	399	181	0	1757
NK cells	14,254	2,571	7,281	6,642	663	553	365	225	0	3075
Pan B cells	13,359	3,003	6,492	6,283	998	1,510	742	2,470	1	3189

** Intron Retention subcategory events were tabulated from within ISM, NIC, NNC, and Fusion categories*

Table 4.9. Antibody markers used for cell-type identification. A fraction of each purified sample was stained with antibodies specific for the myeloid and lymphocyte lineage markers listed. Bolded cell types are the lymphocyte subsets purified. Cell types were defined within the samples by flow cytometric analysis as follows: +, the presence of the lineage marker was required; -, cells with the lineage marker were excluded from the group; +*, either the presence of one or both markers was required. No symbol entry means the marker was not considered in defining the cell type.

Supplier, Catalog #	Marker Specificity	Cell Type Defined								
		CD4$^+$ T	*CD4$^+$ NK T*	***CD8$^+$ T***	*CD8$^+$ NK T*	*CD4$^-$ CD8$^-$ T*	***NK***	***Pan B***	*CD14$^+$CD16$^-$ Monocyte*	*CD14$^+$CD16$^+$ Monocyte*
BD, 560367	**CD45**	+	+	+	+	+	+	+	+	+
ThermoFisher, 58-0038-42	**CD3**	+	+	+	+	+	-	-		
BioLegend, 300534	**CD4**	+				-		-		
ThermoFisher, 64-0088-42	**CD8**			+	+	-		-		
BD, 564057	**CD56**		+		+		+	-		
BD, 563522	**CD43**							+*		
ThermoFisher, 69-0199-42	**CD19**							+*		
Tonbo, 20-0149-T100	**CD14**							-	+	+
BD, 555408	**CD16**							-	-	+

*Pan B cells were defined as positive for CD19 or CD43 or both, and negative for myeloid, T, and NK lineage markers.

CHAPTER IV FIGURES

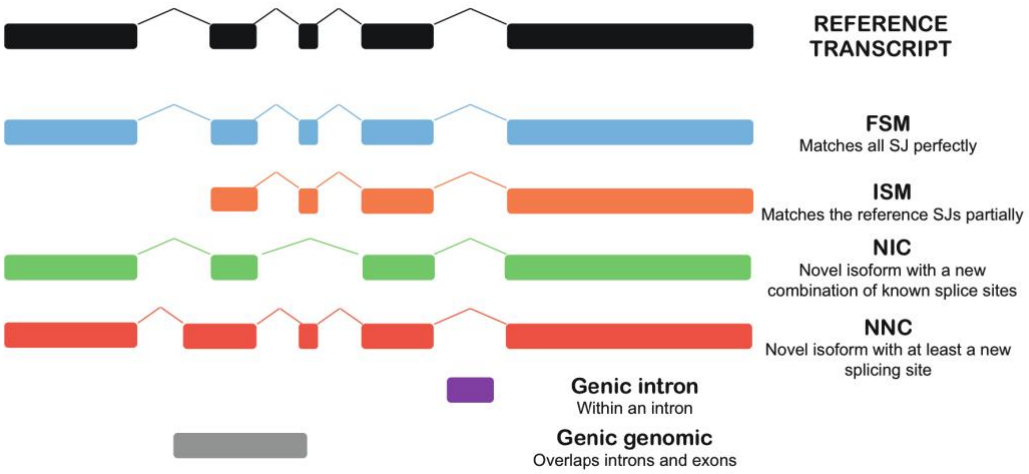


Figure 4.1. Schematic of SQANTI3 isoform structural characterization

(image taken from https://github.com/ConesaLab/SQANTI3/wiki/SQANTI3-isoform-classification:-categories-and-subcategories). FSM, full splice match; ISM, incomplete splice match; NIC, novel in catalog; NNC, novel, not in catalog. Intron retention events are not depicted but would fall within ISM, NIC, NNC, or fusion categories.

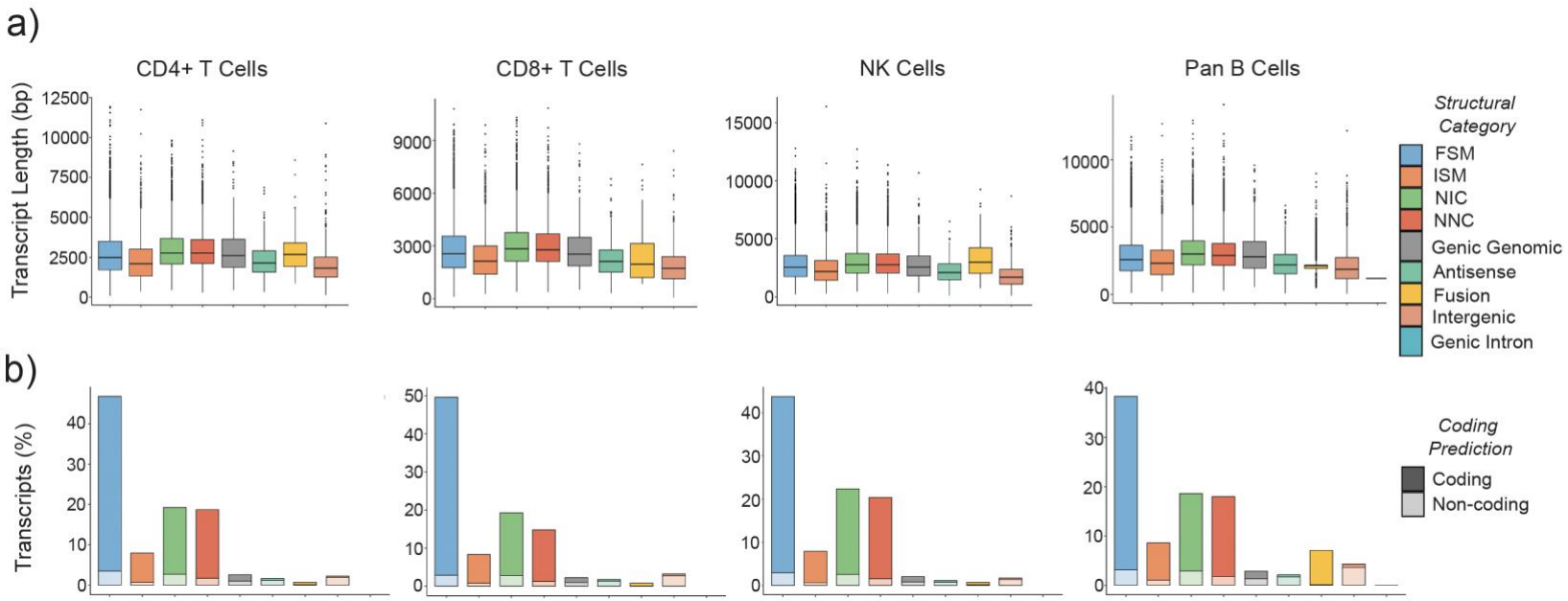


Figure 4.2. Characteristics of transcripts within each structural category. From left to right, each graph represents a different lymphocyte subset. **(a)** Box-and-whisker plots depict the distribution of transcript lengths (bp) for each transcript structural category. Center lines of the plots represent median transcript lengths. **(b)** Bar graphs depict the proportions of transcripts predicted to be protein-coding. Bar height indicates the percent of transcripts within each structural category out of the total transcripts present in the sample. Darker or lighter shading within a bar represents the proportion of the category predicted to be protein-coding or non-coding, respectively.

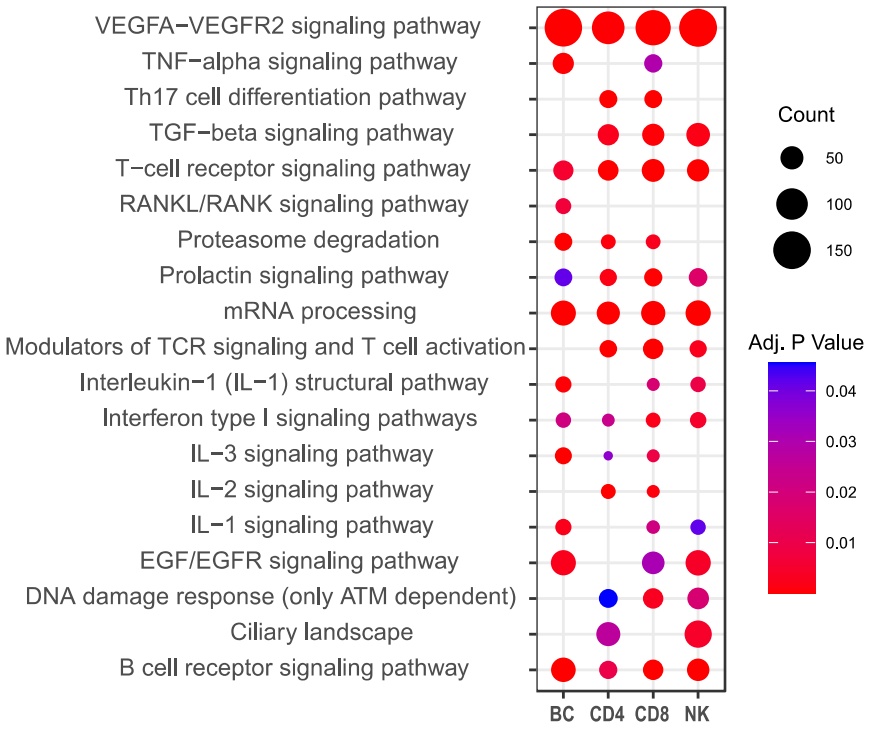


Figure 4.3. WikiPathways enriched for novel isoforms. Dot plot depicting gene-level functional classification and enrichment for novel isoforms in each lymphocyte subset. The size of the dot represents gene count within a category while heatmap-based coloration indicates p-value adjusted for false positive expectation. Only non-disease pathways that returned an adjusted p-value of less than 0.01 in at least one of the cell-type samples are depicted.

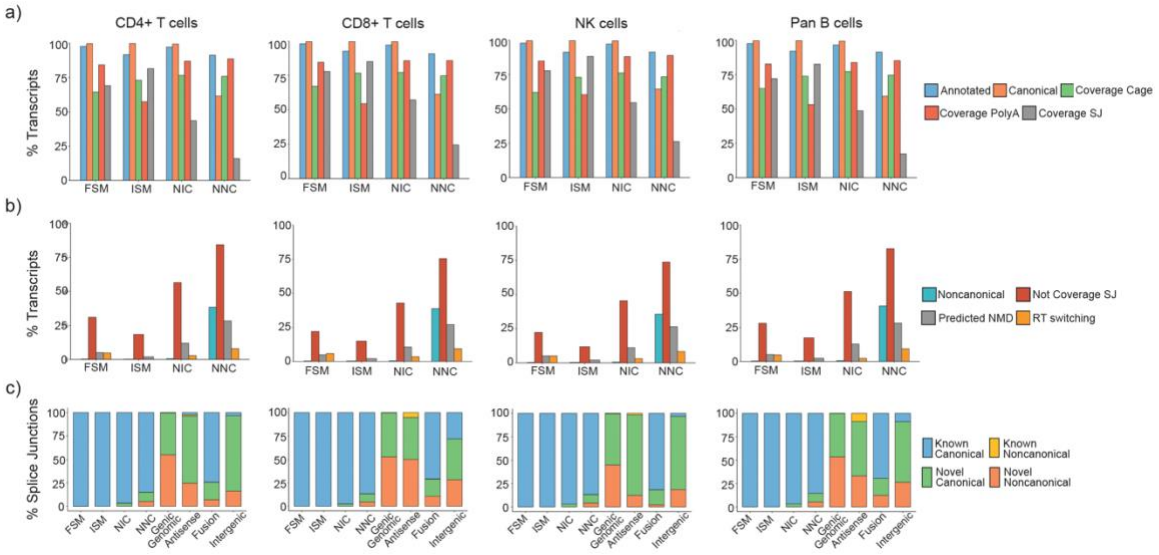


Figure 4.4. Quality control metrics within each isoform sample dataset. From left to right, each graph represents a different lymphocyte subset. **(a)** Metrics of good quality, the percent of transcripts within the noted categories which have: Annotated, mapped to annotated genes; Canonical, all canonical splice junctions; Coverage Cage, an identified CAGE Peak; Coverage PolyA, an identified polyA motif; Coverage SJ, supporting short-read coverage of all splice junctions. **(b)** Metrics of bad quality, the percent of transcripts within the noted categories which have: Noncanonical, at least one noncanonical splice junction; Not Coverage SJ, no short-read coverage of at least one splice junction; Predicted NMD, predicted nonsense mediated decay; RT switching, predicted RT switching occurrence. **(c)** Proportions of all splice junctions identified which are: known canonical, known noncanonical, novel canonical, or novel noncanonical.

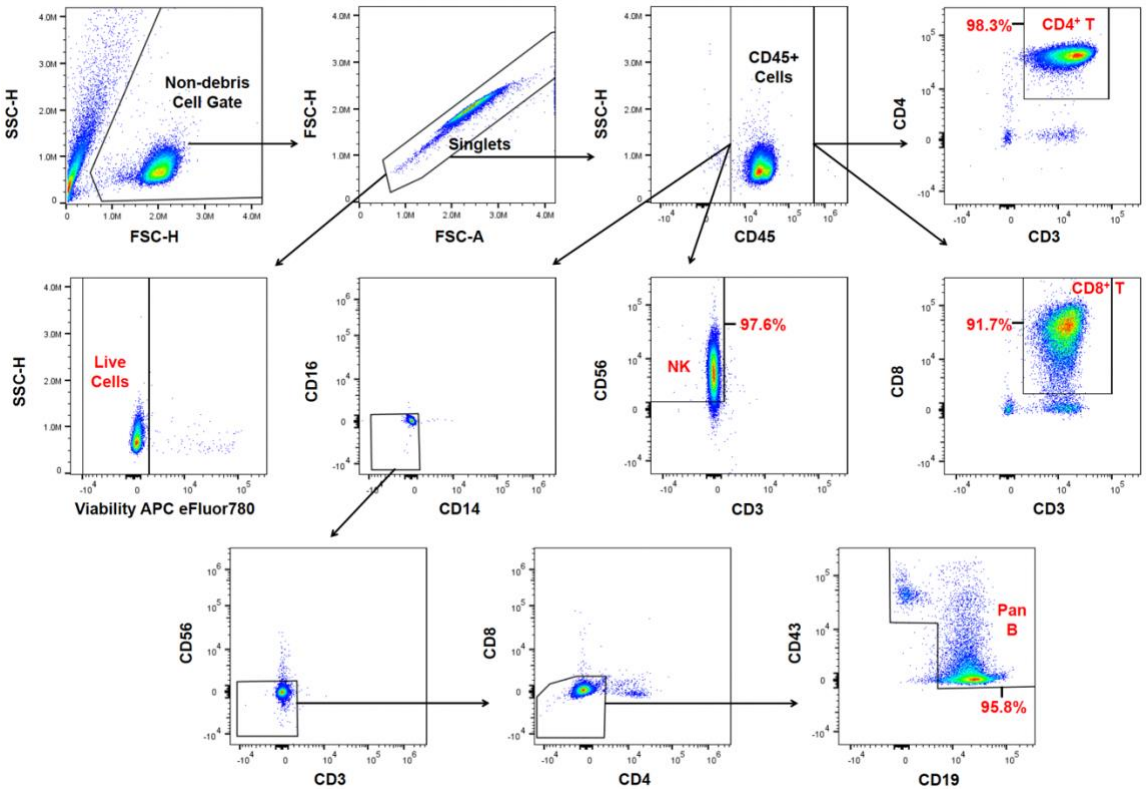


Figure 4.5. Gating strategy used to define live cells and major subsets within each purified cell sample. A fraction of each purified sample was stained with makers specific for myeloid and lymphoid lineages and the gating strategy depicted was used to define live cells, $CD4^+$ T cells, $CD8^+$ T cells, NK cells, and Pan B cells. Inset percentages in red show the final purity of each respective lymphocyte population. Dot plots are representative of gating used upstream of these populations. In calculating viability, the fraction of live cells of total singlets was used. In calculating purity of the desired cell type, the fraction of the corresponding cell type of total $CD45^+$ cells was used.

CHAPTER V

FULL-LENGTH MRNA SEQUENCING RESOLVES 5' END-VARIANT ISOFORMS IN ACTIVATED HUMAN CD4 T CELLS

INTRODUCTION

Only about 40% of the variation in human protein expression can be explained by changes to mRNA levels alone (90). To fully comprehend differential expression across both physiologic and disease states, we must recognize mRNA-intrinsic mechanisms of regulation occurring among isoform variants. Pacific Biosciences Isoform Sequencing (Iso-Seq) has revolutionized our understanding of isoform variation by utilizing single-molecule real-time (SMRT) technology to generate full-length reads of mRNA. In this method, circularized cDNA allows multiple sequencing passes which are ultimately collapsed into highly accurate consensus reads (94, 134). After primary analysis and cleanup, Iso-Seq can achieve > 99.9% sequencing accuracy of complete mRNA molecules (127, 134, 136).

Full-length transcriptomes generated by bulk Iso-Seq can be useful tools to provide isoform references across different cell types and conditions. A comprehensive reference may be especially beneficial for mapping short-read data during isoform prediction, as in the recently developed Scasa method (185). Though single-cell isoform sequencing is also now available, its lower sequencing depth may neglect low-expressed

isoforms (186-189). In comparison, bulk Iso-Seq provides the depth necessary to generate a sufficient roadmap of transcript variants and remains unsurpassed in its ability to broadly identify novel isoforms across all transcript lengths, particularly end-variants (190-192). End-variant isoforms, which differ from canonical isoforms through alteration of their 5' or 3' most ends, are often overlooked by RNA-Seq *de novo* reconstruction algorithms that focus primarily on estimating alternatively spliced transcripts (103-105, 137). Yet, end-variants may possess important regulatory elements within differential untranslated regions (UTRs) that can drastically affect protein expression (97, 98, 162, 163).

For human immune cells, regulation of both internal splicing and transcript UTRs are thought to play important roles in controlling expression during anti-cancer immunity, pathogenic response, inflammation, and autoimmune disease (91, 92, 96, 140, 143, 145, 148-150, 178, 179, 182-184). However, few studies have attempted to investigate the global isoform landscape of activated human CD4 T cells and no bulk Iso-Seq reference yet exists (204-206). To begin to address this gap, we activated human CD4 T cells from a female donor and purified RNA (RIN>8) at four-time points (4, 16, 48, and 120 hrs) of activation for parallel Iso-Seq and RNA-Seq analysis. Full-length non-concatemer reads (FLNC) as well as processed Iso-Seq data files have been deposited to the Gene Expression Omnibus (GEO) database (196) alongside matched data from Illumina short-read RNA-Seq that was performed in parallel (GSE229972 SuperSeries). Each time point and its corresponding sequence dataset met several metrics of quality as assessed by flow cytometric and SQANTI 3 (192) analysis, respectively, suggesting the value of these references for future use. Within these datasets, we further classified multiple novel 5'

end-variant isoforms mapping to immune-important genes. One gene, *CXCR5*, was selected for further study to validate two novel isoforms and assess their relevance in altering protein expression kinetics and mRNA stability.

RESULTS AND DISCUSSION

Quality of purification and metrics of activation

Isolated CD4 T cells to be used as input for activation were sampled and found to be viable (99.9%) and highly pure (98.2%) with contaminating cells primarily consisting of $CD4^{-}CD8^{-}CD3^{+}$ T cells (0.82%) (**Figure 5.1**). In accordance with our goal of creating broadly representative references, the negative selection kit catered to enriching diverse circulating subsets. Of the CD4 T cells present, most were naïve (58.3%) or memory (41.4% total Tcm and Tem), with a low fraction of effector (0.31%) and NK T cells (0.007%) (**Figure 5.1**). This allowed our activated references to reflect both new and re-activated CD4 T cells.

Once cells were isolated, they were cultured with anti-CD3/CD28 beads for various time points of activation. Each time point succeeded in achieving nearly complete activation, supported by almost all sampled $CD4^{+}$ cells expressing CD69 and/or CD25 (**Figure 5.2**). Throughout stimulation, cells also remained viable (viability $\geq$95.9%, **Figure 5.2**) and demonstrated expected kinetics for the primary stages of T-cell activation (**Figure 5.2**). Cells initiated upregulation of the activation marker CD69 by 4 hrs and fully transitioned to a blasting phenotype with increased FSC and SSC by 48 hrs (**Figure 5.2**). Based on comparative calculations of cells per staining well, we further found that cells had not yet doubled at 48 hrs but by 120 hrs had expanded to over six

times the estimated input cell count. These metrics of cellular phenotype and behavior suggest that the references presented are high-quality representations of early activated (4 and 16 hrs), blasting (48 hr), and proliferating (120 hr) CD4 T cells.

RNA-Seq read metrics and genomic mapping

RNA sequencing generated ~58-70 million reads per activation timepoint with alignment rates > 92.9% when considering reads that were uniquely mapped to a single locus (**Table 5.1**). Between 91.5% and 92.0% of sequenced reads were at or above Q30, and sample data mean quality scores were all above Q35 (>99.97%). High alignment and good quality metrics suggest these short-read data may serve as a useful counterpart to each long-read dataset for future analyses.

Iso-Seq initial sequencing data metrics

Isoform sequencing was performed with two pooled samples per 8M SMRTcell (4 and 16 hrs, 48 and 120 hrs). Loading, read length, and yield metrics fell within the expected ranges for both SMRTcells (**Table 5.2**). For all sequencing data, 50% of bases were in reads > 146,000 bp (read length N50), a length allowing > 48 passes of the average 3 kb mRNA. This multi-pass sequencing resulted in highly accurate HiFi reads (≥Q40, 99.99%) that were further processed using primary Iso-Seq analysis to generate ~1-2 million full-length non-concatemer reads (FLNCs) per sample (**Table 5.2**). These FLNCs could then be viewed independently as processed sequence reads or used as input for further downstream analyses.

Transcript annotation and classification of isoforms

To further analyze only the highest confidence and truly unique reads, FLNCs were collapsed and filtered for each sample and a combined collapse of all samples was used to generate a chained analysis group. The chained group was created to assess the sum of unique expression across CD4 T cell activation states, representing a broader, combined reference for this cell state. Resultant transcripts for the chained and per-sample groups were summarized using SQANTI 3 (**Table 5.3**). Unique genes were identified based on the mapped genomic locations of expressed reads, while unique isoforms were categorized as collapsed reads possessing distinct splice junctions and/or distinct transcription start and stop boundaries.

At each activation time point, ~11-15 thousand unique genes and over 39 thousand distinct isoforms were expressed (**Table 5.3)**. When all four samples were considered together, ~18 thousand unique genes were found with over 118 thousand unique isoforms, which is less than the raw total across samples suggesting some overlap of both genes and isoforms throughout CD4 T cell activation (**Table 5.3**). Most genes expressed were previously annotated within Gencode v39. However, between 601 and 3,311 novel genes were found across samples, totaling 4,970 novel genes within the chained sample (**Table 5.3**).

SQANTI3 was utilized to further categorize transcript isoforms into various structural classifications based on their alignment to previously reported reference transcripts and their utilization of known donor and acceptor splice sites. These were either full splice match (FSM), incomplete splice match (ISM), novel in catalog (NIC), novel not in catalog (NNC), intergenic (between annotated genes), antisense (anti-sense

to an annotated gene), fusion (fusion of two annotated genes), genic genomic, or genic intron (**Figure 5.3**). Unexpectedly, transcripts matching their reference completely (FSM) represented only 36-40% of transcripts per sample and were less than a third of unique transcripts identified within the chained analysis group (**Table 5.4**). Rather, a larger percentage of transcripts were novel (e.g., NIC and NNC, **Table 5.4**), supporting the notion that the isoform landscape of activated CD4 T cells has many nuances that remain to be illuminated. Most of these novel transcripts were also predicted to be protein-coding (**Figure 5.4**), further suggesting that, if valid, these transcripts may hold relevant expression-level consequences.

Quality metrics of isoform discovery

The top four isoform categories (FSM, ISM, NIC, NNC) were used to assess metrics of quality via SQANTI 3, with most transcripts reflecting attributes of good quality and few demonstrating potential for bad quality (**Figure 5.5a-b**, respectively). A majority of transcripts mapped to known genes and possessed transcription start and termination sites within annotated cap analysis gene expression (CAGE) peaks or polyadenylation (polyA) motifs, respectively (**Figure 5.5a**). These transcripts also appeared to primarily use canonical splice junction motifs (**Figure 5.5a**), maintaining a high biological capacity for expression. Genuine expression was also supported by the low percentages of transcripts predicted to have undergone reverse transcriptase (RT) switching, a common cause of artifactual novelty (**Figure 5.5b**). For both full and incomplete splice match categories, there were also very few transcripts predicted to have

undergone nonsense-mediated decay (NMD) which would result in incomplete reads (**Figure 5.5b**).

Though still relatively minimal, a fraction of novel transcripts (NIC and NNC) was marked as possible NMD (**Figure 5.5b**). It is important to consider that this NMD flag appears any time an isoform's predicted coding sequence ends at least 50 base pairs before the last junction identified for a transcript (192). Thus, NMD prediction may be artificially inflated by instances in which splicing occurred within 3' UTRs (207, 208) and may not necessarily indicate true decay within these novel categories.

Considered in combination with the quality of purification and subsequent activation achieved, these metrics demonstrate that the datasets provided here consist of representative and high-quality polyadenylated transcripts, a large proportion of which are entirely novel and previously unannotated.

End-variant isoforms within our Iso-Seq datasets

As we examined the novelty of transcripts in our datasets, we noticed that many isoforms were end-variants, meaning they differed from previously annotated references only on their 5' or 3' ends. SQANTI 3 classified some of these end-variant isoforms as subcategories of full splice match transcripts, with alternative 5' or 3' transcripts having novel transcription start sites or termination sites, respectively (**Figure 5.6**). Interestingly, end-variant transcripts made up the majority of the full splice match transcripts identified (**Table 5.5**), suggesting a significant gap in the current annotation of end-variation. While some of these proportions may be inflated due to collapse, particularly for 5' end variants where collapse skews towards longer reads, many of these transcripts were still found

within annotated regions of CAGE Peaks or PolyA motifs, supporting the potential for true expression (**Table 5.5**). It is therefore likely that many of these transcripts represent real events.

End-variant isoforms can alter the regulation of protein expression by modifying elements within the 5' or 3' UTR. While the 3' UTR primarily regulates mRNA stability, the 5' UTR plays an important role in regulating both stability and translational efficiency (97, 98, 162, 163). Therefore, we focused on analyzing 5' end-variants of protein-coding isoforms with longer 5' UTRs than previously annotated for their gene. Longer UTRs are more likely to introduce new regulatory mechanisms and less likely to be sequencing artifacts of processivity or degradation. To classify these isoforms in categories beyond FSM, we developed a custom script that compared the calculated 5' UTR length of each sample isoform to the longest known UTR of its parent gene within MANE (Matched Annotation from NCBI and EMBL-EBI) or Gencode (v39, comprehensive) annotations. Only isoforms that were predicted to be protein-coding, mapped to known genes, and were within the FSM, ISM, NNC, or NIC categories were considered for this analysis.

To validate the use of this method for 5’ UTR comparisons, we first analyzed the distribution of lengths compared to MANE Select transcripts (209). Each MANE Select transcript is a singular transcript created from a convergence of the two major transcriptome annotation references, NCBI and EMBL-EBI, that is representative of biology at its gene locus (209). Based on the assumption that most transcripts expressed are MANE and MANE 5’ annotations are representative, we would expect these differences to cluster around zero with a relatively normal distribution of shortened or lengthened UTRs. Yet, like the findings of SQANTI-classified end-variants, large

proportions of the isoforms analyzed within every sample were found to possess 5' UTRs ≥ 10bp longer than the MANE transcript of their parent gene (**Table 5.6**). Far fewer transcripts were predicted to possess UTRs lengthened by ≥ 100bp (**Table 5.6**), which was further visualized in the overall distribution of lengths compared to MANE UTRs where the modes centered around zero with only a slight skewing towards lengthened reads (**Figure 5.7a**). It is possible this skewing in part reflects an artifact of the 5' collapse to longer reads, as processed by cDNA Cupcake to minimize the impact of 5' degraded reads. However, as previously mentioned for the SQANTI classified end variants, this does not mean these reads were not real but instead suggests that lengthened reads may be slightly overrepresented in the processed datasets. It is also possible that the skewing towards longer reads is actually a true reflection of the UTRs present in the transcripts, contrary to our initial assumptions. This could be because MANE transcripts are not always the predominant transcripts in every cell state, and current annotations for 5' UTRs, even in the MANE set, may be incomplete. We thus continued to compare our transcript 5' UTRs to the longest previously annotated UTRs with an understood caveat that some lengthened reads may be overrepresented.

As expected, far fewer reads had 5' UTRs longer than the longest known UTR of any Gencode annotated transcript for their parent gene (**Table 5.6**). This was also shown in distributions that were similarly centered around zero but heavily skewed towards shorter lengths (**Figure 5.7b**). In further classifying these lengthened UTRs for the chained sample, all structural categories appeared to retain similar proportions of transcripts with novel UTR ≥ 10bp longer, while the NIC and NNC categories had a higher proportion of transcripts with UTR ≥100bp longer (**Table 5.7**). An even higher

proportion of intron retention events within the ISM, NIC, and NNC categories resulted in lengthened 5'UTR (**Table 5.7**), though these events did not dominate because the sum of intron retention events was far less than the sum of events within their parent categories. These results suggest novel lengthened UTRs are represented in a diverse set of transcripts generated from multiple mechanisms of isoform variation. Even with some element of 5' collapse artifact, Iso-Seq appears to have illuminated novel additions to 5' UTRs that could be important to characterize, particularly for enhancing further understanding of UTR deviation in instances of pathology or disease response (91, 92, 148, 149, 183).

Immune-important genes with novel end-variant isoforms

To support the potential relevance of these lengthened 5' UTRs, we looked for variant transcripts within immune-important genes and discovered, among other examples, variants of *CXCR3*, *CXCR5*, and *IL7R* with lengthened 5' UTRs (**Figure 5.8a-c**). Both end-variants of *CXCR5* and two of the three variants of *CXCR3* retained all canonical junctions and the canonical CDS (**Figure 5.8a-b**), suggesting the added UTR length could serve a predominantly regulatory role. For IL7R, the dominant end-variant transcripts did not retain the canonical CDS and instead skipped canonical exon 1, causing a frameshift and a predicted N-terminal truncation of the IL7R protein **(Figure 5.8c**). UTRscan of these exon-skipping transcripts suggested the transcription start of this truncated CDS was downstream of a predicted internal ribosomal entry (IRES) site (97), supporting the biological potential for protein production. These examples, though

unvalidated, demonstrate instances where 5' UTR variation could hold relevance for genes important to the activated CD4 T cell state.

Of these examples, we became particularly interested in the two *CXCR5* isoform variants due to the importance of this gene in T helper cell differentiation (79, 210, 211). Furthermore, we noticed an apparent enrichment in FLNC matching the novel *CXCR5* end-variants at later activation time points (**Figure 5.9a**). The TSS of these novel isoforms mapped to a known promoter that had not previously been associated with any *CXCR5* transcripts in the Eukaryotic Promoter Database (v006) (212) (**Figure 5.8b**). These isoforms also appeared to have a high potential to confer intrinsic regulation of CXCR5 expression through novel 5' UTR elements (**Figure 5.9b**). As only one 5' UTR had been previously documented for *CXCR5* (97, 98), and given the potential importance of these novel *CXCR5* end-variants, we decided to validate these isoforms as an example of the significance of discovering such variants using Iso-Seq.

Evidence of novel CXCR5 end-variants in other donors and datasets

In order to assess the presence of these isoforms in other donors and investigate for CD4 T cell subset preference, we probed for evidence of these isoforms in previously deposited RNA-Seq datasets (**Table 5.8**). Using the isoforms present in our activated references as a guide, we determined the genomic location of the novel 5' junctions of our isoforms (Novel Isoform 1: Chr11:118882935,118883851 and Novel Isoform 2: Chr11:118882935,118883748) and used these to assess junctional coverage, or uniquely mapped reads spanning each junction, within previous RNA-Seq data. Coverage of the shared internal *CXCR5* junction (reads spanning Chr11:118883993,118893595) was used

as a metric of total *CXCR5* expression. Total junctional coverage across the two novel junctions was then divided by coverage across this shared internal junction to estimate per sample fractions of expressed *CXCR5* corresponding to our novel isoforms. The average of approximated isoform expression could then be calculated for samples across different cell subsets and states (**Table 5.9**).

Of the 216 unique donors considered to express *CXCR5*, 166 were found to possess 2 or more junction-spanning reads across our novel *CXCR5* isoforms. Interestingly, our novel isoforms were primarily found among activated subsets of CD4 T cells, particularly those which would be classified as early proliferating (72-120 hrs activation) (**Figure 5.10a-b**). This aligns with our own observations of novel isoform frequency increasing with activation time to 120 hrs and relatively aligns with a pattern of transient CXCR5 protein expression known to occur during early CD4 T cell activation (27, 28). This correlation with activation appeared in all CD4 T cell subsets except in resting and in activated T regulatory cells (**Figure 5.10b**).

We further investigated the frequencies of novel isoforms occurring in resting T follicular helper (Tfh) cell subsets. Lymphoid resident Tfh cells, particularly germinal center Tfh (GC Tfh), express high levels of CXCR5 protein to encourage homing to follicular germinal centers (79, 210, 211). Corresponding high and stable *CXCR5* mRNA expression is achieved through disinhibition of transcription at the canonical *CXCR5* promoter (77, 78, 79, 210). Thus, we did not expect to find our novel isoforms at high frequencies in lymphoid-resident Tfh. However, we considered that our isoforms might play a role in CXCR5 expression within circulating T follicular helper (cTfh) cells, which are thought to express low tonic levels of *CXCR5* under non-traditional transcriptional

control (77). Contrary to this hypothesis, we found low frequencies of the novel isoforms in all resting Tfh, including cTfh (**Figure 5.10c**), suggesting that these novel end-variant isoforms are not the primary transcripts responsible for stable *CXCR5* expression within any of these differentiated cell subsets. Instead, these novel end-variant isoforms might play a part in *CXCR5* expression regulation prior to Tfh commitment. This is supported by the observation of higher novel isoform frequencies in *ex vivo* generated transitional Tfh cells (**Figure 5.10c**). To explore the regulatory potential of these isoforms, we next considered the effects of isoform-specific variation on mRNA stability and protein expression.

A novel CXCR5 transcript is an unstable, low-expressing, protein-invariant isoform

Plasmids expressing each isoform variant were generated with a standardized *CXCR5* CDS and 3' UTR (**Figure 5.11** and **Table 5.10**). These plasmids were then transfected into HEK293T cells to force the unique expression of each *CXCR5* mRNA variant.

To investigate the mRNA stability of each variant and calculate an estimated half-life in this idealized system, we performed an Actinomycin D (ActD) transcription inhibition assay of transfected HEK293T. *FOXO3* was used as an internal control unstable RNA (213), confirming assay success (**Figure 5.12a**). Using this assay, we observed that both variants of *CXCR5* had altered decay curves compared to their canonical counterpart, and for Isoform 2 the half-life of its transcripts was significantly decreased compared to the canonical transcripts (**Figure 5.21b**, $t_{1/2}$ 95% CI shown on

graph). This suggested that the 5' UTR additions were sufficient to confer decreased transcript stability.

We next investigated if either novel variant exhibited altered kinetics of protein expression. We did this using a synchronized transfection system where transfection was allowed to proceed unhindered for 3 hours prior to the removal of transfection reagents followed by flow cytometric measurement of CXCR5 expression at various timepoints thereafter (**Figure 5.12c**). This synchronized transfection approach allowed the continued use of the CMV overexpression system with preservation of relative efficiency (214, 215) while simultaneously preventing an overload of the system with *CXCR5* mRNA that would hinder the observation of more nuanced differences in expression.

Using this approach, we observed significantly altered CXCR5 expression at all time points for Isoform 2 compared to the Canonical Isoform (**Figure 5.12c**), suggesting novel mechanisms of mRNA-intrinsic translational regulation occurring. Interestingly, Isoform 1 did not show a difference in protein expression compared to the Canonical Isoform (**Figure 5.12c**), distinct from apparent differences in mRNA stability (**Figure 5.12b**). We believe this may be in part due to the CMV overexpression system allowing for sustained availability of mRNA which masked changes in expression caused by lower RNA stability. However, Isoform 2 revealed significantly lower protein expression even when overexpressed. Thus, we hypothesized that elements present in the 5' UTR of Novel Isoform 2 could contribute to its decreased ability to express CXCR5 protein.

Specific elements of the novel CXCR5 5' UTR decrease the efficiency of protein expression

In previously scanning the 5' UTR of each novel isoform for regulatory elements, we found Novel Isoform 2 contained multiple uORFs and an upstream adenine-rich (A-rich) region (**Figure 5.9c**). To test whether these elements caused differences in expression, we generated plasmids expressing mutant variants of Isoform 2 lacking uORFs or the A-rich region. To ensure the complete negation of uORF activity, we altered the sequence of all three upstream start codons possessing an in-frame stop (Isoform 2 w/o uORFs, **Table 5.10**). Separately, in removing the A-rich region, we kept only the minimum A-repeat necessary to retain the frame of upstream elements present with respect to the downstream CXCR5 coding sequence (Isoform 2 w/o A-rich**, Table 5.10**). This ultimately produced two separate mutants with minimally altered sequences. We then tested these mutants' expression kinetics against the Canonical Isoform and unaltered Novel Isoform 2.

When the A-rich region of Isoform 2 was removed, we observed a late but nonsignificant divergence of CXCR5 expression from the unaltered isoform which trended towards a similar expression to the Canonical Isoform (**Figure 5.12d**). This suggests that the A-rich element does not play a primary role in inhibiting CXCR5 translation. However, this variant may still play a small role in mRNA intrinsic regulation that remains to be further investigated, perhaps only in altering mRNA stability which seems less easily visualized in this system.

Removal of all uORF regions, on the other hand, rescued CXCR5 expression by Isoform 2-derived constructs to levels mirroring those of the Canonical Isoform (**Figure**

5.12e). CXCR5 expression was significantly increased compared to unaltered Isoform 2 at all but the 3 hr time points (**Figure 5.12e**). This pattern suggests that intact uORF regions are the primary elements responsible for decreased expression of CXCR5 protein by Novel Isoform 2.

Previous studies have shown the presence of uORFs can decrease transcript expression by 30-80% (173) through mechanisms such as ribosomal stalling or by encoding small regulatory peptides that independently affect translation (163, 171, 172, 174, 176, 177). Though we confirmed expression hindrance from the uORFs within Novel Isoform 2, further investigation is needed to explore the impact of the single uORF within Novel Isoform 1. Novel Isoform 1 did not appear to directly regulate CXCR5 protein expression in our study, but our investigations were conducted in an overexpression system that likely masks some effects. Thus, while we may not have observed the full range of expression differences that would be present in an activated CD4 T cell system where the isoforms are naturally expressed, the experiments with Novel Isoform 2-derived constructs demonstrated a functional consequence for protein expression was conferred by uORF regions.

Ultimately, our investigation into these novel CXCR5 end-variants emphasizes the significance of isoform-level differences that may be overlooked despite evidence of expression in RNA-Seq datasets. Many such novel isoforms, discovered using Iso-Seq, may have important expression-level consequences and should not be excluded from transcriptome annotations. The high-quality isoform references provided here will help fill gaps in the known transcriptome of activated CD4 T cells and enhance our understanding of the isoform landscape to guide future differential expression analyses.

MATERIALS AND METHODS

Human blood collection and CD4 T cell isolation

Using standard phlebotomy, venous blood, 100 mL, from a healthy, consented 24-year-old female donor, was collected in and adjusted to 6mM K3EDTA. Blood collection was approved by the University of Louisville Institutional Review Board under expedited review (IRB 14.0661).

PBMCs were isolated from blood using Sepmate™ PBMC Isolation Tubes (StemCell Technologies, cat no. 85450) as directed by the manufacturer. Cell yield was determined using the count per μL feature of a Cytek® Northern Lights flow cytometer as previously described (56), immediately after which the PBMCs were further processed to isolate CD4 T cells.

CD4 T cells were purified from freshly prepared PBMC via negative magnetic selection using EasySep™ Human $CD4^+$ T Cell Isolation Kit (StemCell Technologies, cat no. 17952) per manufacturer's instructions. A fraction of these cells was used for flow cytometric staining of markers to assess cell viability and purity as described below. The remaining isolated cells were immediately plated for activation.

CD4 T cell culture and activation

Isolated CD4 T cells were plated in a 96-well U-bottom plate at 0.3 million (for 4, 16, or 48 hr activation) or 0.1 million (for 120 hr activation) cells per well in complete RPMI. Complete RPMI was generated using RPMI 1640 media (Thermo Fisher Scientific, cat no. 21870-076) with the addition of 1X Glutamax (Thermo Fisher Scientific, cat no. 35050-061), 100 U/mL penicillin and 100 μg/mL streptomycin

(Thermo Fisher Scientific, cat no. 15140-122), and 10% sterile-filtered, heat-inactivated male AB serum (Sigma-Aldrich, H3667). Cells were activated for the indicated times using Dynabeads Human T-Activator CD3/CD28 (Thermo Fisher Scientific, cat no. 11161D) according to the manufacturer's instructions. Cells were plated with a bead:cell ratio of 1:1 for 4, 16, or 48 hr activation, or 3:1 for 120 hr activation. No additional cytokines or stimulants were added during this time. Cells stimulated for 120 hrs were split 1:1 on day 2 of stimulation with 50% new complete RPMI added to prevent overgrowth and media exhaustion.

At the indicated time points, approximately 3 million (4, 16, and 120 hrs) or 1.8 million (48 hours) activated cells were lysed in-well with Buffer RLT Plus (RNeasy® Plus Mini Kit, Qiagen, cat no. 74134), per the manufacturer's protocol. Lysates were then transferred to a microfuge tube and Dynabeads were removed before immediately proceeding with RNA extraction. One well of cells per time point was also used for flow cytometric staining to assess the success of activation as described below.

CXCR5 Plasmid Preparations

All plasmids were produced through the services of BlueHeron Bio (Bothell, WA) using standard synthesis and cloning techniques. A plasmid expressing CXCR5 in the pCMV6-XL4 backbone (OriGene, cat no. SC309454) was used as a base. Three plasmids expressing a canonical CXCR5 mRNA isoform or either of two novel 5' end-variant isoforms were created (**Figure 5.11**). Insert sequences differed only in their 5' UTR region and used a standard CXCR5 coding sequence and 3' UTR (**Table 5.10**, CDS and 3' UTR from NM_001716.2). Later, two additional mutant variants of the Novel Isoform

2 were generated in the same backbone, again differing only from their parent only in the 5' UTR region (**Table 5.10**). All inserts were sequence-verified by BluHeron Bio.

Plasmids were maintained in GC10 *E. coli* (containing the canonical transcript sequence), DH10B *E. coli* (Isoform 1, Isoform 2, and Isoform 2 w/o A-rich), or Stbl3 *E. coli* (Isoform 2 w/o uORF). For each plasmid, multiple independent DNA preparations were performed. All plasmid DNA was purified using ZymoPURE II Plasmid Midiprep (Zymogen, cat no. D4201) or Maxiprep (Zymogen, cat no. D4202) kits per the manufacturer's instructions with optional endotoxin removal included. Plasmid DNA was eluted in DNA-ase-free water and stored at -20^{o}C until use in transfection experiments.

HEK293T cell culture and plasmid transfection

HEK293T cells were maintained in complete DMEM (Thermo Fisher Scientific, cat no. 10569-010) supplemented with 10% heat-inactivated fetal bovine serum (Thermo Fisher Scientific) and 0.1 mg/mL Normocin (InvivoGen, cat no. ant-nr-1). Cells were passaged to achieve 50-80% confluency on the day of transfection, and directly prior to transfection media was replaced with Opti-MEM® reduced serum media (Thermo Fisher Scientific, cat no. 31985-070). All transfections were performed using the Lipofectamine 3000 kit (Thermo Fisher Scientific, cat no. L3000-008) according to the manufacturer's recommended protocol.

For RNA stability experiments, transfection was performed with a pooled mixture of preps and allowed to proceed for a full 20 hours of incubation to achieve a standardized maximum transfection. RNA stability was assessed using an Actinomycin D (ActD) transcription inhibition assay based on a previously described protocol (216).

Briefly, after 20 hours of transfection, the media was replaced with complete DMEM containing 5 μg/mL of ActD. At 0, 1, 2, 4, 8, 12, and 24 hours post-ActD addition, cells were lysed in-well with Buffer RLT Plus (RNeasy® Plus Mini Kit, Qiagen, cat no. 74134), per the manufacturer's protocol. Lysates were then transferred to a microfuge tube and kept at -80^{o}C until same-day RNA extraction and cDNA synthesis as described below.

For protein expression experiments, transfections were performed with independent plasmid preps and allowed to proceed for 3 hours prior to the replacement of media with complete DMEM. This synchronized transfection allowed a clearer assessment of surface expression kinetics while maintaining a minimum time of transfection to preserve relative efficiency (214, 215). At the indicated times, cells were stained for CXCR5, and flow cytometric analysis was performed as described below.

Flow cytometric staining

All flow cytometric staining was performed in 96-well plates as previously described (56) with minor modifications.

To evaluate the purity and viability of the initial CD4 T cell sample prior to activation, cells were first washed twice with PBS lacking calcium and magnesium ($PBS^{-/-}$) and then resuspended in 100 μL of $PBS^{-/-}$ containing eBioscience™ Fixable Viability Dye eFluor™ 780 (Thermo Fisher Scientific, cat no. 65-0865-14). Cells were incubated for 30 minutes prior to two washes with standard stain buffer ($PBS^{-/-}$ with 0.09% NaN_3 and 2% human serum) and subsequent resuspension in a cocktail of antibodies specific for markers of myeloid and lymphoid lineages. The antibody cocktail was generated by

first adding appropriate amounts of fluorescent-conjugated monoclonal antibodies (**Table 5.11**) to Brilliant Stain Buffer Plus (BD Biosciences, cat no. 566385), and subsequently adding standard stain buffer to a cumulative 100 µL per test. After resuspension in this antibody cocktail, cells were incubated for 30 minutes prior to wash and resuspension in stain buffer. Cells were then fixed to 1% formaldehyde and transferred to 12x75 mm flow cytometry tubes which were kept on ice until flow cytometric analysis.

At each time point of RNA extraction, a single parallel well of activated CD4 T cells was stained as described above except with an antibody cocktail containing a separate panel of markers, which differed from the initial panel used through the addition of CD25 and CD69 and the omission of CD45, CD19, CD14, and CD16 (**Table 5.11**). Staining and flow cytometric analysis was performed without removing Dynabeads from the sample. This was possible because of the ability to extract the Dynabead fluorescence as its own marker using spectral flow cytometric analysis, described below.

When staining transfected HEK293T, cells were first detached using TrypleE Express (Thermo Fisher Scientific, cat no. 12605010) and washed twice with $PBS^{-/-}$ before resuspension in 100 µL of standard stain buffer with anti-CXCR5 antibody (Biolegend cat no. 356920). Cells were then incubated for 30 minutes prior to subsequent wash and fixation as described above. Fixed cells were then transferred to flow cytometry tubes that were kept on ice until flow cytometric analysis.

Flow cytometric analysis

Flow cytometry was performed with a Cytek® Northern Lights 3-laser flow cytometer, and spectral profiles of each fluorophore were unmixed using SpectroFlo®

software (Cytek Biosciences) and appropriate single-stain and unstained controls to account for autofluorescence. The count per μL feature of the cytometer was also used to calculate cells per staining well and estimate cell expansion during activation. For activation staining, Dynabead fluorescence was accounted for per the advice of Cytek Biosciences technical specialists by independently unmixing the autofluorescence of Dynabeads from activated cells. To do this, the software was instructed to consider Dynabeads as a cell marker and was provided a Dynabead-only control in addition to the other unmixing controls. All processed data files were analyzed in FlowJo™ (BD Biosciences).

For the CD4 T cell experiment, initial purity and viability are reported as the average of three technical replicates calculated using a fraction of isolated cells prior to plating for activation. Quality metrics of activated cells are reported as values for a single parallel well stained at each time point of RNA extraction.

For analyses of HEK293T cells in synchronized transfection experiments, total CXCR5 MFI was calculated across two technical replicates for each independent transfection. These values were then normalized per experiment by dividing by the average 21-hour MFI of Canonical plasmid transfections. Per the figure legend, between 6 and 13 independent transfections per condition were averaged across experiments. Significance is reported based on the adjusted p-value of multiple T-test analyses using the Holm-Sidak method to account for multiple comparisons. Statistical tests were performed using GraphPad Prism (v8.4.3).

RNA extraction

RNA was purified from corresponding cell lysates using RNeasy® Plus Mini Kit (Qiagen, cat no. 74134) following the manufacturer's protocol with the addition of on-column DNAase digestion (Qiagen, cat no. 79254). RNA from each sample was eluted in 30 µL of RNAase-free water. For activated CD4 T cell RNA, samples were aliquoted and kept at -80^{o}C until Iso-Seq library preparation or until shipping to the HudsonAlpha Institute for Biotechnology (Huntsville, AL) for RNA-Seq library preparation. For HEK293T RNA, the quality and amount of RNA were assessed using nanodrop, and 1 µg of eluted RNA was immediately used as input for cDNA synthesis for relative quantification by qPCR (see below).

RNA-Seq library preparation and sequencing

RNA-Seq libraries were prepared by HudsonAlpha Institute for Biotechnology (Huntsville, AL) according to an in-house protocol. Approximately 500 ng of high-quality RNA (all RIN>8 measured by HudsonAlpha directly prior to library construction) was used as input for a PolyA-based mRNA enrichment and library prep (NEB). The resulting cDNA libraries possessed fragment sizes ranging from 616 to 628 base pairs and passed all metrics of quality per company standards, assessed using Kapa qPCR prior to sequencing. Samples were pooled and paired-end 100bp sequencing (50M) was performed for each sample on an Illumina NovaSeq instrument.

Iso-Seq library preparation, sequencing, and initial data analysis

IsoSeq libraries were generated as previously described (217). 500ng of high-quality RNA (RIN$\geq$9.7) was used as input into oligodT primed cDNA synthesis using

commercially available NEB Next reagents (New England Biolabs). A Template Switch Oligo (Pacific Biosciences) was utilized during the reverse transcriptase reaction for second-strand cDNA synthesis. The resulting double-stranded cDNA was amplified with unique barcoded primers and equimolar pooled to include two samples per SMRTbell library preparation (Pool 1: 4 hr and 16 hr; Pool 2: 48 hr and 120 hr). SMRTbell libraries were generated according to manufacturer protocols (Pacific Biosciences). Briefly, the amplified and pooled cDNA underwent enzymatic DNA damage and end repair prior to the ligation of SMRTbell hairpin adapters. An enzymatic cocktail was used to remove the non-ligated template. Final libraries were then purified with magnetic beads prior to annealing to sequencing primer (v4) and binding to polymerase (v2.1). Sequencing was performed using one SMRTcell 8M per pool of barcoded samples on a Sequel IIe system in the UofL Sequencing Technology Center. Following data generation, multiple, iterative sequences covering a single molecule were collapsed to generate highly accurate circular consensus sequence (CCS) reads. These reads were demultiplexed and used as input into the IsoSeq 3 pipeline within SMRTLink (v10.1). CCS reads were further filtered on those reads with quality > 99%, producing "HiFi" reads. The IsoSeq 3 pipeline demultiplexed HiFi reads per individual sample based on barcode from the pooled sequencing data, filtered out amplification artifacts, trimmed primers and polyA tails, and produced *de novo* full-length non-concatemer (FLNC) transcripts for downstream mapping and annotation.

RNA-Seq data analysis

For newly generated RNA-Seq data, the quality of the sequenced reads was first assessed using FastQC v.0.10.1 (200) which indicated high-quality reads such that no sequence trimming was necessary. The sequences were then directly aligned to the *Homo sapiens* reference genome assembly (hg38) using the STAR aligner v2.6 (201) two-pass method with Gencode (v39) annotations (133). The STAR option –outFilterMultimapNmax was set to 1 to allow only uniquely mapped reads. Read counts for gene regions were obtained with HTSeq v.0.10.0 (218) using Gencode annotations.

To analyze previously deposited RNA-Seq data (**Table 5.8**) (144, 219-234), a custom script was written to pull FASTQ files from the NIH Sequence Read Archive (SRA) (235) and directly align sequences to the reference genome assembly (hg38) using STAR aligner. Resulting SJ.out.tab files generated for each sample were used to assess uniquely mapped read coverage across *CXCR5* junctions. Studies selected contained human samples with an average read length $\geq$ 85, and all samples used for analysis were pre-screened within their respective study to include only those without known autoimmune conditions. Corresponding study data and associated publications were used to assign each sample to a cell subset category (**Table 5.9**). After primary analysis, only samples with > 5 reads mapping to the previously annotated internal *CXCR5* junction (Chr11:118883993,118893595) were recognized as expressing *CXCR5* and included in calculations. This resulted in 216 unique donors considered as expressing *CXCR5* with some donors repeated across samples and cell types (**Table 5.9**). The proportion of novel isoform expression was estimated per sample by dividing the coverage of novel end-variant isoform junctions (Chr11:118882935,118883748 and

Chr11:118882935,118883851) by the total coverage at the shared internal *CXCR5* junction (Chr11:118883993,118893595). The average of this calculation per subset category was reported. For all calculations, two or more supporting reads across either novel junction were required to be considered as true expression.

Tertiary Iso-Seq data analyses and isoform characterization

A modified version of a previously described custom pipeline (217) was used to integrate steps in the cDNA Cupcake protocol for post-processing of Iso-Seq v3 FLNC reads (https://github.com/Magdoll/cDNA_Cupcake). The alignment files for the FLNC reads were converted to fastq (198) and fasta format. Fasta files were aligned to the *Homo sapiens* reference genome assembly (hg38) using minimap2 v2.18-r1015 (199) with the option -secondary=no. The resulting alignment files were collapsed into isoforms based on sequence similarity using the cDNA Cupcake Python script collapse_isoforms_by_sam.py. Isoform abundance was calculated using get_abundance_post_collapse.py, and isoforms were filtered to include those with at least two supporting reads with filter_by_count.py. Isoforms were also filtered to remove possible non-full length reads by removing those with indications of a degraded 5' prime region, i.e., apparent 5' shortened isoforms with otherwise equivalent long reads (filter_away_subset.py). For chained analyses, the samples were chained together using chain_samples.py. SQANTI3 v4.0 (192) was used to summarize the results for each of the individual samples and the chained dataset. SQANTI3 input included the cDNA Cupcake GTF, Gencode (v39) annotations (133), the hg38 assembly, long read transcript abundance, a polyA motif list, and a CAGE peak (TSS sites) BED file for human and

mouse. UCSC Genome Browser tracks were created using guidelines and utilities available on the UCSC Genome Browser website (202). Custom tracks for the individual sample Iso-Seq and matched RNA-Seq data are available to view at http://genome.ucsc.edu/cgi-bin/hgHubConnect?hgHub_do_redirect=on&hgHubConnect.remakeTrackHub=on&hgHub_do_firstDb=on&position=chr1:206,903,317-206,921,941&hubUrl=http://162.215.210.70/~tracks/Mitchell_IsoSeq_Stim/hub.txt

Calculating and comparing 5' UTR length for sample isoforms

A custom script was written to compare the 5' UTR length of each isoform in a sample to the longest known 5' UTR of its SQANTI 3 assigned parent gene. The SQANTI3 annotated coding sequence (CDS) start was used as an equivalent to isoform 5' UTR length. The longest 5' UTR length among annotated transcripts for each gene was calculated from transcripts within Gencode (v39, comprehensive) or Matched annotations between NCBI and EMBL-EBI (MANE Select) (209). Only predicted protein-coding isoforms that mapped to a known gene were compared in this analysis. Publicly available R packages ggplot2 and dplyr were used to visualize distributions of UTR differences across transcripts.

Where indicated, novel 5' UTR discovered for immune-important genes were aligned to each other and to previously annotated 5' UTR of their parent gene using CLUSTAL Multiple Sequence Alignment by Muscle 3.8 (236-238). Additionally, upstream elements within novel 5' UTR were identified using independent observation, UTRscan (97) or the NCBI ORF finder (239, 240).

Two-step RT-qPCR and relative quantification analysis

Immediately following RNA extraction, 1 μg of RNA was used as input to cDNA synthesis. cDNA synthesis was performed using QuantiTect Reverse Transcription kit (Qiagen, cat no. 205313) following the manufacturer's protocol. Samples with low-quality RNA were excluded based on a low A260/A280 ratio, leading to exclusion of the 12 hr timepoint for all preps in one experiment and a single experiment 24 hr time point for Isoform 1. cDNA was aliquoted and stored at -20^{o}C until qPCR.

qPCR was performed using Quantinova SYBR® Green PCR kit (Qiagen, cat no. 208054) on a Quantstudio 3 (Thermo Fisher Scientific) instrument following the manufacturer's protocol. All independent experimental replicates were analyzed with four technical replicates across two qPCR runs. All primer pairs (Bio-Rad PrimePCR Assay^{TM}, cat no. 10025636, **Table 5.12**) were certified as wet lab validated and optimized for qPCR. An un-transfected HEK293T cDNA control was used to assess off-target amplification of the *CXCR5* primer pair, and none was found. No template controls and controls that had no reverse transcriptase added during cDNA synthesis were also run in parallel to further screen for the quality of each experiment.

The qPCR data were analyzed using the QuantStudio 3 Design and Analysis software (Thermo Fisher Scientific) with the Relative Quantification app (v4.3). Only wells with a raw Ct between 15 and 35 were kept, and those with low-quality amplification curves were excluded. Endogenous controls *SDHA, ACTB,* and *B2M* were used for initial normalization. The average gene stability score across experiments for these controls was 0.002, with no single score greater than 0.008, suggesting highly stable expression. ΔCt values were calculated based on these controls, and relative

quantification of *CXCR5* and *FOXO3* was performed using ΔΔCt normalized to 1 hr post-ActD addition. Relative quantification values of *CXCR5* or *FOXO3* at 1, 2, 4, 8, 12, and 24 hrs post-ActD addition were graphed using GraphPad Prism. For each condition, one phase decay nonlinear regression was performed with the constraints of plateau = 0 and shared Y_o. Curves generated were used to estimate the 95% confidence interval (CI) of RNA half-life ($t_{1/2}$). The 1 hr time point was consistently the peak of CXCR5 expression, so 0 hr data were excluded from these analyses and instead inferred by the shared Y_o restriction. Analyses were performed for three independent experiments of the Canonical Isoform or four independent experiments for each of Novel Isoform 1 or 2. One experimental outlier at the 4-hour time point for Isoform 1 was detected by GraphPad Prism and thus excluded.

Data Availability Statement

All Sequencing data sets were deposited to the Gene Expression Omnibus (GEO). For each time point of CD4 T cell activation, FLNC files were deposited in BAM format alongside further processed data files (Iso-Seq only: GSE229971, SuperSeries: GSE229972). The processed files consist of raw gene isoform counts, annotation (GFF) files, and UCSC Genome Browser tracks. Raw Illumina RNA-Seq data for each dataset were deposited as fastq.gz files (RNA-Seq only: GSE229969, SuperSeries: GSE229972). Processed RNA-Seq files deposited consist of raw gene counts and UCSC Genome Browser tracks.

CHAPTER V TABLES

Table 5.1. Total RNA-Seq Reads and Uniquely Mapped Alignment.

Sample (Activation Time Point)	*Total Reads*	*% of reads ≥ Q30*	*Mean Read Quality*	*Uniquely Aligned Reads*	*Unique Alignment Rate*
4 hr	58,284,865	91.73%	> Q35 (>99.97%)	54,674,774	93.81%
16 hr	69,361,924	92.02%		64,992,643	93.70%
48 hr	68,319,565	91.52%		63,499,850	92.95%
120 hr	70,477,487	91.80%		65,908,266	93.52%

Table 5.2. Number of FLNCs, clusters, and non-zero transcripts for Iso-Seq. HiFi reads are filtered on quality > 99%. FLNC (full-length, non-concatemer) reads are generated from HiFi reads.

Sample (Activation Time Point)	*SMRTcell Pool*	*Quality Loading Efficiency (P1)*	*Read Length N50 (bp)*	*Number of HiFi Reads*	*Mean HiFi Read Quality*	*Number of FLNC*
4 hr	**1**	75.3%	147,250	3,666,219	Q41 (99.99%)	1,714,773
16 hr						1,252,437
48 hr	**2**	71.0%	146,750	3,327,662	Q40 (99.99%)	1,586,335
120 hr						1,351,430

Table 5.3. Genes and isoforms detected by Iso-Seq.

Sample (Activation Time Point)	*Unique Genes*	*Unique Isoforms*	*Annotated Genes*	*Novel Genes*
4 hr	10,999	43,269	9,603	1,396
16 hr	10,698	39,690	9,576	1,122
48 hr	14,615	52,538	11,304	3,311
120 hr	11,463	45,246	10,862	601
Chained	18,277	118,287	13,307	4,970

Table 5.4. Transcripts identified in each structural category. FSM, full splice match; ISM, incomplete splice match; NIC, novel in catalog; NNC, novel, not in catalog. Entries show the percentage each category represents amongst total unique transcripts, with the number of transcripts in each category in parentheses.

Sample (Activation Time Point)	*FSM*	*ISM*	*NIC*	*NNC*	*Genic Genomic*	*Intergenic*	*Antisense*	*Fusion*	*Genic Intron*
4 hr	**38%** (16,367)	**10%** (4,241)	**25%** (10,686)	**22%** (9,411)	**2%** (722)	**2%** (1,042)	**1%** (425)	**1%** (375)	0
16 hr	**40%** (16,036)	**10%** (3,821)	**28%** (11,128)	**17%** (6,831)	**1%** (511)	**2%** (872)	**1%** (273)	**1%** (218)	0
48 hr	**36%** (18,872)	**12%** (6,072)	**23%** (12,263)	**20%** (10,493)	**2%** (887)	**5%** (2,486)	**2%** (969)	**1%** (496)	0
120 hr	**40%** (17,967)	**9%** (4,028)	**27%** (12,319)	**20%** (9,206)	**1%** (551)	**1%** (401)	**1%** (227)	**1%** (547)	0
Chained	**31%** (36,831)	**11%** (12,483)	**27%** (31,937)	**24%** (28,297)	**2%** (1,982)	**3%** (3,812)	**1%** (1,486)	**1%** (1,459)	0

Table 5.5. FSM transcripts identified as end-variant isoforms by SQANTI 3. The number of specific end-variant transcripts within each FSM subcategory is reported alongside the proportion of those transcripts whose start or termination sites map within annotated CAGE peaks or to known PolyA motifs, respectively.

Sample (Activation Time Point)	**FSM Subcategory (% CAGE, % PolyA)**			
	Reference Match	***Alternative 3' end***	***Alternative 5' end***	***Alternative 3' and 5' ends***
4 hr	4,718 (77%, 94%)	5,911 (81%, 78%)	2,372 (43%, 94%)	2,764 (50%, 79%)
16 hr	4,780 (73%, 94%)	5,758 (76%, 77%)	2,285 (42%, 94%)	2,683 (46%, 79%)
48 hr	5,300 (70%, 94%)	6,407 (76%, 73%)	3,110 (40%, 93%)	3,348 (47%, 74%)
120 hr	5,453 (74%, 94%)	6,191 (79%, 79%)	2,796 (46%, 94%)	2,927 (52%, 80%)
Chained	7,873 (68%, 93%)	13,895 (75%, 74%)	5,879 (35%, 93%)	7,879 (44%, 76%)

Table 5.6. Transcripts with lengthened 5' UTR. Only predicted protein-coding transcripts that were within FSM, ISM, NIC, or NNC categories and mapped to a known gene were considered for UTR comparisons.

Sample (Activation Time Point)	**Predicted protein-coding and map to known gene**	***5' UTR ≥ 10 bp longer than***		***5' UTR ≥ 100 bp longer than***	
		MANE 5' UTR	**Longest 5' UTR in Gencode^**	**MANE 5' UTR**	**Longest 5' UTR in Gencode^**
4 hr	37,190	21,516	6,625	10,691	3,869
16 hr	35,112	21,037	6,166	9,606	3,469
48 hr	43,970	27,102	8,070	12,277	4,324
120 hr	40,757	24,853	7,534	11,569	4,151
Chained	100,986	63,176	20,752	32,664	12,635

^Gencode v39 Comprehensive dataset was used for comparison

Table 5.7. Breakdown of transcripts with long 5' UTR within the chained sample dataset. SQANTI3 categories (FSM, ISM, NIC, NNC) assigned to the transcripts discovered to possess long 5' UTR. Only predicted protein-coding transcripts that mapped to a known gene were considered for UTR comparisons. Proportions among transcripts compared within each category are shown in parentheses.

Chained Sample Statistic		*Within Structural Category*				
		FSM	**ISM**	**NIC**	**NNC**	**Intron Retention***
Predicted protein-coding and mapped to known gene		33,954	11,449	29,402	26,181	15,305
5' UTR ≥ 10 bp longer than:	**MANE 5' UTR**	23,199 (68%)	6,439 (56%)	18,473 (62%)	15,065 (58%)	10,265 (67%)
	Longest 5' UTR in Gencode^	6,551 (19%)	2,292 (20%)	6,283 (21%)	5,625 (21%)	4,783 (31%)
5' UTR ≥ 100 bp longer than:	**MANE 5' UTR**	9,127 (27%)	2,820 (25%)	11,550 (39%)	9,166 (35%)	7,409 (48%)
	Longest 5' UTR in Gencode^	2,760 (8%)	1,042 (9%)	4,596 (15%)	3,875 (15%)	4,028 (26%)

^Gencode v39 Comprehensive dataset was used for comparison

** Intron Retention subcategory events were tabulated from within ISM, NIC, and NNC categories*

Table 5.8. BioSamples used to assess novel *CXCR5* isoform junctions. FASTQ files from the NIH Sequence Read Archive used to validate end-variants detected by Iso-Seq. Some samples were later found to have no short-read sequences that mapped to the canonical splice junction of CXCR5 Such that they were judged to have no detectable CXCR5 expression (summarized in Table 5.9).

BioProject Accession Number (PRJNA#)	Assigned cell subset category[1]: BioSample Accession Number (SRR#)
30709 **ENCODE Consortium:** Yijun Ruan, JAX Barbara Wold, Caltech John Stamatoyannopoulos, UW Bradley Bernstein, Broad	**Activated B cell**: 22521771, 22521772 **Activated CD8 cell:** 16808559, 16808560, 16811460, 16811461, 16811937, 16811938, 16815211, 16815212, 16809015, 16809016, 16809791, 16809792, 16811789, 16811790, 16812600, 16812601, 16811164, 16811165, 16812437, 16812438, 16813254, 16813255, 16815213, 16815214 **Activated Th17:** 22521503, 22522091 **Activated Th2:** 22521643, 22521644 **B Cell**: 5048157, 5048158, 5048159, 5048160, 5048161, 5048162, 5048163, 5048164, 5048165, 5048166, 5048167, 5048168, 16809078, 16812085, 16809493, 16811118, 16815804, 22521972, 22521973 **Blasting CD4:** 16810817, 16810818, 16815643, 16815644 **Blasting from Memory**: 16814387, 16814388, 16816454, 16816455 **Blasting from Naïve**: 16814815, 16814816, 16815923, 16815924 **Early Proliferating from Naïve**: 22521312, 22521436 **Late Proliferating CD4**: 16811841, 16811842, 16814981, 16814982, 22520975, 22520976 **Late Proliferating from Memory**: 16811048, 16811049, 16813298, 16813299 **Late Proliferating from Naïve**: 16808326, 16808327, 16814027, 16814028 **Memory CD4**: 16809042, 16810750, 16811702, 16808891, 16808892, 16810759, 16810760 **Naïve CD4**: 16810660, 16812052, 16813424, 16815452, 16815453, 16816234, 16816235, 22520801, 22520802, 22521433, 22521556, 22521557 **Resting CD4:** 16808496, 16808497, 16811207, 16811208 **Resting CD8**: 16807998, 16808870, 16809699, 16808195, 16808325, 16808561, 16814544, 16814545, 16816236, 16816237, 16813430, 16813431, 16814783, 16814784, 16808323, 16808324, 16811451, 16811452 **Th17**: 22521821, 22522178 **Th2**: 22521125, 22521126 **Treg**: 16807314, 16815055, 16815210, 22521050
252962	**GC Tfh**: 1747966, 1747967, 1747968 **Naïve CD4**: 1422909, 1422906, 1422907, 1422908
264229	**Activated Th1**: 1615173, 1615182, 1615177, 1615186 **Activated Th2**: 1615175, 1615179, 1615184, 1615188 **Naïve CD4**: 1615171, 1615180 **Th1**: 1615172, 1615176, 1615181, 1615185 **Th2**: 1615174, 1615178, 1615183, 1615187
369563	**Blasting from Naïve**: 5223499, 5223500, 5223501, 5223521, 5223522, 5223542, 5223543, 5223544, 5223568 **Late Proliferating from Naïve**: 5223502, 5223523, 5223545 **Naïve CD4**: 5223498, 5223519, 5223540 **Treg**: 5223518, 5223561, 5223573
464049	**Early Proliferating from Naïve**: 7123400, 7123402, 7123404, 7123406, 7123408, 7123410, 7123412, 7123414, 7123416, 7123418, 7123420, 7123422, 7123424, 7123426, 7123428, 7123430, 7123432, 7123434, 7123436, 7123438, 7123440, 7123442, 7123444, 7123446, 7123448, 7123450, 7123452, 7123454, 7123456, 7123458, 7123460, 7123462, 7123464, 7123466, 7123468, 7123470, 7123472, 7123474, 7123476, 7123478, 7123480, 7123482, 7123484, 7123486, 7123487, 7123489, 7123491, 7123493, 7123495, 7123497, 7123499, 7123501, 7123503, 7123505, 7123506, 7123507, 7123509, 7123511, 7123513, 7123515, 7123517, 7123519, 7123521, 7123523, 7123525, 7123527, 7123529, 7123531, 7123533, 7123535 **Naïve CD4:** 7123401, 7123403, 7123405, 7123407, 7123409, 7123411, 7123413, 7123415, 7123417, 7123419, 7123421, 7123423, 7123425, 7123427, 7123429, 7123431, 7123433, 7123435, 7123437, 7123439, 7123441, 7123443, 7123445, 7123447, 7123449, 7123451, 7123453, 7123455, 7123457, 7123459, 7123461, 7123463, 7123465, 7123467, 7123469, 7123471, 7123473, 7123475, 7123477, 7123479, 7123481, 7123483, 7123485, 7123488, 7123490, 7123492, 7123494, 7123496, 7123498, 7123500, 7123502, 7123504, 7123508, 7123510, 7123512, 7123514, 7123516, 7123518, 7123520, 7123522, 7123524, 7123526, 7123528, 7123530, 7123532, 7123534, 7123536
484735	**Activated B cell:** 7647655, 7647697, 7647732, 7647657, 7647699, 7647733, 7647659, 7647734, 7647802, 7647817 **Activated CD8 cell:** 11007097, 11007128, 11007062, 11007161, 11007064, 11007130, 11007098, 11007163, 11007100, 11007165, 11007066, 11007132, 11007068, 11007103, 11007134, 11007167 **Activated cTfh**: 11007140, 11007173, 11007074, 11007107 **Activated Th1**: 11007085, 11007116, 11007151, 11007183

	Activated Th17: 11007184, 11007087, 11007118, 11007153 **Activated Th2**: 11007089, 11007120, 11007155, 11007185 **Activated Treg**: 11007181, 7647804, 7647810, 11007114, 11007148, 11007083 **B cell:** 7647654, 7647696, 7647731, 7647767, 7647656, 7647698, 7647768, 7647807, 7647658, 7647700, 7647769, 7647808 **Blasting from Memory**: 11007142, 11007076, 11007109, 11007175 **Blasting from Naïve**: 11007080, 11007112, 11007179, 11007146, 7647803, 7647818 **cTfh**: 11007106, 11007139, 11007073, 11007172 **Memory CD4**: 11007075, 11007108, 11007174, 11007141 **Naïve CD4**: 11007145, 11007079, 11007111, 11007178 **Resting CD8**: 11007061, 11007096, 11007160, 11007127, 11007162, 11007063, 11007129, 11007065, 11007099, 11007164, 11007131, 11007067, 11007102, 11007133, 11007166 **Th1**: 11007084, 11007115, 11007149, 7647811 **Th17**: 11007152, 11007086, 11007117, 7647812 **Th2**: 11007088, 11007154, 11007119, 7647813 **Treg**: 11007081, 11007180, 11007113, 11007147, 11007082, 11007182
516647	**Activated Treg**: 8477715, 8477716, 8477717, 8477718, 8477727, 8477728, 8477729, 8477730, 8477739, 8477740, 8477741, 8477742
521046	**Activated CD8:** 8534318, 8534324 **B cell:** 8534319, 8534325 **Early Proliferating from Naïve**: 8534317, 8534323 **Naïve CD4**: 8534320, 8534321, 8534326 **Resting CD8**: 8534322, 8534327
541437	**cTfh:** 9021737, 9021740, 9021745 **GC Tfh:** 9021723, 9021727, 9021731 **Naïve CD4**: 9021725, 9021729, 9021733, 9021739, 9021742, 9021747 **Non-GC Tfh**: 9021724, 9021728, 9021732
542640	**cTfh:** 9047607, 9047610, 9047613, 9047616, 9047620, 9047623, 9047626, 9047629, 9047631, 9047634, 9047637, 9047640, 9047643, 9047646, 9047649
555109	**cTfh:** 9698907, 9698909, 9698913, 9698915, 9698919, 9698921, 9698925, 9698927, 9698931, 9698933, 9698937, 9698939, 9698943, 9698945, 9698949, 9698951, 9698955, 9698957, 9698961, 9698963, 9698967, 9698969, 9698973, 9698975, 9698979, 9698981, 9698985, 9698987 **Naïve CD4:** 9698911, 9698917, 9698923, 9698929, 9698935, 9698941, 9698947, 9698953, 9698959, 9698965, 9698971, 9698977, 9698983, 9698989
557467	**Blasting from Naïve:** 9861798, 9861799, 9861800, 9861801, 9861802, 9861803 **Naïve CD4:** 9861777, 9861778, 9861779
562144	**GC Tfh:** 10033933, 10033934, 10033937, 10033938, 10033939, 10033940, 10033941, 10033942, 10033943, 10033944, 10033945, 10033946, 10033947, 10033948, 10033949, 10033950, 10033953, 10033954, 10033957, 10033958, 10033959, 10033960, 10033961, 10033962, 10033963, 10033964, 10033967, 10033968 **Naïve CD4:** 10033931, 10033935, 10033951, 10033955, 10033965 **Non-GC Tfh**: 10033932, 10033936, 10033952, 10033956, 10033966
627949	**Early Proliferating CD4:** 11607045, 11607047, 11607049, 11607051, 11607053, 11607055, 11607057, 11607059, 11607061, 11607063, 11607065, 11607067 **Transitional Tfh:** 11607046, 11607048, 11607050, 11607052, 11607054, 11607056, 11607058, 11607060, 11607062, 11607064, 11607066, 11607068
682651	**cTfh**: 14369428, 14369430, 14369433, 14369436, 14369439, 14369442, 14369445
704268	**cTfh:** 13768446, 13768447, 13768448, 13768449 **Memory CD4**: 13768442, 13768443, 13768444, 13768445 **Naïve CD4**: 13768440, 13768441
743946	**Activated B cell:** 15046299, 15046302, 15046310, 15046311, 15046318, 15046321, 15046322, 15046327, 15046332, 15046335, 15046339, 15046341, 15046347 **B cell:** 15046306, 15046307, 15046309, 15046316, 15046317, 15046331, 15046334, 15046338, 15046346, 15046350, 15046304, 15046305, 15046314, 15046315, 15046320, 15046324, 15046325, 15046329, 15046330, 15046333, 15046337, 15046340, 15046344, 15046345, 15046349 **cTfh:** 15046300, 15046301, 15046303, 15046308, 15046312, 15046313, 15046319, 15046323, 15046326, 15046328, 15046336, 15046342, 15046343, 15046348
744261	**cTfh:** 15055402, 15055406, 15055575, 15055579, 15055583, 15055590 **Naïve CD4**: 15055400, 15055404, 15055470, 15055473, 15055475, 15055478, 15055481, 15055484, 15055487, 15055490, 15055567, 15055570, 15055573, 15055577, 15055581, 15055585, 15055588 **Resting CD8**: 15055408, 15055414, 15055420, 15055431, 15055439, 15055449, 15055456, 15055463, 15055495, 15055509, 15055517, 15055527, 15055536, 15055543, 15055551, 15055558, 15055594, 15055600, 15055607, 15055426, 15055409, 15055415, 15055421, 15055427, 15055440, 15055450, 15055457, 15055464, 15055501, 15055510, 15055518, 15055528, 15055537, 15055544, 15055552, 15055559, 15055592, 15055595, 15055601, 15055608, 15055432, 15055412, 15055419, 15055425, 15055436, 15055444, 15055454, 15055461, 15055468, 15055496, 15055505, 15055514, 15055522, 15055532, 15055541, 15055548, 15055556, 15055562, 15055605, 15055411, 15055418, 15055424, 15055429, 15055435, 15055443, 15055453, 15055460, 15055467, 15055504, 15055513, 15055521, 15055531, 15055540, 15055547, 15055555, 15055561, 15055604, 15055611, 15055417, 15055423, 15055434, 15055442, 15055452, 15055459, 15055466, 15055503, 15055512, 15055520, 15055530, 15055539, 15055546, 15055554, 15055603, 15055610, 15055410, 15055416, 15055422, 15055428, 15055433, 15055441, 15055451, 15055458, 15055465, 15055502, 15055511, 15055519, 15055529, 15055538, 15055545, 15055553, 15055560, 15055602, 15055609, 15055413, 15055430, 15055437, 15055445, 15055462, 15055469, 15055497, 15055506, 15055515, 15055523, 15055533, 15055542, 15055549, 15055557, 15055563, 15055593, 15055606

	Treg: 15055403, 15055407, 15055472, 15055477, 15055480, 15055483, 15055486, 15055489, 15055492, 15055494, 15055569, 15055572, 15055576, 15055580, 15055584, 15055587
781654	**Early Proliferating from Naïve**: 16976861, 16976862, 16976863, 16976864, 16976865, 16976866, 16976867, 16976868, 16976869, 16976870, 16976871, 16976872, 16976873, 16976874, 16976875, 16976876, 16976877, 16976878, 16976879, 16976880, 16976881, 16976882, 16976883, 16976884, 16976885, 16976886, 16976887, 16976888, 16976889, 16976890, 16976891, 16976892, 16976893, 16976894, 16976895, 16976896, 16976897, 16976898, 16976899, 16976900, 16976901, 16976902, 16976903, 16976904, 16976905, 16976906, 16976907, 16976908

[1]Subsets were assigned according to Table 5.4.

Table 5.9. Description of cell subset assignments and donor numbers for analysis of previous RNA-Seq data. The time point of activation was only considered where indicated and activated T cell categories do not distinguish samples activated in the presence or absence of supplemented IL-2. Independent donor # is not reflective of the total number of samples investigated as some samples were found to be repeat sequencing of the same donor. More than 5 reads spanning the shared internal (canonical) *CXCR5* junction were required to be considered as expressing *CXCR5*.

Cell Subset Category	***Description of Parameters***	***# Donors Analyzed***	***# Donors with Detectable CXCR5 Expression****
Resting CD4	Whole CD4 T cells; unactivated.	***1***	***0***
Blasting CD4	Whole CD4 T cells activated for times ranging from 4-48 hours. (Early Activation)	***1***	***1***
Early Proliferating CD4	Whole CD4 T cells activated for 72-120 hours. (Mid Activation).	***12***	***12***
Late Proliferating CD4	Whole CD4 T cells activated for > 120 hours. (Late Activation)	***2***	***1***
Naïve CD4	Isolated naïve CD4 T cells; resting condition, unactivated.	***113***	***29***
Blasting from Naïve	Isolated naïve CD4 T cells activated for times ranging from 4-48 hours. (Early Activation)	***13***	***5***
Early Proliferating from Naïve	Isolated naïve CD4 T cells activated for 72-120 hours. (Mid Activation)	***96***	***94***
Late Proliferating from Naïve	Isolated naïve CD4 T cells activated for > 120 hours. (Late Activation)	***4***	***0***
Memory CD4	Isolated memory CD4 T cells (both Tcm and Tem); unactivated.	***10***	***8***
Blasting from Memory	Isolated memory CD4 T cells reactivated for 24-48 hours. (Early re-activation)	***5***	***4***
Late Proliferating from Memory	Isolated memory CD4 T cells reactivated for > 120 hours. (Late re-activation)	***1***	***0***
Th1	Th1 polarized or isolated Th1 cells.	***6***	***1***
Activated Th1	Isolated Th1 or pre-polarized Th1 that were re-activated immediately prior to sequencing.	***6***	***4***
Th2	Th2 polarized or isolated Th2 cells.	***7***	***0***
Activated Th2	Isolated Th2 or pre-polarized Th2 that were re-activated immediately prior to sequencing.	***7***	***5***
Th17	Th17 polarized or isolated Th17 cells.	***6***	***2***
Activated Th17	Isolated Th17 or pre-polarized Th17 that were re-activated immediately prior to sequencing.	***6***	***4***
Treg	Isolated circulating T regulatory cells (nTreg).	***27***	***10***
Activated Treg	Isolated T regulatory cells activated *ex vivo.*	***9***	***3***
cTfh	Circulating T follicular helper cells: CXCR5+ CD4 T cells from human peripheral blood.	***63***	***61***
Activated cTfh	Isolated cTfh re-activated directly prior to sequencing.	***4***	***3***
Transitional Tfh	Isolated CD4 cells stimulated *ex vivo* with polarizing cytokines to create a transitional state of T follicular helper differentiation; not yet stable Tfh.	***12***	***12***

GC Tfh	Germinal center T follicular helper cells: CXCR5hiPD1hi CD4 T cells isolated from human spleen, lymph node, or tonsil.	***21***	***20***
Non-GC Tfh	Non-germinal center T follicular helper cells: CXCR5lo CD4 T cells isolated from human spleen, lymph node, or tonsil.	***8***	***7***
Resting CD8	Isolated CD8 T cells (resting condition).	***32***	***7***
Activated CD8 cell	Isolated CD8 T cells activated directly prior to sequencing.	***8***	***4***
B cell	Isolated B cell (any subtype).	***24***	***22***
Activated B cell	Isolated B cell activated directly prior to sequencing.	***17***	***14***

**** > 5 reads spanning the canonical CXCR5 splice junction***

Table 5.10. cDNA sequences of 5' UTR of plasmid inserts. All 5' UTR sequences were cloned upstream of the same *CXCR5* coding sequence and 3' UTR (CXCR5 Transcript Variant 1, NM_001716.2, Origene cat no. SC309454). For Isoform 2 mutants, specific modifications are bolded and red. Sequences in brackets were removed. For A-rich removal, the maximum amount of sequence was removed while still maintaining the frame of other UTR elements to the *CXCR5* coding sequence. The location of the *CXCR5* start codon is capitalized.

Isoform (plasmid)	5' UTR cDNA Sequence used for isoform inserts
Canonical	cctctcaacataagacagtgaccagtctggtgactcacagccggcacagccATG
Isoform 1	agacaggacagagttgagggaaaggacagaggttatgagtgcctgcaagagtggcagcctggag tagagaaaacactaaaggtggagtcaaaagacctgagttcaagtcccagctctgccactggttagct gtgggatctcggctgacggctgccacctctctagaggcacctggcggggagcctctcaacataag acagtgaccagtctggtgactcacagccggcacagccATG
Isoform 2	agacaggacagagttgagggaaaggacagaggttatgagtgcctgcaagagtggcagcctggag tagagaaaacactaaaggtggagtcaaaagacctgagttcaagtcccagctctgccactggttagct gtgggatctcggaaaagacccagtggaaaaaaaaaaaaaaagtgatgagttgtgaggcaggtcgc ggccctactgcctcaggagacgatgcgcagctcatttgcttaaatttgcagctgacggctgccacct ctctagaggcacctggcggggagcctctcaacataagacagtgaccagtctggtgactcacagcc ggcacagccATG
Isoform 2 w/o uORFs	agacaggacagagttgagggaaaggacagaggtt**gta**agtgcctgcaagagtggcagcctgga gtagagaaaacactaaaggtggagtcaaaagacctgagttcaagtcccagctctgccactggttag ctgtgggatctcggaaaagacccagtggaaaaaaaaaaaaaaagtg**gta**agttgtgaggcaggtcg cggccctactgcctcaggagacg**gta**cgcagctcatttgcttaaatttgcagctgacggctgccacc tctctagaggcacctggcggggagcctctcaacataagacagtgaccagtctggtgactcacagcc ggcacagccATG
Isoform 2 w/o A-rich	agacaggacagagttgagggaaaggacagaggttatgagtgcctgcaagagtggcagcctggag tagagaaaacactaaaggtggagtcaaaagacctgagttcaagtcccagctctgccactggttagct gtgggatctcggaaaagacccagtggaa**[aaaaaaaaaaaa]**gtgatgagttgtgaggcaggtc gcggccctactgcctcaggagacgatgcgcagctcatttgcttaaatttgcagctgacggctgccac ctctctagaggcacctggcggggagcctctcaacataagacagtgaccagtctggtgactcacagc cggcacagccATG

Table 5.11. Antibody markers used for flow cytometric analysis. To assess purity, a fraction of the initial CD4 T cell sample was stained with antibodies specific to the myeloid and lymphocyte lineage markers listed. Cell types and subsets were defined within the sample by flow cytometric analysis as follows: +, the presence of the lineage marker was required; -, cells with the lineage marker were excluded from the group. No symbol entry means the marker was not considered in defining the cell type. To assess the success of activation, staining was performed at indicated time points on parallel activated wells. *, markers were excluded from activation staining analysis. ^, markers were introduced as indicators for post-activation staining only and not used for cell type or subset identification.

Supplier, Catalog #	Marker Specificity	Cell Type Defined in Purity and Subset Staining											
		$CD4^+$ T	*CD4+ Naïve*	*CD4+ Tcm*	*CD4+ Tem*	*CD4+ Teff*	*$CD4^+$ NK T*	*$CD8^+$ T*	*$CD4^-$ $CD8^-$ T*	*NK*	*B*	*$CD14^+$ $CD16^-$ Monocyte*	*$CD14^+$ $CD16^+$ Monocyte*
BD, 560367	**CD45***	+	+	+	+	+	+	+	+	+	+	+	+
ThermoFisher, 58-0038-42	**CD3**	+	+	+	+	+	+	+	+	-	-		
BioLegend, 300534	**CD4**	+	+	+	+	+	+	-	-				
ThermoFisher, 64-0088-42	**CD8**	-	-	-	-	-	-	+	-				
BD, 564057	**CD56**						+			+		-	
ThermoFisher, 69-0199-42	**CD19***	-	-	-	-	-	-	-	-	-	+	-	
Tonbo, 20-0149-T100	**CD14***										-	+	+
BD, 555408	**CD16***										-	-	+
Invitrogen, 25-0457-42	**CD45RO**		-	+	+	-							
Biolegend, 304820	**CD62L**		+	+	-	-							
BD, 1056723	**CD25^**												
BD, 0288235	**CD69^**												

*, included only for purity staining, not included in activation staining panel
^, included only in activation staining panel, not used for cell type identification

Table 5.12. Primer pairs used in qPCR. All primers were from PrimePCR™ Assay (Bio-Rad cat no. 10025636) designed for SYBR® Green experiments and wet lab validated for qPCR.

Target	Bio-Rad unique assay ID	Chromosome Location	Amplicon Length
CXCR5	qHsaCID0020761	11:118754662-118764394	100
FOXO3	qHsaCID0023235	6:108986007-109001107	134
B2M	qHsaCID0015347	15:45003754-45007715	123
SDHA	qHsaCED0057393	5:256526-256746	191
ACTB	qHsaCED0036269	7:5568936-5569027	62

CHAPTER V FIGURES

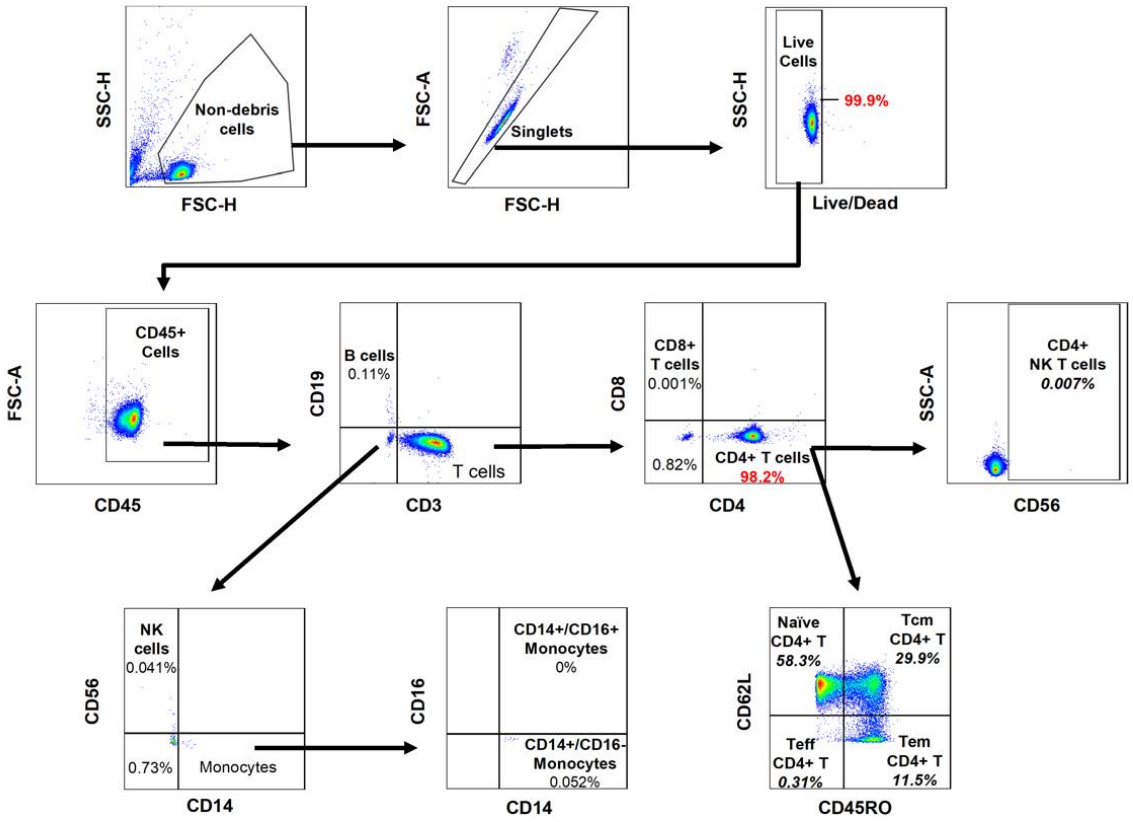


Figure 5.1. Live cells and major subsets present within the purified CD4 T cell sample. A fraction of the initial purified sample was stained with makers specific for myeloid and lymphoid lineages and the gating strategy depicted was used to define live cells, $CD4^+$ T cells, contaminating cells, and CD4 T cell subsets. Inset percentages show the final calculation averaged across three technical replicates. Dot plots are representative of the gating of one technical replicate. In calculating viability, the live cell fraction of total singlets was used. In calculating CD4 T cell subset percentages (Naïve, Tcm, Tem, Teff, and NK T), the fraction of CD4+ T cells was used. All other percentages are reported as the fraction of total $CD45^+$ cells.

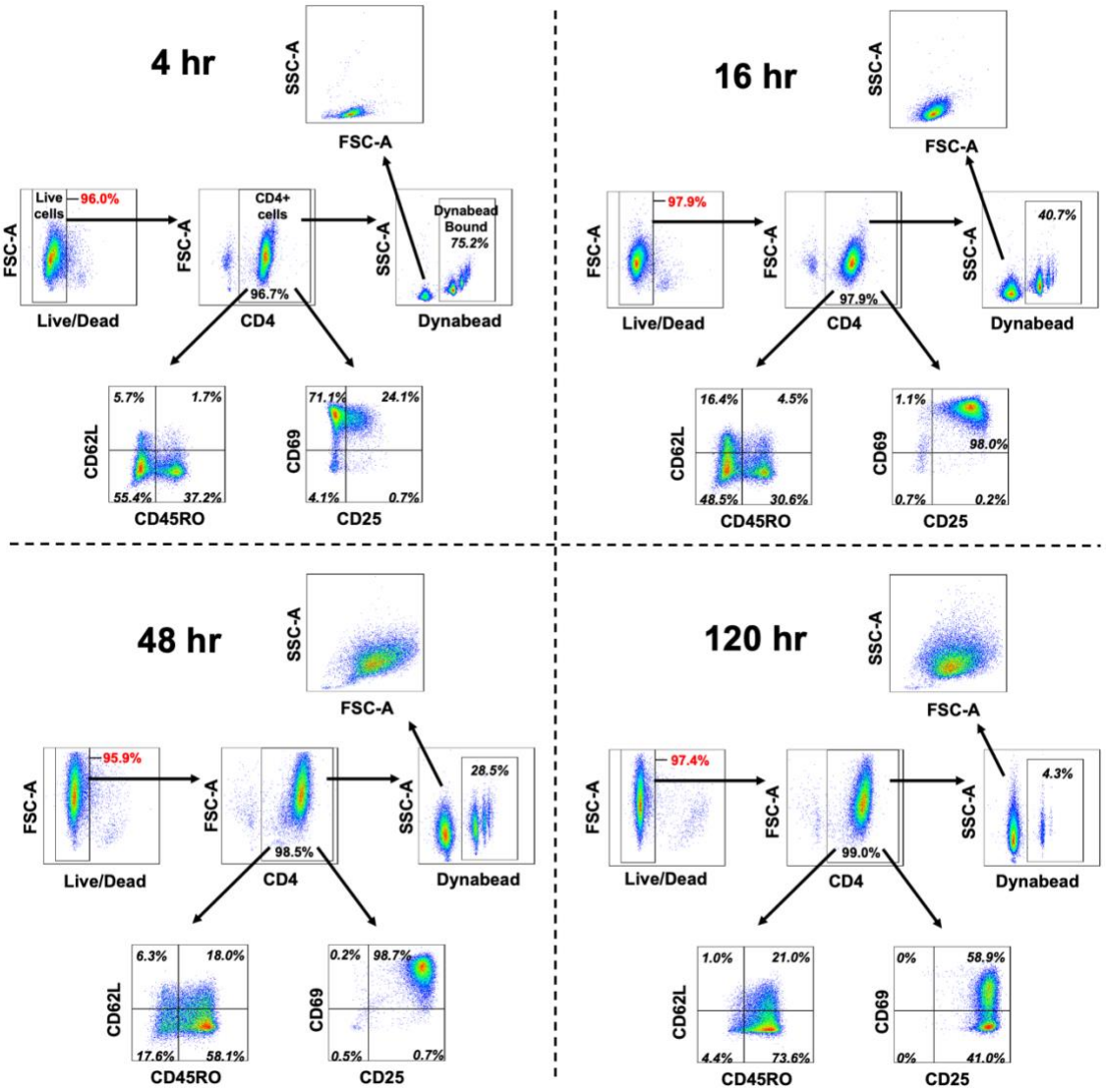


Figure 5.2. Cell viability and activation markers of CD4 T cells stained in parallel to RNA extraction. At each time point of RNA extraction, a single parallel well of activated CD4 T cells was stained with makers relevant to T cell activation. The gating strategy used was the same as that upstream of Live/Dead differentiation depicted in Figure 5.2. In calculating viability, the live cell fraction of total singlets was used. All other inset percentages are reported as a fraction of the respective parent population. Forward-by-side scatter plots are shown only for those cells not bound by Dynabeads to allow adequate assessment of blasting phenotype.

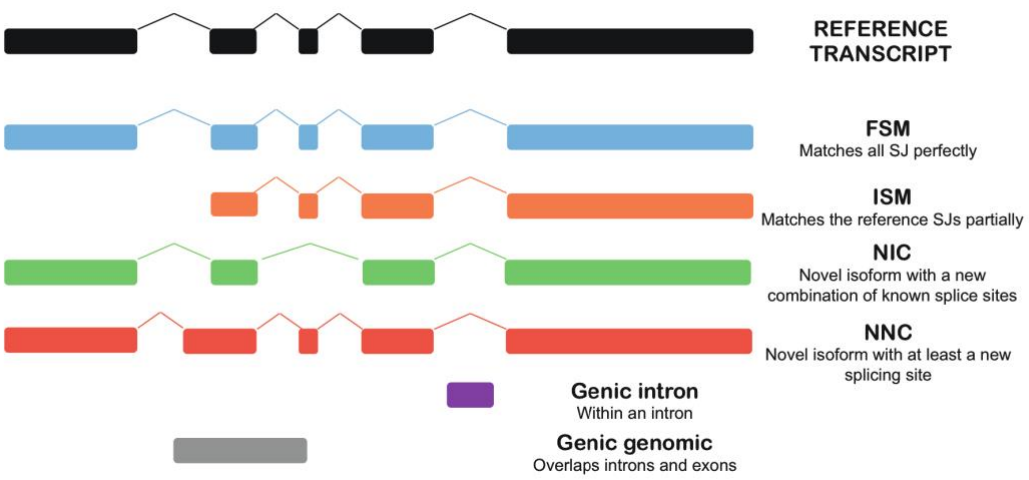


Figure 5.3. Schematic of SQANTI3 isoform structural characterization

(image taken from https://github.com/ConesaLab/SQANTI3/wiki/SQANTI3-isoform-classification:-categories-and-subcategories). FSM, full splice match; ISM, incomplete splice match; NIC, novel in catalog; NNC, novel, not in catalog. Intron retention events are not depicted but would fall within ISM, NIC, NNC, or fusion categories.

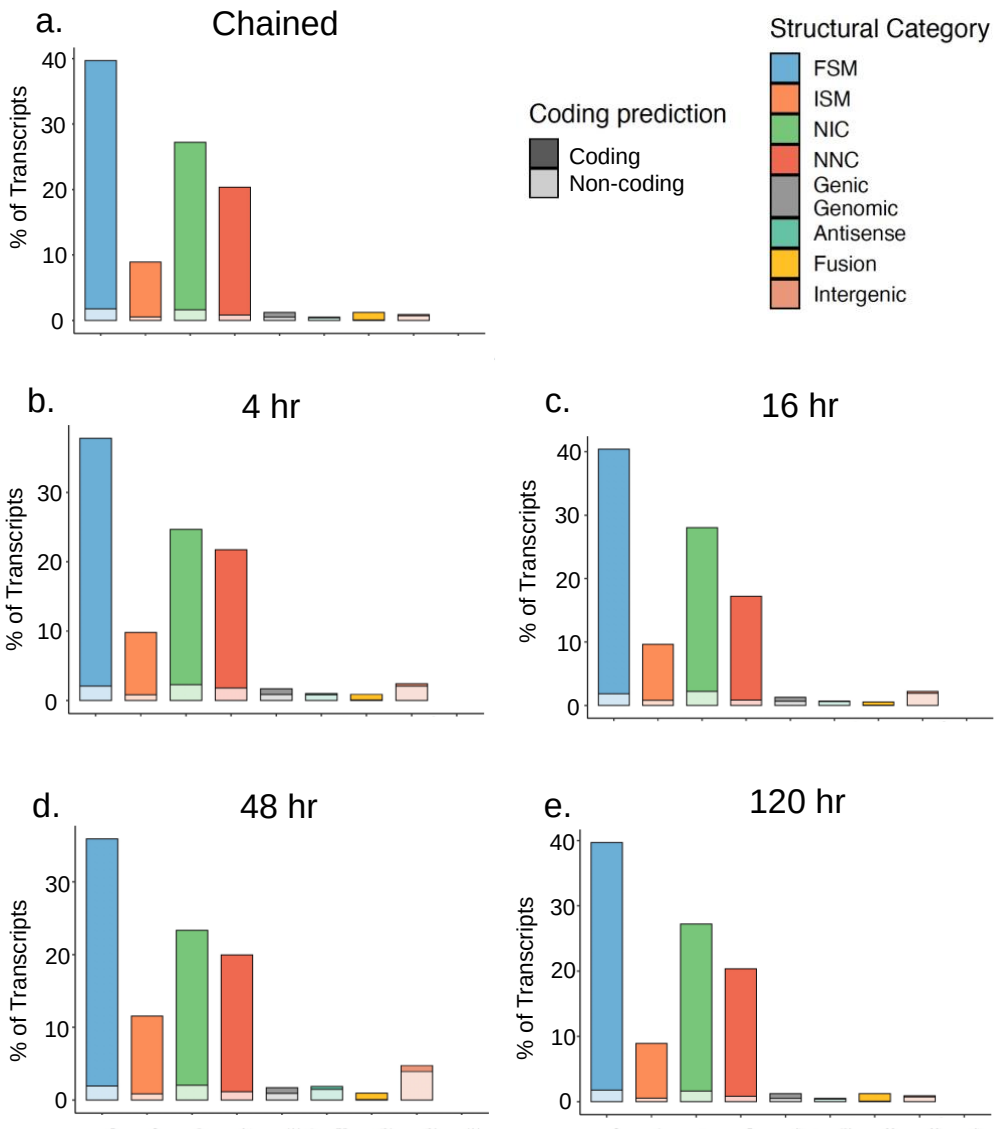


Figure 5.4. Proportions of predicted protein-coding isoforms within each structural category. Each graph represents data from a chained **(a)** or per-sample time point (4-120 hrs, **(b-e)** analysis of Iso-Seq performed at varying time points of CD4 T cell activation. Bar graphs depict the proportions of transcripts predicted to be protein-coding. Bar height indicates the percent of transcripts within each structural category out of the total transcripts present. Darker or lighter shading within a bar represents the proportion of the category predicted to be protein-coding or non-coding, respectively. Bar order from left to right matches the structural category listing from top to bottom.

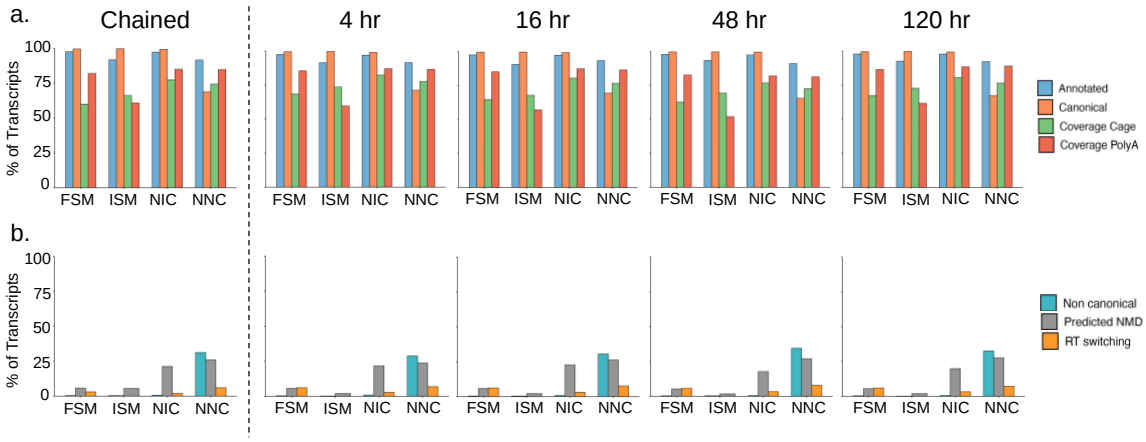


Figure 5.5. Quality metrics of major isoforms categories. Each graph represents data from a chained (far left) or per-sample time point (4-120 hrs, left to right) analysis of Iso-Seq performed at varying time points of CD4 T cell activation. **(a)** Metrics of good quality, the percent of transcripts within the noted categories which have: Annotated, mapped to annotated genes; Canonical, all canonical splice junctions; Coverage Cage, an identified CAGE Peak; Coverage PolyA, an identified polyA motif; Coverage SJ, supporting short-read coverage of all splice junctions. **(b)** Metrics of bad quality, the percent of transcripts within the noted categories which have: Non canonical, at least one noncanonical splice junction; Not Coverage SJ, no short-read coverage of at least one splice junction; Predicted NMD, predicted nonsense-mediated decay; RT switching, predicted RT switching occurrence.

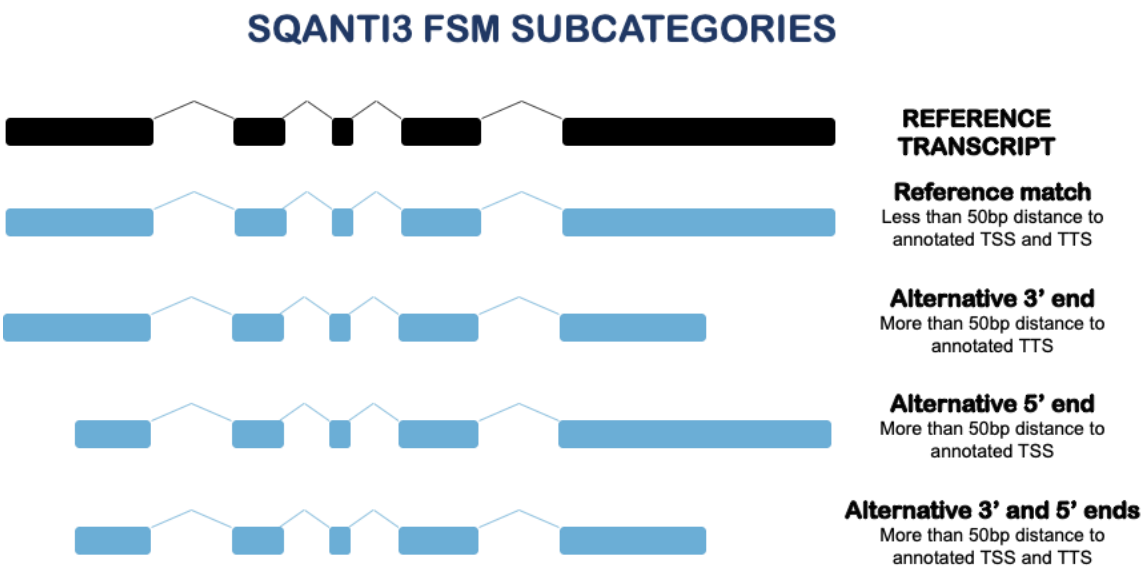


Figure 5.6. Schematic of SQANTI3 FSM subcategory classifications

(image taken from https://github.com/ConesaLab/SQANTI3/wiki/SQANTI3-isoform-classification:-categories-and-subcategories).

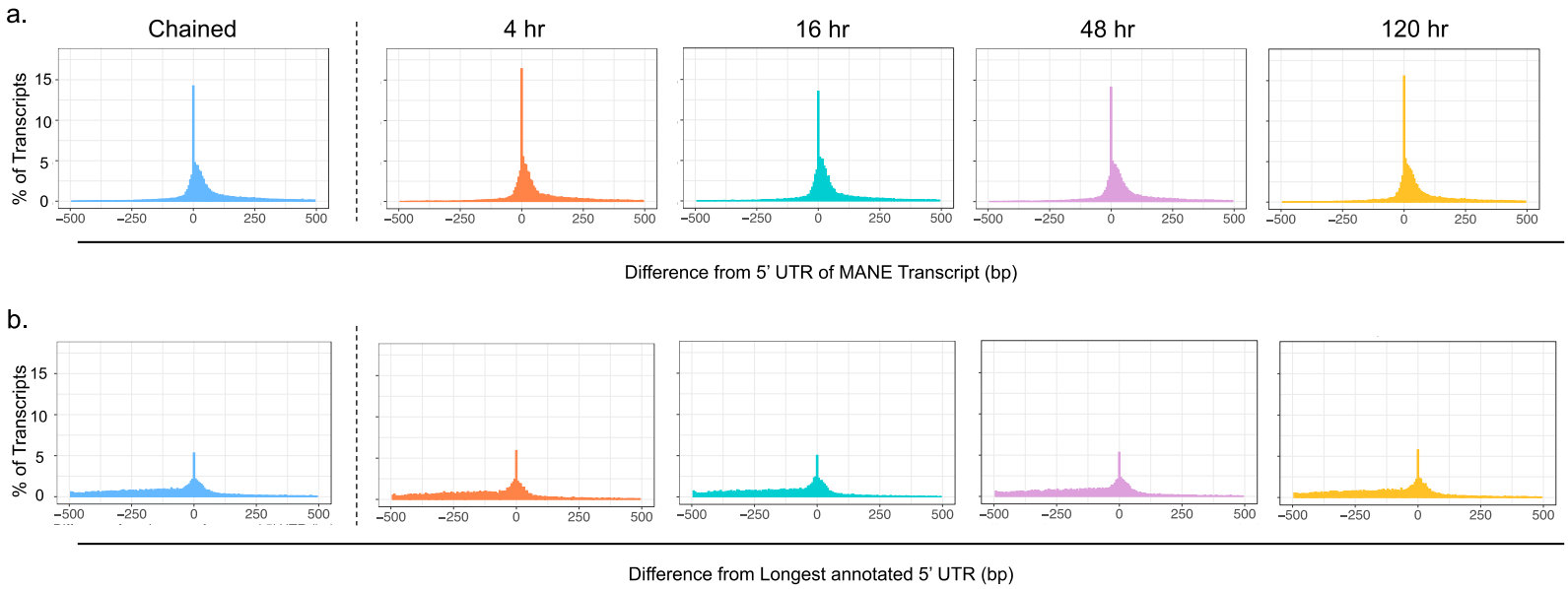


Figure 5.7. Distribution of transcripts with lengthened 5' UTR. A custom script was written to compare 5' UTR lengths of individual sample transcripts to either the MANE transcript 5' UTR **(a)** or the longest Genocode (v39, comprehensive) annotated 5' UTR **(b)** of the parent gene. Only protein-coding transcripts mapping to a known gene and falling within the FSM, ISM, NNC, or NIC categories were compared in this manner. Histograms were generated using publicly available R packages ggplot2 and dplyr. The % of total transcripts possessing specific UTR differences (bp) were calculated per sample by counting the number of transcripts possessing each difference value and dividing by the total number of transcripts analyzed. UTR differences outside of +/- 500 are not shown.

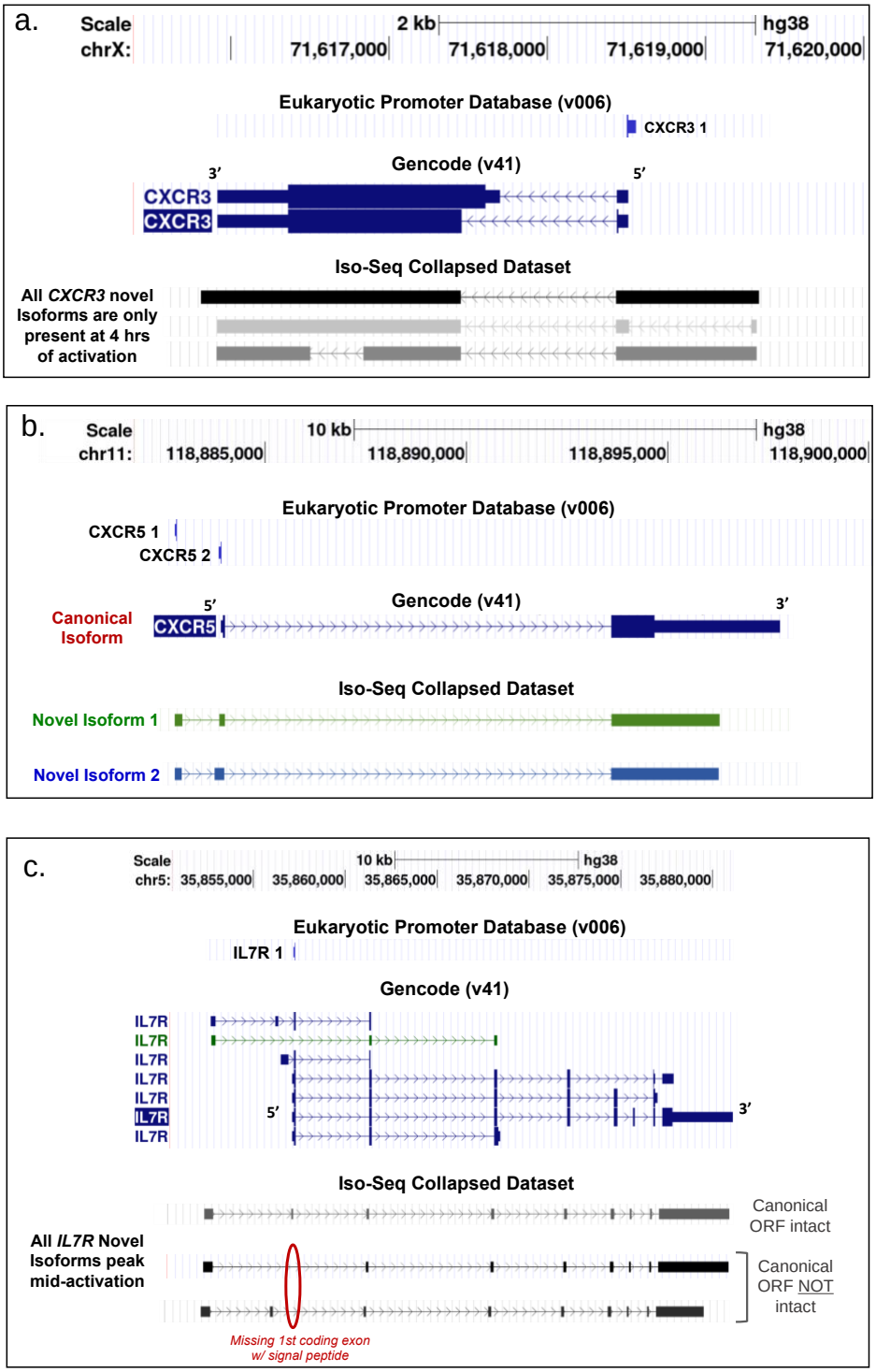


Figure 5.8. Immune-relevant genes with novel 5' end-variant transcripts. 5' end-variant isoforms of *CXCR3* **(a)**, *CXCR5* **(b)**, and *IL7R* **(c)** found within our activated Iso-Seq datasets. Images were taken from the UCSC Genome Browser. Tracks shown top to

bottom are the chromosomal location for the field of view, promoter locations documented in the human Eukaryotic Promoter Database (v006), Gencode (v41) reference transcripts, and finally, representative collapsed isoforms taken from the 4 hr time point of activation. For all transcripts depicted, bars represent exons, lines represent introns, and arrows indicate 5' to 3' directionality. For the negative-stranded gene *CXCR3* **(a)**, transcripts read 5' to 3' from right to left. For the positive-stranded genes *CXCR5* **(b)** and *IL7R* **(c)**, transcripts read 5' to 3' from left to right. Protein coding sequences are indicated on Gencode tracks by thicker bars within exonic regions.

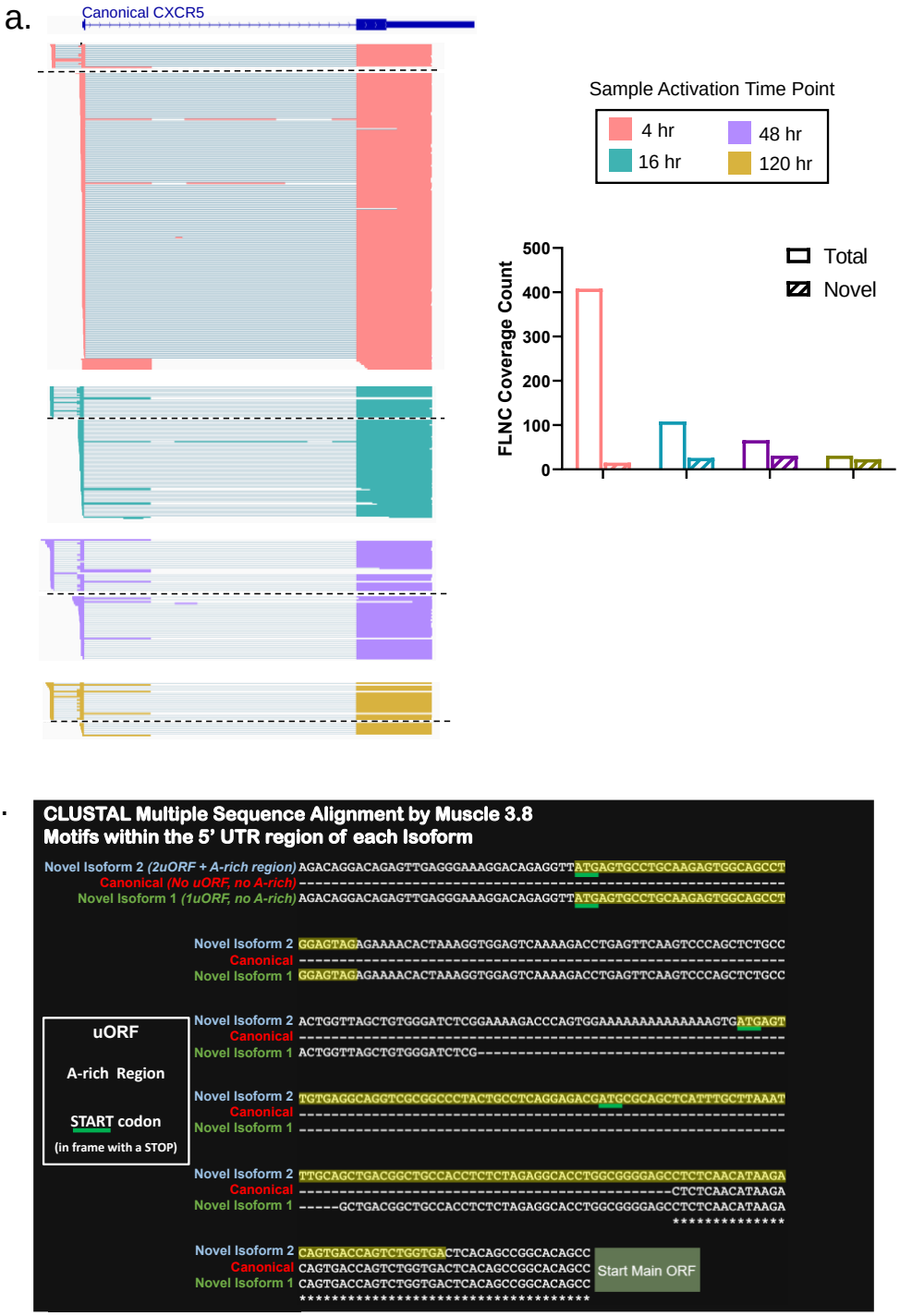


Figure 5.9. Novel end-variant transcripts of *CXCR5* increase in frequency with activation and possess potential intrinsic regulatory elements. (a) Mapped FLNC reads observed in the integrated genome viewer (IGV). Reads were grouped according to

their coverage of the novel upstream exon found only within end-variants. The canonical *CXCR5* Isoform is shown for reference. Each un-collapsed full-length read is depicted with exons as shaded boxes and anticipated introns as connecting lines. Direct FLNC coverage at the canonical *CXCR5* transcription start locus and coverage of the last base in the novel upstream exon were graphed as total and novel FLNC coverage counts, respectively. Sequenced time points of activation are represented by different colors both in the IGV view and in the read count graph. **(b)** Motifs within the 5' UTR of novel isoforms. CLUSTAL Multiple Sequence Alignment by Muscle 3.8 was performed for the collapsed 5' UTR sequences of Novel Isoform 1, Novel Isoform 2, and the Canonical *CXCR5* Isoform. Motifs were identified by NCBI ORF finder or independent observation. Identified motifs within the 5' UTR of each isoform are highlighted. uORF regions are highlighted yellow, A-rich regions are highlighted orange, and upstream start codons possessing an in-frame STOP are underlined with green.

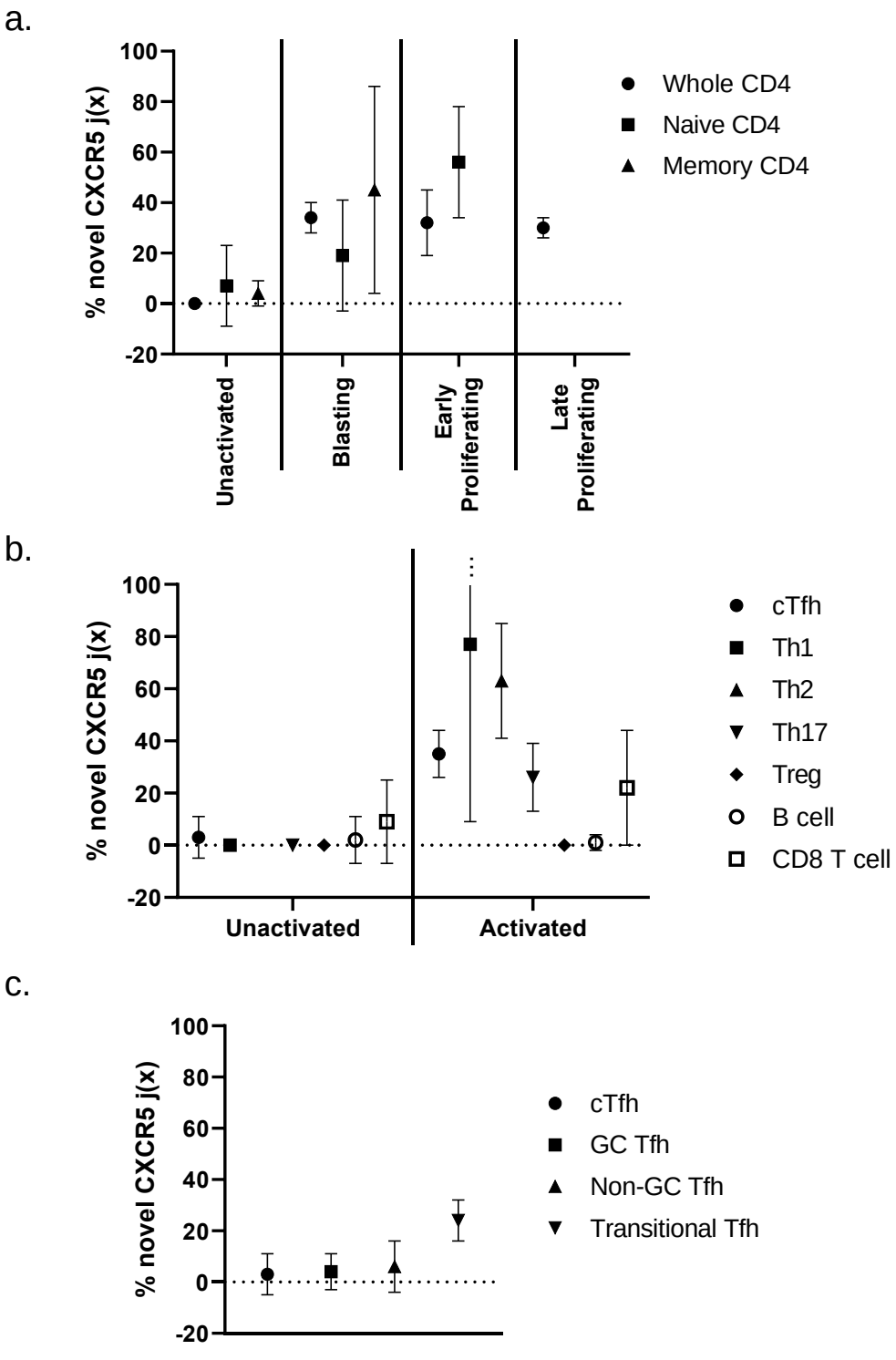


Figure 5.10. Evidence of novel *CXCR5* isoform junctional coverage across previous RNA-Seq samples. Previous RNA-Seq data were pulled from the NIH Sequence Read

Archive. Samples from 216 unique donors expressing *CXCR5* were categorized according to cell subset and activation state. Unique donors per category are listed in **Table 5.9**. Only donors expressing *CXCR5* were considered for analysis. The % novel *CXCR5* j(x) is an estimated proportion of novel isoform expression, calculated per sample by dividing the coverage of our novel end-variant isoform junctions (Chr11:118882935,118883748 and Chr11:118882935,118883851) by the total coverage at the shared *CXCR5* junction (Chr11:118883993,118893595). The average of this metric per subset category is shown with error bars representing standard deviation. No additional statistical tests were performed. **(a)** Unactivated or activated whole CD4, naïve CD4, or memory CD4 T cells. Activated cells were further classified by activation time as blasting (4-48 hr), early proliferating (72-120 hr), or late proliferating (>120 hr). **(b)** Unactivated or activated samples of B cells, CD8 T cells, or various subsets of CD4 T helper (Th) cells: circulating T follicular helper (cTfh), Th1, Th2, Th17, and T regulatory cells (Treg). **(c)** Samples of T follicular helper (Tfh) cells: cTfh, GC Tfh (germinal center Tfh), Non-GC Tfh (non-germinal center Tfh), or *ex-vivo* transitional Tfh.

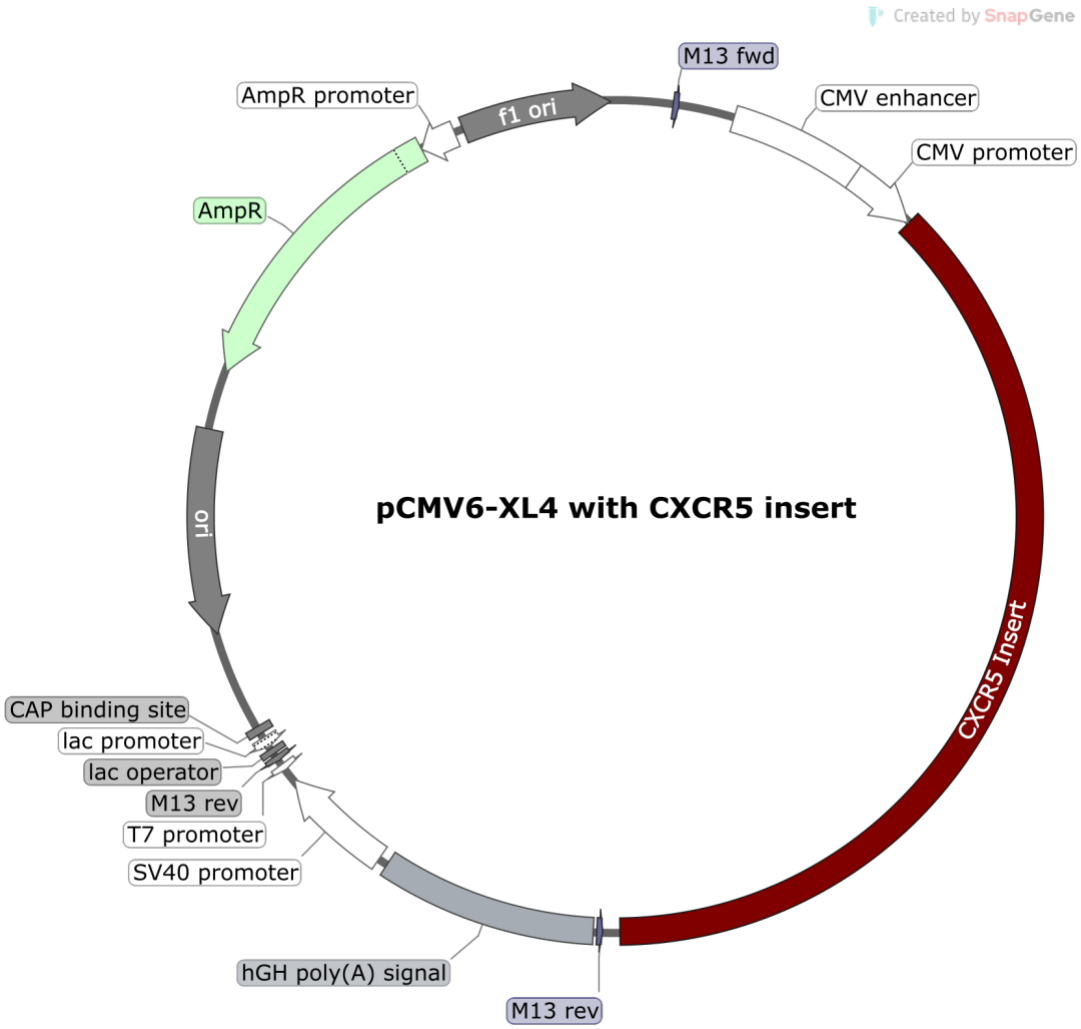


Figure 5.11. Plasmid elements of pCMV6-XL4. Plasmids were generated through the services of BlueHeron Bio (Bothell, WA). The location of *CXCR5* inserts is shown. *CXCR5* inserts contained variant 5' UTR upstream of the standard *CXCR5* CDS and 3' UTR of NM_001716.2.

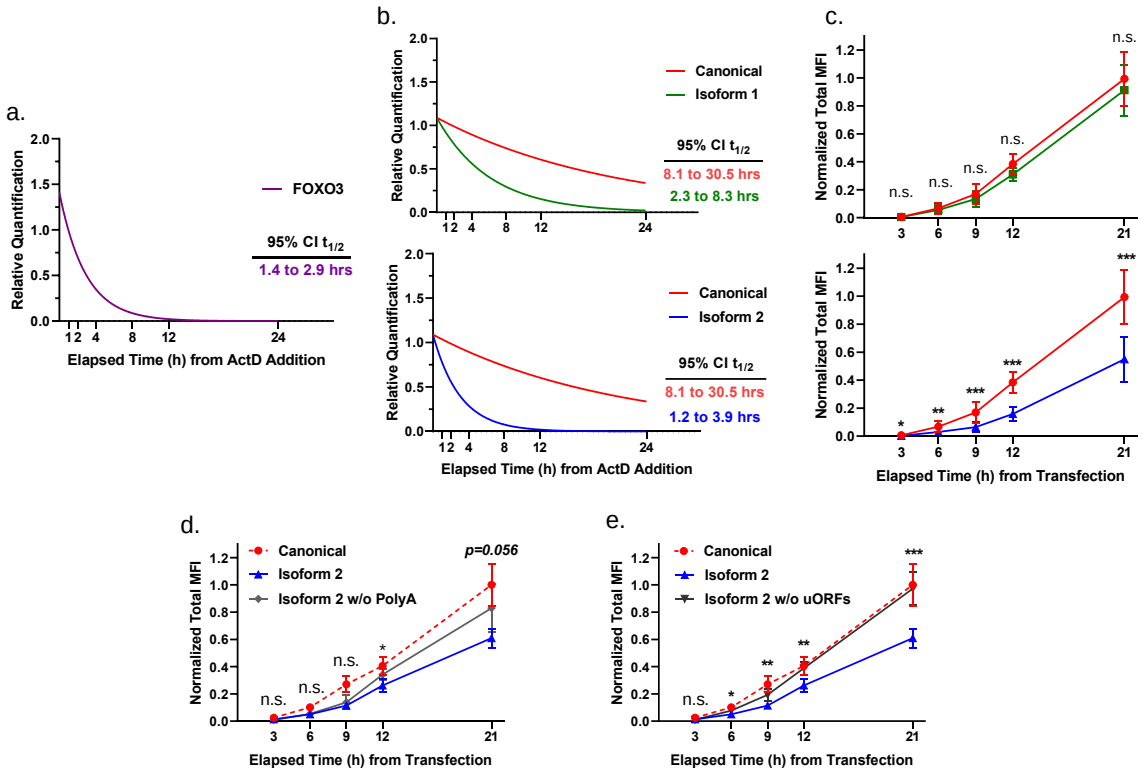


Figure 5.12. Differential mRNA stability and protein expression conferred by novel *CXCR5* isoforms. (a-b) mRNA stability assessed by an Actinomycin D (ActD) transcription inhibition assay. Relative quantification was performed with normalization to 1-hour post-ActD addition. Graphs depict nonlinear regression curves for one-phase decay with constraints of plateau = 0 and shared Y_o. 95% confidence intervals (CI) for the curves are shaded. 95% CI for mRNA half-lives ($t_{1/2}$) were calculated from regression analyses. Datapoints included in analyses were 1, 2, 4, 8, 12, and 24 hr time points after ActD addition. **(a)** One-phase decay of a known unstable transcript, *FOXO3*, which is endogenously expressed within HEK293T cells. *FOXO3* was quantified alongside *CXCR5* in every experiment as a metric of assay quality. The curve depicted was generated using data points collected across all 11 experiments. **(b)** One-phase decay of *CXCR5* mRNA isoforms. Isoforms were uniquely expressed in HEK293T cells by

transfection of plasmids engineered to transcribe only one mRNA variant. Nonlinear regression curves incorporate data across three (Canonical) or four (Novel Isoforms 1 and 2) independent experiments. One outlier at the 4-hour time point for Isoform 2 was excluded. **(c-e)** CXCR5 display kinetics for variant mRNA isoforms occurring during synchronized transfection. Plasmids expressing each mRNA isoform were transfected into HEK293T cells. Transfection was allowed to proceed unhindered for 3 hours prior to removal of transfection reagents and observation of protein expression kinetics using flow cytometric staining for surface CXCR5. CXCR5 expression was observed at 3, 6, 9, 12, and 21 hours after initial transfection. Data were normalized per experiment to the 21-hour average of Canonical isoform expression. Significance is reported based on adjusted p-values of multiple T-tests using the Holm-Sidak method to correct for multiple comparisons; n.s. not significant, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. **(c)** Expression kinetics from plasmids transcribing either Canonical, Isoform 1, or Isoform 2. For each plasmid, N=13 transfections (3 experiments, 3-5 independent plasmid preps per experiment). **(d-e)** Expression kinetics from plasmids transcribing either Canonical, intact Isoform 2, or mutant variants of Isoform 2. Canonical and intact Isoform 2 were transfected in parallel to mutant variants of Isoform 2 for these experiments. For canonical and intact Isoform 2, N=6 transfections (2 experiments, 3 independent plasmid preps). For the mutant variants of Isoform 2, N= 8 transfections (2 experiments, 4 independent plasmid preps). Significance is reported for mutant variants compared to intact Isoform 2 expression.

CHAPTER VI

DISCUSSION AND IMPLICATIONS

SIGNIFICANCE OF FINDINGS IN FCMR REGULATION

Implications for hypotheses of FcMR function in human T cells

Even with well-documented occurrences of ligand-induced endocytosis (50-53), we found that soluble IgM is not a modulator of FcMR surface expression on lymphocytes. Instead, FcMR surface expression appeared to be regulated by a nontraditional mechanism involving cell-density-driven inhibition. Aside from lending better information to shape future *in vitro* studies of this receptor, this discovery has implications in both forming, testing, and refining FcMR functional hypotheses, particularly for human T cells where the receptor's function is not well understood.

The prevailing hypothesis for the role of FcMR in human T cells is that internalization of IgM-FcMR complexes upregulates T cell costimulatory molecules (53). In this hypothesis, circulating levels of IgM would be of considerable importance in determining T-cell sensitivity to activation. During the early stages of an infection, when IgM titers are high, increased FcMR stimulation may heighten T-cell sensitivity to activation before the initiation of the T-cell-dependent antibody response and class switching. However, based on the previous understanding of FcMR regulation (53, 106, 108, 110), FcMR surface expression was predicted to be lower in these high IgM environments, decreasing the likelihood of its functionality. In contrast, our findings

predict relatively constant levels of FcMR on circulating T cells, even in high IgM environments. This would allow increased sensitivity to broad ranges of IgM that could more directly couple the IgM-productive B cell response to FcMR-mediated T cell priming. A T cell could thus keep FcMR high until additional stimuli, such as activation or activation-adjacent signals (53), lead to the downregulation of the receptor when it may no longer be needed.

Outside of adding layers to the prevailing hypothesis of FcMR function in human T cells, our findings support additional hypotheses involving the continuous sampling of IgM, which would not be possible with direct ligand-mediated regulation. One such hypothesis is that FcMR-mediated internalization of pathogen-associated molecular patterns (PAMPs) triggers internal pattern-recognition receptors (PRRs). In non-T cells, antigen may be internalized through mechanisms such as Fc gamma receptor (FcγR) binding and endocytosing complexed antigens (241, 242). Once inside the cell, PAMPs can bind internal PRRs such as TLR3, TLR7, TLR9, or other internal detection molecules such as Caspase-4/5 which may directly bind internalized LPS (243-245). FcMR is the only constitutively expressed Fc receptor in human T cells (108), and naive T cells where FcMR is most expressed do not have many highly expressed cell-surface PRRs (246). It is possible that FcMR in these cells provides a mechanism to sample IgM-bound PAMPs from the extracellular, thus providing further context during T cell activation.

The lack of previous consideration for FcMR to play a role in sensing the extracellular environment is likely due, in part, to the assumption that FcMR is not available in high IgM environments if it is directly downregulated through the binding and internalization of IgM. However, based on our study of FcMR regulation, a

continued sampling of environmental IgM by patrolling T cells becomes more biologically likely. Investigations into this and other hypotheses requiring FcMR-mediated IgM sampling would be interesting directions for future studies. Further analyses of FcMR-mediated regulation in T cells might also help to support or deny such hypotheses and lend clues to additional theories into the roles of FcMR.

Implications for FcMR as a target in Chronic Lymphocytic Leukemia

The ability to predict receptor availability is important in the design of targeted therapeutics and in understanding receptors as biomarkers of disease. FcMR is a known biomarker, and potential target for the treatment, of Chronic Lymphocytic Leukemia (CLL), where it is elevated on leukemic B cells as well as on non-leukemic B and T cells (108, 118). Our studies into mechanisms regulating the expression of FcMR on healthy lymphocytes provide insight into how this receptor may behave during FcMR-targeted treatments (54, 55).

Though CLL is primarily a hematopoietic disease, states of natural cell crowding exist within the primary reservoirs of bone marrow and lymph nodes (247). Yet, CLL cells taken from circulation have been most often used for *ex vivo* measurement of FcMR and to predict therapeutic efficacy (52, 54, 55). If FcMR on CLL cells behaves similarly to FcMR in our *in vitro* studies of the effects of cell density on healthy lymphocytes, CLL cells in cell-crowded bone marrow and lymph node environments may have less cell surface FcMR available than has been assumed. Less available FcMR could lessen therapeutic efficacy in these reservoirs and prevent the complete elimination of leukemic cells. Future studies integrating our newly illuminated mechanisms of receptor regulation

with previous understandings of FcMR expression are needed to estimate the true availability of this receptor as a target of therapeutic interventions.

SIGNIFICANCE OF FINDINGS TOWARDS ISOFORM-BASED REGULATION

Iso-Seq in the discovery of novel end-variant isoforms

Using Iso-Seq, we generated isoform-aware transcriptomes for human lymphocytes and activated CD4 T cells that revealed many novel isoforms, a large proportion of which were end-variants with novel UTRs. It is known that mRNA and protein expression are not directly equivalent, and end-variant regulation may be an under-explored but important mechanism contributing to this disassociation (90-93). Several examples of specific end-variant regulation in human immune cells have been found (95, 153, 157-159, 163, 168, 248), but we may still be overlooking many end variants and further missing changes in global end-variation. With few human immune cell types and states well-annotated with isoform-aware sequencing, there is a continued need for Iso-Seq studies cataloging the end-variant landscape during the human immune response.

In addition to enhancing our understanding of post-transcriptional regulatory mechanisms, a comprehensive understanding of end-variation would also add to our ability to adequately annotate transcriptionally relevant regions such as promoters and enhancers. Many aspects of promoter and enhancer regions have been shown to overlap, including the ability to start transcription near the element's locus (249, 250). Recent models have even proposed that promoter and enhancer elements might exist on a spectrum rather than as independent categories (249, 250). Iso-Seq can provide additional

insight into distinguishing these elements by providing valuable information on transcription start sites and respective transcript characteristics, as in visualizing bidirectional starts that would be typical of more enhancer-like TSS (249, 250). The end-variant transcripts discovered with bulk Iso-Seq may be particularly helpful in corroborating previous CAGE Peak and chromatin accessibility data that predict specific regulatory loci (219, 249, 251).

The ability of Iso-Seq to unambiguously detect end-variation may also be important in mapping cell-state-specific transcriptional control of immune-important genes. For example, the novel *CXCR5* end-variant isoforms discovered here appear to map to a known promoter whose role in regulating Tfh differentiation has not yet been cataloged or studied (212). Future studies could investigate the mechanisms of regulation occurring at this locus, particularly as *CXCR5* expression at the time of end-variant appearance is thought to be independent of the transcription factor Bcl6 which is oft considered vital in controlling *CXCR5* expression (77-79, 88, 89).

Without appropriate end-variant characterization, current annotations of transcriptional regulation for many genes are likely incomplete. It is probable many TSSs remain to be discovered and potentially important variants of the transcripts produced at even recognized TSS remain undocumented. End variants found in our and future Iso-Seq studies will thus help better define the complete transcriptional and translational regulatory atlas of human immune receptors.

Implications for our understanding of receptors as biomarkers of disease

Surface receptors are often used as biomarkers of disease as changes in their expression reflect the intrinsic and extrinsic factors that affect various cell states. Since the advent of next-generation sequencing and a decrease in the cost of sequencing technology, genetic and transcriptional markers have also become widely used in diagnostics (252, 253). Differential expression is frequently considered in this context, though differential RNA splicing events have also been appreciated as a novel area to look for biomarkers of cancer or infectious disease (138, 179, 181, 254). Even differential end-variation of isoforms has been suggested for possible diagnostic capability (96, 149, 255, 256). Single-molecule sequencing approaches such as Iso-Seq offer the unique opportunity to enhance investigations into the differential expression of splice- and end-variant isoforms, which may not only be recognized as biomarkers but may also have potential consequences in controlling levels of disease-associated proteins.

For example, global 5' UTR alterations and differential first exon usage have been identified as likely mechanisms influencing protein expression in cancer (148, 255, 256). For prostate cancer, point mutations within the 5' UTR have even been associated with varying clinical outcomes (255). As cancer cells constantly adapt to evade immune surveillance, slight alterations in regulatory regions that drastically affect protein expression without creating neo-antigens, could be highly advantageous to these cells. End-variation and mutations within 5' UTR are likely more widespread in cancer cells than is currently appreciated, where they could thus play an overlooked role in the evasion of immune surveillance.

Post-transcriptional regulation is thought to play a role in the dysfunction and exhaustion of T-cells in cancer patients (257), suggesting end-variants might serve to distinguish amongst various immune states that could potentially predict immunotherapy responsiveness. Yet, the landscape of end-variant isoforms in tumor-invading and circulating cells of cancer patients is relatively understudied. Additional Iso-Seq surveys in cancer are warranted to further distinguish end-variants as possible biomarkers of disease or therapeutic efficacy.

Aside from enlightening end-aware nuances in differential expression across cancer, Iso-Seq discovery of end-variation holds great potential for clarifying the significance of single nucleotide polymorphisms (SNPs) found in Genome-Wide Association Studies (GWAS). For many disease-associated SNPs, the exact implications of their locus remain unknown (258). This is particularly true for SNPs in regions outside of annotated transcripts that are not within documented promoter, enhancer, or other regulatory elements. The significance of these intergenic SNPs is often chalked up to the interruption of an undocumented long-range regulatory region (259). However, it is possible some of these SNPs could be within the UTR of novel and unexplored end-variant isoforms or adjacent to as-yet-undiscovered transcription start sites. These hypotheses are supported by a recent study involving the re-annotation of transcript expression that found many previously “intergenic” SNPs were actually within unannotated, novel transcripts (260). Though few novel transcripts found in this study were predicted to be protein-coding (260), this highlights the importance of continuous annotation of possible isoforms in various cell states to best predict the consequences of disease-associated polymorphisms.

Alongside disease-associated SNPs, rare diseases that are directly caused by mutations in the 5’ UTR of mRNA also exist that are often missed by diagnostics methods centered on detecting mutations in internal exons (261). In such cases, visualization of intact mRNA molecules with Iso-Seq may be useful to classify and characterize both mutations occurring in the protein-coding regions as well as in the noncoding regions of the mRNA. Iso-Seq thus offers unique opportunities for biomarker discovery, clarity of genetic disease associations, and direct diagnostics.

Implications of end-variant isoforms in designing targeted therapeutics

A well-informed catalog of end-variants improves our ability to design therapeutics that use or target elements within the 5’ UTR. In generating novel mRNA-based therapeutics, for example, annotated UTRs are often used to predict the stability and expression outcomes of prospective constructs (262, 263). By enhancing 5’ UTR annotation using Iso-Seq, we may also improve the design of UTRs in mRNA-based therapeutics. Improved annotation of 5’ UTRs could further enhance the design of catered treatments targeting specific aspects of UTR-based regulation in disease. For instance, in Alzheimer’s disease, the amyloid precursor-protein (APP) mRNA is controlled by an iron-responsive element within the 5’ UTR that increases protein expression and directly contributes to disease pathology (264, 265). Taking advantage of this mechanism, drugs targeting the chelation of iron to decrease APP expression have been tested in clinical trials (264, 265). Because of its ability to pick up subtle alterations in the 5’ regulatory regions, Iso-Seq may be a useful tool in laying the groundwork for this and other UTR sequence-informed treatments.

One promising 5’ UTR-associated target involves cancer-specific changes to uORFs. Certain uORFs have been shown to produce small peptides that may be displayed to patrolling T cells via the MHC I presentation pathway that, when altered by mutations, serve as neoantigens that can initiate anti-cancer T-cell response (266). Such uORF-encoded peptides could serve as novel targets for cellular therapies. However, without understanding the extent of end-variants in cancer, we cannot fully understand and reconstruct the landscape of uORF neoantigens that may be present. Few studies have attempted to catalog end-variants with uORFs that may encode peptides used in MHC-based surveillance. Iso-Seq offers a unique opportunity to characterize end-variation within diseased and healthy tissue, catalog 5’ UTRs, and observe potential uORFs to predict short coding sequences that may be used for MHC I display.

CONCLUSIONS

Together, these studies have contributed to a more comprehensive understanding of immune-relevant receptors by positing novel mechanisms of receptor regulation. When integrated with previous research, our findings may help to anticipate patterns of receptor expression during targeted treatments and further contribute to the identification of efficient disease biomarkers or therapeutic targets. Overall, our research has provided a valuable contribution to the field of immunology and laid the groundwork for future studies investigating how the mechanisms described here could affect many aspects of immune-related receptor regulation.

REFERENCES

1. Jung, T. M., W. M. Gallatin, I. L. Weissman, and M. O. Dailey. 1988. Down-regulation of homing receptors after T cell activation. *The Journal of Immunology* 141: 4110-4117.
2. Ansel, K. M., L. J. Mcheyzer-Williams, V. N. Ngo, M. G. Mcheyzer-Williams, and J. G. Cyster. 1999. In Vivo–Activated Cd4 T Cells Upregulate Cxc Chemokine Receptor 5 and Reprogram Their Response to Lymphoid Chemokines. *Journal of Experimental Medicine* 190: 1123-1134.
3. Sallusto, F., E. Kremmer, B. Palermo, A. Hoy, P. Ponath, S. Qin, R. Förster, M. Lipp, and A. Lanzavecchia. 1999. Switch in chemokine receptor expression upon TCR stimulation reveals novel homing potential for recently activated T cells. *European Journal of Immunology* 29: 2037-2045.
4. Kunkel, E. J., J. Boisvert, K. Murphy, M. A. Vierra, M. C. Genovese, A. J. Wardlaw, H. B. Greenberg, M. R. Hodge, L. Wu, E. C. Butcher, and J. J. Campbell. 2002. Expression of the Chemokine Receptors CCR4, CCR5, and CXCR3 by Human Tissue-Infiltrating Lymphocytes. *The American Journal of Pathology* 160: 347-355.
5. Pereira, J. P., L. M. Kelly, and J. G. Cyster. 2010. Finding the right niche: B-cell migration in the early phases of T-dependent antibody responses. *Int Immunol* 22: 413-419.

6. Gatto, D., K. Wood, and R. Brink. 2011. EBI2 Operates Independently of but in Cooperation with CXCR5 and CCR7 To Direct B Cell Migration and Organization in Follicles and the Germinal Center. *The Journal of Immunology* 187: 4621-4628.

7. Shipkova, M., and E. Wieland. 2012. Surface markers of lymphocyte activation and markers of cell proliferation. *Clinica Chimica Acta* 413: 1338-1349.

8. Kalina, T., K. Fišer, M. Pérez-Andrés, D. Kuzílková, M. Cuenca, S. J. W. Bartol, E. Blanco, P. Engel, and M. C. van Zelm. 2019. CD Maps—Dynamic Profiling of CD1–CD100 Surface Expression on Human Leukocyte and Lymphocyte Subsets. *Frontiers in Immunology* 10.

9. Shilts, J., Y. Severin, F. Galaway, N. Müller-Sienerth, Z.-S. Chong, S. Pritchard, S. Teichmann, R. Vento-Tormo, B. Snijder, and G. J. Wright. 2022. A physical wiring diagram for the human immune system. *Nature* 608: 397-404.

10. Bausch-Fluck, D., U. Goldmann, S. Müller, M. Van Oostrum, M. Müller, O. T. Schubert, and B. Wollscheid. 2018. The in silico human surfaceome. *Proceedings of the National Academy of Sciences* 115: E10988-E10997.

11. Overington, J. P., B. Al-Lazikani, and A. L. Hopkins. 2006. How many drug targets are there? *Nature Reviews Drug Discovery* 5: 993-996.

12. Yao, S., Y. Zhu, and L. Chen. 2013. Advances in targeting cell surface signalling molecules for immune modulation. *Nature Reviews Drug Discovery* 12: 130-146.

13. Allman, D., and S. Pillai. 2008. Peripheral B cell subsets. *Current Opinion in Immunology* 20: 149-157.

14. Laurent, C., N. Fazilleau, and P. Brousset. 2010. A novel subset of T-helper cells: follicular T-helper cells and their markers. *Haematologica* 95: 356-358.

15. Kim, C. H., L. Rott, E. J. Kunkel, M. C. Genovese, D. P. Andrew, L. Wu, and E. C. Butcher. 2001. Rules of chemokine receptor association with T cell polarization in vivo. *Journal of Clinical Investigation* 108: 1331-1339.

16. Sallusto, F. 2016. Heterogeneity of Human CD4⁺ T Cells Against Microbes. *Annual Review of Immunology* 34: 317-334.

17. Zielinski, C. E. 2023. T helper cell subsets: diversification of the field. *European Journal of Immunology*: 2250218.

18. Kiner, E., E. Willie, B. Vijaykumar, K. Chowdhary, H. Schmutz, J. Chandler, A. Schnell, P. I. Thakore, G. Legros, S. Mostafavi, D. Mathis, C. Benoist, O. Aguilar, R. Allan, J. Astarita, K. F. Austen, N. Barrett, A. Baysoy, C. Benoist, B. D. Brown, M. Buechler, J. Buenrostro, M. A. Casanova, K. Choi, K. Chowdhary, M. Colonna, T. Crowl, T. Deng, J. V. Desai, F. Desland, M. Dhainaut, J. Ding, C. Dominguez, D. Dwyer, M. Frascoli, S. Gal-Oz, A. Goldrath, R. Grieshaber-Bouyer, B. Jia, T. Johanson, S. Jordan, J. Kang, V. Kapoor, E. Kenigsberg, J. Kim, K. Wook Kim, E. Kiner, M. Kronenberg, L. Lanier, C. Laplace, C. Lareau, A. Leader, J. Lee, A. Magen, B. Maier, A. Maslova, D. Mathis, A. Mcfarland, M. Merad, E. Meunier, P. Monach, S. Mostafavi, S. Muller, C. Muus, H. Ner-Gaon, Q. Nguyen, P. A. Nigrovic, G. Novakovsky, S. Nutt, K. Omilusik, A. Ortiz-Lopez, M. Paynich, V. Peng, M. Potempa, R. Pradhan, S. Quon, R. Ramirez, D. Ramanan, G. Randolph, A. Regev, S. A. Rose, K. Seddu, T. Shay, A. Shemesh, J. Shyer, C. Smilie, N. Spidale, A. Subramanian, K. Sylvia, J. Tellier, S. Turley, B.

Vijaykumar, A. Wagers, C. Wang, P. L. Wang, A. Wroblewska, L. Yang, A. Yim, and H. Yoshida. 2021. Gut CD4+ T cell phenotypes are a continuum molded by microbes, not by TH archetypes. *Nature Immunology* 22: 216-228.

19. Murphy, K. M., and B. Stockinger. 2010. Effector T cell plasticity: flexibility in the face of changing circumstances. *Nature Immunology* 11: 674-680.

20. Sheikh, A. A., and J. R. Groom. 2021. Transcription tipping points for T follicular helper cell and T-helper 1 cell fate commitment. *Cellular & Molecular Immunology* 18: 528-538.

21. Haynes, N. M., C. D. C. Allen, R. Lesley, K. M. Ansel, N. Killeen, and J. G. Cyster. 2007. Role of CXCR5 and CCR7 in Follicular Th Cell Positioning and Appearance of a Programmed Cell Death Gene-1^{High} Germinal Center-Associated Subpopulation. *The Journal of Immunology* 179: 5099-5108.

22. Hardtke, S., L. Ohl, and R. FöRster. 2005. Balanced expression of CXCR5 and CCR7 on follicular T helper cells determines their transient positioning to lymph node follicles and is essential for efficient B-cell help. *Blood* 106: 1924-1931.

23. Groom, J. R., and A. D. Luster. 2011. CXCR3 in T cell function. *Experimental Cell Research* 317: 620-631.

24. Olatunde, A. C., J. S. Hale, and T. J. Lamb. 2021. Cytokine-skewed Tfh cells: functional consequences for B cell help. *Trends in Immunology* 42: 536-550.

25. Yu, S., L. Jia, Y. Zhang, J. Zhong, B. Yang, and C. Wu. 2015. IL-12 induced the generation of IL-21- and IFN-γ-co-expressing poly-functional CD4+ T cells from human naive CD4+ T cells. *Cell Cycle* 14: 3362-3372.

26. Bentebibel, S.-E., S. Khurana, N. Schmitt, P. Kurup, C. Mueller, G. Obermoser, A. K. Palucka, R. A. Albrecht, A. Garcia-Sastre, H. Golding, and H. Ueno. 2016. ICOS+PD-1+CXCR3+ T follicular helper cells contribute to the generation of high-avidity antibodies following influenza vaccination. *Scientific Reports* 6: 26494.

27. Schaerli, P., P. Loetscher, and B. Moser. 2001. Cutting Edge: Induction of Follicular Homing Precedes Effector Th Cell Development. *The Journal of Immunology* 167: 6082-6086.

28. Nakayamada, S., Y. Kanno, H. Takahashi, D. Jankovic, Kristina, Thomas, H.-W. Sun, G. Vahedi, O. Hakim, R. Handon, Pamela, Gordon, and John. 2011. Early Th1 Cell Differentiation Is Marked by a Tfh Cell-like Transition. *Immunity* 35: 919-931.

29. Bromley, S. K., S. Y. Thomas, and A. D. Luster. 2005. Chemokine receptor CCR7 guides T cell exit from peripheral tissues and entry into afferent lymphatics. *Nature Immunology* 6: 895-901.

30. Edner, N. M., G. Carlesso, J. S. Rush, and L. S. K. Walker. 2020. Targeting co-stimulatory molecules in autoimmune disease. *Nature Reviews Drug Discovery* 19: 860-883.

31. Chen, L., and D. B. Flies. 2013. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nature Reviews Immunology* 13: 227-242.

32. Hochweller, K., and S. M. Anderton. 2005. Kinetics of costimulatory molecule expression by T cells and dendritic cells during the induction of tolerance versus immunityin vivo. *European Journal of Immunology* 35: 1086-1096.

33. Zhu, Y., S. Yao, and L. Chen. 2011. Cell Surface Signaling Molecules in the Control of Immune Responses: A Tide Model. *Immunity* 34: 466-478.

34. Fujii, M., K. Sugamura, K. Sano, M. Nakai, K. Sugita, and Y. Hinuma. 1986. High-affinity receptor-mediated internalization and degradation of interleukin 2 in human T cells. *Journal of Experimental Medicine* 163: 550-562.

35. Hémar, A., A. Subtil, M. Lieb, E. Morelon, R. Hellio, and A. Dautry-Varsat. 1995. Endocytosis of interleukin 2 receptors in human T lymphocytes: distinct intracellular localization and fate of the receptor alpha, beta, and gamma chains. *Journal of Cell Biology* 129: 55-64.

36. Morelon, E., and A. Dautry-Varsat. 1998. Endocytosis of the Common Cytokine Receptor γcChain. *Journal of Biological Chemistry* 273: 22044-22051.

37. Puré, E., and L. Tardelli. 1992. Tyrosine phosphorylation is required for ligand-induced internalization of the antigen receptor on B lymphocytes. *Proceedings of the National Academy of Sciences* 89: 114-117.

38. Martínez-Martín, N., E. Fernández-Arenas, S. Cemerski, P. Delgado, M. Turner, J. Heuser, Darrell, B. Huang, Xosé, A. Shaw, and B. Alarcón. 2011. T Cell Receptor Internalization from the Immunological Synapse Is Mediated by TC21 and RhoG GTPase-Dependent Phagocytosis. *Immunity* 35: 208-222.

39. Dietrich, J., X. Hou, A. M. Wegener, and C. Geisler. 1994. CD3 gamma contains a phosphoserine-dependent di-leucine motif involved in down-regulation of the T cell receptor. *The EMBO Journal* 13: 2156-2166.

40. Valitutti, S., S. Müller, M. Salio, and A. Lanzavecchia. 1997. Degradation of T Cell Receptor (TCR)–CD3-ζ Complexes after Antigenic Stimulation. *Journal of Experimental Medicine* 185: 1859-1864.

41. Alcover, A., and B. Alarcón. 2000. Internalization and intracellular fate of TCR-CD3 complexes. *Crit Rev Immunol* 20: 325-346.

42. Linsley, P. S., J. Bradshaw, M. Urnes, L. Grosmaire, and J. A. Ledbetter. 1993. CD28 engagement by B7/BB-1 induces transient down-regulation of CD28 synthesis and prolonged unresponsiveness to CD28 signaling. *The Journal of Immunology* 150: 3161-3169.

43. Riha, P., and C. E. Rudd. 2010. CD28 co-signaling in the adaptive immune response. *Self/Nonself* 1: 231-240.

44. Lee, K.-H., A. R. Dinner, C. Tu, G. Campi, S. Raychaudhuri, R. Varma, T. N. Sims, W. R. Burack, H. Wu, J. Wang, O. Kanagawa, M. Markiewicz, P. M. Allen, M. L. Dustin, A. K. Chakraborty, and A. S. Shaw. 2003. The Immunological Synapse Balances T Cell Receptor Signaling and Degradation. *Science* 302: 1218-1222.

45. Keam, S. J. 2023. Teplizumab: First Approval. *Drugs*.

46. Kuhn, C., and H. L. Weiner. 2016. Therapeutic anti-CD3 monoclonal antibodies: from bench to bedside. *Immunotherapy* 8: 889-906.

47. Mignogna, C., E. Maddaloni, L. D'Onofrio, and R. Buzzetti. 2021. Investigational therapies targeting CD3 for prevention and treatment of type 1 diabetes. *Expert Opinion on Investigational Drugs* 30: 1209-1219.

48. Shultes, K. C. 2022. Loncastuximab Tesirine-lpyl (Zynlonta®). *Oncology Times* 44: 14.

49. Esapa, B., J. Jiang, A. Cheung, A. Chenoweth, D. E. Thurston, and S. N. Karagiannis. 2023. Target Antigen Attributes and Their Contributions to Clinically Approved Antibody-Drug Conjugates (ADCs) in Haematopoietic and Solid Cancers. *Cancers* 15: 1845.

50. Nguyen, T. T. T., K. Kläsener, C. Zürn, P. A. Castillo, I. Brust-Mascher, D. M. Imai, C. L. Bevins, C. Reardon, M. Reth, and N. Baumgarth. 2017. The IgM receptor FcμR limits tonic BCR signaling by regulating expression of the IgM BCR. *Nature Immunology* 18: 321-333.

51. Lloyd, K. A., J. Wang, B. C. Urban, D. M. Czajkowsky, and R. J. Pleass. 2017. Glycan-independent binding and internalization of human IgM to FCMR, its cognate cellular receptor. *Scientific Reports* 7: 42989.

52. Vire, B., A. David, and A. Wiestner. 2011. TOSO, the Fcμ Receptor, Is Highly Expressed on Chronic Lymphocytic Leukemia B Cells, Internalizes upon IgM Binding, Shuttles to the Lysosome, and Is Downregulated in Response to TLR Activation. *The Journal of Immunology* 187: 4040-4050.

53. Meryk, A., L. Pangrazzi, M. Hagen, F. Hatzmann, B. Jenewein, B. Jakic, N. Hermann-Kleiter, G. Baier, J. Jylhävä, M. Hurme, K. Trieb, and B. Grubeck-Loebenstein. 2019. Fcμ receptor as a Costimulatory Molecule for T Cells. *Cell Reports* 26: 2681-2691.e2685.

54. Vire, B., M. Skarzynski, J. D. Thomas, C. G. Nelson, A. David, G. Aue, T. R. Burke, Jr, C. Rader, and A. Wiestner. 2014. Harnessing the Fcμ Receptor for

Potent and Selective Cytotoxic Therapy of Chronic Lymphocytic Leukemia. *Cancer Research* 74: 7510-7520.

55. Faitschuk, E., A. A. Hombach, L. P. Frenzel, C.-M. Wendtner, and H. Abken. 2016. Chimeric antigen receptor T cells targeting Fc μ receptor selectively eliminate CLL cells while sparing healthy B cells. *Blood* 128: 1711-1722.

56. Woolley, C. R., N. C. Brinkman, E. D. Cash, S. K. Chandran, and T. C. Mitchell. 2022. An Unexpected Role for Cell Density Rather than IgM in Cell-Surface Display of the Fc Receptor for IgM on Human Lymphocytes. *ImmunoHorizons* 6: 47-63.

57. Kuksin, M., D. Morel, M. Aglave, F.-X. Danlos, A. Marabelle, A. Zinovyev, D. Gautheret, and L. Verlingue. 2021. Applications of single-cell and bulk RNA sequencing in onco-immunology. *European Journal of Cancer* 149: 193-210.

58. Chen, H., F. Ye, and G. Guo. 2019. Revolutionizing immunology with single-cell RNA sequencing. *Cellular & Molecular Immunology* 16: 242-249.

59. Monaco, G., B. Lee, W. Xu, S. Mustafah, Y. Y. Hwang, C. Carré, N. Burdin, L. Visan, M. Ceccarelli, M. Poidinger, A. Zippelius, J. Pedro De Magalhães, and A. Larbi. 2019. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Reports* 26: 1627-1640.e1627.

60. Bouquet, J., J. L. Gardy, S. Brown, J. Pfeil, R. R. Miller, M. Morshed, A. Avina-Zubieta, K. Shojania, M. McCabe, S. Parker, M. Uyaguari, S. Federman, P. Tang, T. Steiner, M. Otterstater, R. Holt, R. Moore, C. Y. Chiu, D. M. Patrick, and f. t. C. C. D. S. Group. 2017. RNA-Seq Analysis of Gene Expression, Viral Pathogen,

and B-Cell/T-Cell Receptor Signatures in Complex Chronic Disease. *Clinical Infectious Diseases* 64: 476-481.

61. Bannister, A. J., and T. Kouzarides. 2011. Regulation of chromatin by histone modifications. *Cell Research* 21: 381-395.

62. Schmidl, C., M. Delacher, J. Huehn, and M. Feuerer. 2018. Epigenetic mechanisms regulating T-cell responses. *Journal of Allergy and Clinical Immunology* 142: 728-743.

63. Moore, L. D., T. Le, and G. Fan. 2013. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* 38: 23-38.

64. Ogunshola, F. J., W. Smidt, A. F. Naidoo, T. Nkosi, T. Ngubane, T. Khaba, O. O. Baiyegunhi, B. Mahlobo, S. Rasehlo, N. Ngema, I. Jajbhay, K. L. Dong, V. Ramsuran, J. Pansegrouw, T. Ndung'U, B. D. Walker, T. D. Oliveria, and Z. M. Ndhlovu. 2022. Hypermethylation at the CXCR5 gene locus limits trafficking potential of CD8+ T cells into B-cell follicles during HIV-1 infection. *Blood Advances* 6: 1904-1916.

65. Liao, J., S. Luo, M. Yang, and Q. Lu. 2020. Overexpression of CXCR5 in CD4(+) T cells of SLE patients caused by excessive SETD3. *Clin Immunol* 214: 108406.

66. Coit, P., M. G. Dozmorov, J. T. Merrill, W. J. Mccune, K. Maksimowicz-Mckinnon, J. D. Wren, and A. H. Sawalha. 2016. Epigenetic Reprogramming in Naive CD4+ T Cells Favoring T Cell Activation and Non-Th1 Effector T Cell Immune Response as an Early Event in Lupus Flares. *Arthritis & Rheumatology* 68: 2200-2209.

67. Pitaksalee, R., A. N. Burska, S. Ajaib, J. Rogers, R. Parmar, K. Mydlova, X. Xie, A. Droop, J. S. Nijjar, P. Chambers, P. Emery, R. Hodgett, I. B. Mcinnes, and F. Ponchel. 2020. Differential CpG DNA methylation in peripheral naïve CD4+ T-cells in early rheumatoid arthritis patients. *Clinical Epigenetics* 12.

68. De Obaldia, M. E., and A. Bhandoola. 2015. Transcriptional Regulation of Innate and Adaptive Lymphocyte Lineages. *Annual Review of Immunology* 33: 607-642.

69. Gupta, S., H. Yan, L. H. Wong, S. Ralph, J. Krolewski, and C. Schindler. 1996. The SH2 domains of Stat1 and Stat2 mediate multiple interactions in the transduction of IFN-alpha signals. *The EMBO Journal* 15: 1075-1084.

70. Wesoly, J., Z. Szweykowska-Kulinska, and H. A. R. Bluyssen. 2007. STAT activation and differential complex formation dictate selectivity of interferon responses. *Acta Biochimica Polonica* 54: 27-38.

71. Au-Yeung, N., R. Mandhana, and C. M. Horvath. 2013. Transcriptional regulation by STAT1 and STAT2 in the interferon JAK-STAT pathway. *JAK-STAT* 2: e23931.

72. Leviyang, S. 2021. Interferon stimulated binding of ISRE is cell type specific and is predicted by homeostatic chromatin state. *Cytokine X* 3: 100056.

73. Schneider, W. M., M. D. Chevillotte, and C. M. Rice. 2014. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annual Review of Immunology* 32: 513-545.

74. Revilla-I-Domingo, R., I. Bilic, B. Vilagos, H. Tagoh, A. Ebert, I. M. Tamir, L. Smeenk, J. Trupke, A. Sommer, M. Jaritz, and M. Busslinger. 2012. The B-cell

identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *The EMBO Journal* 31: 3130-3146.

75. Xue, K., J. Song, Y. Yang, Z. Li, C. Wu, J. Jin, and W. Li. 2016. PAX5 promotes pre-B cell proliferation by regulating the expression of pre-B cell receptor and its downstream signaling. *Molecular Immunology* 73: 1-9.

76. Cobaleda, C., A. Schebesta, A. Delogu, and M. Busslinger. 2007. Pax5: the guardian of B cell identity and function. *Nature Immunology* 8: 463-470.

77. Hart, A. P., and T. M. Laufer. 2022. A review of signaling and transcriptional control in T follicular helper cell differentiation. *Journal of Leukocyte Biology* 111: 173-195.

78. Choi, J., and S. Crotty. 2021. Bcl6-Mediated Transcriptional Regulation of Follicular Helper T cells (TFH). *Trends in Immunology* 42: 336-349.

79. Crotty, S. 2019. T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. *Immunity* 50: 1132-1148.

80. Zhu, X., and J. Zhu. 2020. CD4 T Helper Cell Subsets and Related Human Immunological Disorders. *International Journal of Molecular Sciences* 21: 8011.

81. Lord, G. M., R. M. Rao, H. Choe, B. M. Sullivan, A. H. Lichtman, F. W. Luscinskas, and L. H. Glimcher. 2005. T-bet is required for optimal proinflammatory CD4+ T-cell trafficking. *Blood* 106: 3432-3439.

82. Beima, K. M., M. M. Miazgowicz, M. D. Lewis, P. S. Yan, T. H.-M. Huang, and A. S. Weinmann. 2006. T-bet Binding to Newly Identified Target Gene Promoters Is Cell Type-independent but Results in Variable Context-dependent Functional Effects. *Journal of Biological Chemistry* 281: 11992-12000.

83. Choi, J., H. Diao, C. E. Faliti, J. Truong, M. Rossi, S. Bélanger, B. Yu, A. W. Goldrath, M. E. Pipkin, and S. Crotty. 2020. Bcl-6 is the nexus transcription factor of T follicular helper cells via repressor-of-repressor circuits. *Nature Immunology* 21: 777-789.

84. Shaw, L. A., S. Bélanger, K. D. Omilusik, S. Cho, J. P. Scott-Browne, J. P. Nance, J. Goulding, A. Lasorella, L.-F. Lu, S. Crotty, and A. W. Goldrath. 2016. Id2 reinforces TH1 differentiation and inhibits E2A to repress TFH differentiation. *Nature Immunology* 17: 834-843.

85. Crotty, S. 2012. The 1-1-1 fallacy. *Immunological Reviews* 247: 133-142.

86. Oestreich, K. J., and A. S. Weinmann. 2012. Master regulators or lineage-specifying? Changing views on CD4+ T cell transcription factors. *Nature Reviews Immunology* 12: 799-804.

87. Wang, C., M. Collins, and V. K. Kuchroo. 2015. Effector T cell differentiation: are master regulators of effector T cells still the masters? *Current Opinion in Immunology* 37: 6-10.

88. Liu, X., X. Yan, B. Zhong, R. I. Nurieva, A. Wang, X. Wang, N. Martin-Orozco, Y. Wang, S. H. Chang, E. Esplugues, R. A. Flavell, Q. Tian, and C. Dong. 2012. Bcl6 expression specifies the T follicular helper cell program in vivo. *Journal of Experimental Medicine* 209: 1841-1852.

89. Liu, X., X. Chen, B. Zhong, A. Wang, X. Wang, F. Chu, R. I. Nurieva, X. Yan, P. Chen, L. G. Van Der Flier, H. Nakatsukasa, S. S. Neelapu, W. Chen, H. Clevers, Q. Tian, H. Qi, L. Wei, and C. Dong. 2014. Transcription factor achaete-scute homologue 2 initiates follicular T-helper-cell development. *Nature* 507: 513-518.

90. Schwanhäusser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. 2011. Global quantification of mammalian gene expression control. *Nature* 473: 337-342.

91. Nicolet, B. P., and M. C. Wolkers. 2022. The relationship of mRNA with protein expression in CD8+ T cells associates with gene class and gene characteristics. *PLOS ONE* 17: e0276294.

92. Nicolet, B. P., N. D. Zandhuis, V. M. Lattanzio, and M. C. Wolkers. 2021. Sequence determinants as key regulators in gene expression of T cells. *Immunological Reviews* 304: 10-29.

93. Lichti, J., C. Gallus, and E. Glasmacher. 2018. Immune Responses – Transcriptional and Post-Transcriptional Networks Pass the Baton. *Trends in Biochemical Sciences* 43: 1-4.

94. Xie, S. Q., Y. Han, X. Z. Chen, T. Y. Cao, K. K. Ji, J. Zhu, P. Ling, and C. L. Xiao. 2018. ISOdb: A Comprehensive Database of Full-Length Isoforms Generated by Iso-Seq. *Int J Genomics* 2018: 9207637.

95. Robinson, E. K., P. Jagannatha, S. Covarrubias, M. Cattle, V. Smaliy, R. Safavi, B. Shapleigh, R. Abu-Shumays, M. Jain, S. M. Cloonan, M. Akeson, A. N. Brooks, and S. Carpenter. 2021. Inflammation drives alternative first exon usage to regulate immune genes including a novel iron-regulated isoform of Aim2. *eLife* 10.

96. Kandhari, N., C. A. Kraupner-Taylor, P. F. Harrison, D. R. Powell, and T. H. Beilharz. 2021. The Detection and Bioinformatic Analysis of Alternative 3(') UTR Isoforms as Potential Cancer Biomarkers. *Int J Mol Sci* 22.

97. Grillo, G., A. Turi, F. Licciulli, F. Mignone, S. Liuni, S. Banfi, V. A. Gennarino, D. S. Horner, G. Pavesi, E. Picardi, and G. Pesole. 2010. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Research* 38: D75-D80.

98. Lo Giudice, C., F. Zambelli, M. Chiara, G. Pavesi, Marco A. Tangaro, E. Picardi, and G. Pesole. 2022. UTRdb 2.0: a comprehensive, expert curated catalog of eukaryotic mRNAs untranslated regions. *Nucleic Acids Research* 51: D337-D344.

99. Mayr, C. 2017. Regulation by 3'-Untranslated Regions. *Annu Rev Genet* 51: 171-194.

100. Salloum, D., K. Singh, N. R. Davidson, L. Cao, D. Kuo, V. R. Sanghvi, M. Jiang, M. T. Lafoz, A. Viale, G. Ratsch, and H.-G. Wendel. 2022. A Rapid Translational Immune Response Program in CD8 Memory T Lymphocytes. *The Journal of Immunology* 209: 1189-1199.

101. PATEL, P. S., S. PÉREZ-BAOS, B. WALTERS, M. ORLEN, A. VOLKOVA, K. RUGGLES, C. Y. PARK, and R. J. SCHNEIDER. 2022. Translational regulation of TFH cell differentiation and autoimmune pathogenesis. *Science Advances* 8.

102. Lu, F., H. Chen, Y. Hong, Y. Lin, L. Liu, N. Wei, Q. Wu, S. Liao, S. Yang, J. He, and Y. Shao. 2021. A gain-of-function NLRP3 3′-UTR polymorphism causes miR-146a-mediated suppression of NLRP3 expression and confers protection against sepsis progression. *Scientific Reports* 11.

103. Liu, Y., J. Wang, S. Wu, and J. Yang. 2022. A model for isoform-level differential expression analysis using RNA-seq data without pre-specifying isoform structure. *PLOS ONE* 17: e0266162.

104. Dimopoulos, A. C., K. Koukoutegos, F. E. Psomopoulos, and P. Moulos. 2021. Combining Multiple RNA-Seq Data Analysis Algorithms Using Machine Learning Improves Differential Isoform Expression Analysis. *Methods and Protocols* 4: 68.

105. Merino, G. A., and E. A. Fernandez. 2020. Differential splicing analysis based on isoforms expression with NBSplice. *J Biomed Inform* 103: 103378.

106. Kubagawa, H., S. Oka, Y. Kubagawa, I. Torii, E. Takayama, D. W. Kang, G. L. Gartland, L. F. Bertoli, H. Mori, H. Takatsu, T. Kitamura, H. Ohno, and J. Y. Wang. 2009. Identity of the elusive IgM Fc receptor (FcmuR) in humans. *J Exp Med* 206: 2779-2793.

107. Kubagawa, H., S. Oka, Y. Kubagawa, I. Torii, E. Takayama, D.-W. Kang, D. Jones, N. Nishida, T. Miyawaki, L. F. Bertoli, S. K. Sanders, and K. Honjo. 2014. The Long Elusive IgM Fc Receptor, FcμR. *Journal of Clinical Immunology* 34: 35-45.

108. Kubagawa, H., K. Honjo, N. Ohkura, S. Sakaguchi, A. Radbruch, F. Melchers, and P. K. Jani. 2019. Functional Roles of the IgM Fc Receptor in the Immune System. *Front Immunol* 10: 945.

109. Hideaki Shima, Hiroyuki Takatsu, Shinji Fukuda, Masumi Ohmae, Koji Hase, Hiromi Kubagawa, Ji-Yang Wang, and Hiroshi Ohno. 2010. Identification of TOSO/FAIM3 as an Fc receptor for IgM. *International Immunology* 22: 149-156.

110. Kubagawa, H., C. M. Skopnik, K. Al-Qaisi, R. A. Calvert, K. Honjo, Y. Kubagawa, R. Teuber, P. M. Aliabadi, P. Enghard, A. Radbruch, and B. J. Sutton.

2021. Differences between Human and Mouse IgM Fc Receptor (FcμR). *International Journal of Molecular Sciences* 22: 7024.

111. Kubli, S. P., L. Vornholz, G. Duncan, W. Zhou, P. Ramachandran, J. Fortin, M. Cox, S. Han, R. Nechanitzky, D. Nechanitzky, B. E. Snow, L. Jones, W. Y. Li, J. Haight, A. Wakeham, M. R. Bray, and T. W. Mak. 2019. Fcmr regulates mononuclear phagocyte control of anti-tumor immunity. *Nature Communications* 10.

112. Lang, P. A., A. Meryk, A. A. Pandyra, D. Brenner, A. Brüstle, H. C. Xu, K. Merches, F. Lang, V. Khairnar, P. Sharma, P. Funkner, M. Recher, N. Shaabani, G. S. Duncan, V. Duhan, B. Homey, P. S. Ohashi, D. Häussinger, P. A. Knolle, N. Honke, T. W. Mak, and K. S. Lang. 2015. Toso regulates differentiation and activation of inflammatory dendritic cells during persistence-prone virus infection. *Cell Death & Differentiation* 22: 164-173.

113. Lang, K. S., P. A. Lang, A. Meryk, A. A. Pandyra, L.-M. Boucher, V. I. Pozdeev, M. W. Tusche, J. R. Gothert, J. Haight, A. Wakeham, A. J. You-Ten, D. R. Mcilwain, K. Merches, V. Khairnar, M. Recher, G. P. Nolan, Y. Hitoshi, P. Funkner, A. A. Navarini, A. Verschoor, N. Shaabani, N. Honke, L. Z. Penn, P. S. Ohashi, D. Haussinger, K.-H. Lee, and T. W. Mak. 2013. Involvement of Toso in activation of monocytes, macrophages, and granulocytes. *Proceedings of the National Academy of Sciences* 110: 2593-2598.

114. Lang, K. S., P. A. Lang, A. Meryk, A. A. Pandyra, K. Merches, K.-H. Lee, and T. W. Mak. 2013. Reply to Honjo et al.: Functional relevant expression of Toso on

granulocytes. *Proceedings of the National Academy of Sciences* 110: E2542-E2543.

115. Brenner, D., A. Brustle, G. H. Y. Lin, P. A. Lang, G. S. Duncan, C. B. Knobbe-Thomsen, M. St. Paul, C. Reardon, M. W. Tusche, B. Snow, S. R. Hamilton, A. Pfefferle, S. O. Gilani, P. S. Ohashi, K. S. Lang, and T. W. Mak. 2014. Toso controls encephalitogenic immune responses by dendritic cells and regulatory T cells. *Proceedings of the National Academy of Sciences* 111: 1060-1065.

116. Honjo, K., Y. Kubagawa, and H. Kubagawa. 2013. Is Toso/IgM Fc receptor (Fc R) expressed by innate immune cells? *Proceedings of the National Academy of Sciences* 110: E2540-E2541.

117. Jani, P. K., H. Kubagawa, and F. Melchers. 2020. A rheostat sets B-cell receptor repertoire selection to distinguish self from non-self. *Current Opinion in Immunology* 67: 42-49.

118. Li, F. J., Y. Kubagawa, M. K. Mccollum, L. Wilson, T. Motohashi, L. F. Bertoli, J. C. Barton, S. Barnes, R. S. Davis, and H. Kubagawa. 2011. Enhanced levels of both the membrane-bound and soluble forms of IgM Fc receptor (FcμR) in patients with chronic lymphocytic leukemia. *Blood* 118: 4902-4909.

119. Gupta, S., S. Agrawal, S. Gollapudi, and H. Kubagawa. 2016. FcμR in human B cell subsets in primary selective IgM deficiency, and regulation of FcμR and production of natural IgM antibodies by IGIV. *Human Immunology* 77: 1194-1201.

120. Gonzalez-Quintela, A., R. Alende, F. Gude, J. Campos, J. Rey, L. M. Meijide, C. Fernandez-Merino, and C. Vidal. 2008. Serum levels of immunoglobulins (IgG,

IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities. *Clinical & Experimental Immunology* 151: 42-50.

121. StemCell. Using EasySep™ Positive Selection Kits for Cell Depletion. Tech Tips and Protocols.

122. Johnston, A., S. L. Sigurdardottir, and J. J. Ryon. 2009. Isolation of Mononuclear Cells from Tonsillar Tissue. *Current Protocols in Immunology* 86.

123. Carswell, K. S., J. W. Weiss, and E. T. Papoutsakis. 2000. Low oxygen tension enhances the stimulation and proliferation of human T lymphocytes in the presence of IL-2. *Cytotherapy* 2: 25-37.

124. Carreau, A., B. E. Hafny-Rahbi, A. Matejuk, C. Grillon, and C. Kieda. 2011. Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia. *Journal of Cellular and Molecular Medicine* 15: 1239-1253.

125. Andersen, M. N., S. N. H. Al-Karradi, T. W. Kragstrup, and M. Hokland. 2016. Elimination of erroneous results in flow cytometry caused by antibody binding to Fc receptors on human monocytes and macrophages. *Cytometry Part A* 89: 1001-1009.

126. Nakamura, T., H. Kubagawa, T. Ohno, and M. D. Cooper. 1993. Characterization of an IgM Fc-binding receptor on human T cells. *Journal of Immunology* 151: 6933-6941.

127. Wegner, J., S. Hackenberg, C.-J. Scholz, S. Chuvpilo, D. Tyrsin, A. A. Matskevich, G. U. Grigoleit, S. Stevanović, and T. Hünig. 2015. High-density

preculture of PBMCs restores defective sensitivity of circulating CD8 T cells to virus- and tumor-derived antigens. *Blood* 126: 185-194.

128. Römer, P. S., S. Berr, E. Avota, S.-Y. Na, M. Battaglia, I. Ten Berge, H. Einsele, and T. Hünig. 2011. Preculture of PBMCs at high cell density increases sensitivity of T-cell responses, revealing cytokine release by CD28 superagonist TGN1412. *Blood* 118: 6772-6782.

129. Ma, Q., Y. Wang, A. S. Lo, E. M. Gomes, and R. P. Junghans. 2010. Cell density plays a critical role in ex vivo expansion of T cells for adoptive immunotherapy. *J Biomed Biotechnol* 2010: 386545.

130. Stefanová, I., J. R. Dorfman, and R. N. Germain. 2002. Self-recognition promotes the foreign antigen sensitivity of naive T lymphocytes. *Nature* 420: 429-434.

131. Kubagawa, H., C. Clark, C. M. Skopnik, P. Mahmoudi Aliabadi, K. Al-Qaisi, R. Teuber, P. K. Jani, A. Radbruch, F. Melchers, N. Engels, and J. Wienands. 2023. Physiological and Pathophysiological Roles of IgM Fc Receptor (FcμR) Isoforms. *International Journal of Molecular Sciences* 24: 5728.

132. Cunningham, F., J. E. Allen, J. Allen, J. Alvarez-Jarreta, M R. Amode, Irina M. Armean, O. Austine-Orimoloye, Andrey G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, Jose G. Martinez, C. Guijarro-Clarke, A. Gymer, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. C. Marugán, S. Mohanan, A. Mushtaq, M. Naven, Denye N. Ogeh, A. Parker, A. Parton, M. Perry, I. Piližota, I.

Prosovetskaia, Manoj P. Sakthivel, Ahamed Imran A. Salam, Bianca M. Schmitt, H. Schuilenburg, D. Sheppard, José G. Pérez-Silva, W. Stark, E. Steed, K. Sutinen, R. Sukumaran, D. Sumathipala, M.-M. Suner, M. Szpak, A. Thormann, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, Thomas A. Walsh, B. Walts, N. Willhoft, A. Winterbottom, E. Wass, M. Chakiachvili, B. Flint, A. Frankish, S. Giorgetti, L. Haggerty, Sarah E. Hunt, Garth R. IIsley, Jane E. Loveland, Fergal J. Martin, B. Moore, Jonathan M. Mudge, M. Muffato, E. Perry, M. Ruffier, J. Tate, D. Thybert, Stephen J. Trevanion, S. Dyer, Peter W. Harrison, Kevin L. Howe, Andrew D. Yates, Daniel R. Zerbino, and P. Flicek. 2021. Ensembl 2022. *Nucleic Acids Research* 50: D988-D995.

133. Frankish, A., M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O. G. Izuogu, J. Lagarde, F. J. Martin, L. Martínez, S. Mohanan, P. Muir, F. C P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B. M. Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J. S. Choudhary, M. Gerstein, R. Guigó, T. J P. Hubbard, M. Kellis, B. Paten, A. Reymond, M. L. Tress, and P. Flicek. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research* 47: D766-D773.

134. Leung, S. K., A. R. Jeffries, I. Castanho, B. T. Jordan, K. Moore, J. P. Davies, E. L. Dempster, N. J. Bray, P. O'Neill, E. Tseng, Z. Ahmed, D. A. Collier, E. D.

Jeffery, S. Prabhakar, L. Schalkwyk, C. Jops, M. J. Gandal, G. M. Sheynkman, E. Hannon, and J. Mill. 2021. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Reports* 37: 110022.

135. Corney, D., and G. Basturea. 2016. RNA-seq Using Next Generation Sequencing A comprehensive review of RNA-seq methodologies. *MATER METHODS 2013* 3.

136. Sharon, D., H. Tilgner, F. Grubert, and M. Snyder. 2013. A single-molecule long-read survey of the human transcriptome. *Nature Biotechnology* 31: 1009-1014.

137. Sims, D., I. Sudbery, N. E. Ilott, A. Heger, and C. P. Ponting. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics* 15: 121-132.

138. Wang, C., L. Chen, Y. Chen, W. Jia, X. Cai, Y. Liu, F. Ji, P. Xiong, A. Liang, R. Liu, Y. Guan, Z. Cheng, Y. Weng, W. Wang, Y. Duan, D. Kuang, S. Xu, H. Cai, Q. Xia, D. Yang, M.-W. Wang, X. Yang, J. Zhang, C. Cheng, L. Liu, Z. Liu, R. Liang, G. Wang, Z. Li, H. Xia, and T. Xia. 2022. Abnormal global alternative RNA splicing in COVID-19 patients. *PLOS Genetics* 18: e1010137.

139. Kim, E.-Y., Y. Che, H. J. Dean, R. Lorenzo-Redondo, M. Stewart, C. K. Keller, D. Whorf, D. Mills, N. N. Dulin, T. Kim, M. Votoupal, M. Walter, A. Fernandez-Sesma, H. Kim, and S. M. Wolinsky. 2022. Transcriptome-wide changes in gene expression, splicing, and lncRNAs in response to a live attenuated dengue virus vaccine. *Cell Reports* 38: 110341.

140. Chen, Y., G. Wang, J. Li, L. Xia, L. Zhu, W. Li, Q. Luo, Y. Liao, Y. Lin, L. Bi, H. Chen, J. Chu, Y. Li, J. Su, L. Ye, J.-J. Jiang, H. Liang, W. Li, and S. An. 2022. CASA: a comprehensive database resource for the COVID-19 Alternative Splicing Atlas. *Journal of Translational Medicine* 20.

141. Marrone, L., M. D'Agostino, E. Cesaro, V. Di Giacomo, S. Urzini, M. F. Romano, and S. Romano. 2023. Alternative splicing of FKBP5 gene exerts control over T lymphocyte expansion. *Journal of Cellular Biochemistry*.

142. Lee, F. F. Y., and S. Alper. 2022. Alternative pre-mRNA splicing as a mechanism for terminating Toll-like Receptor signaling. *Front Immunol* 13: 1023567.

143. Su, Z., and D. Huang. 2021. Alternative Splicing of Pre-mRNA in the Control of Immune Activity. *Genes* 12: 574.

144. Blake, D., C. M. Radens, M. B. Ferretti, M. R. Gazzara, and K. W. Lynch. 2022. Alternative splicing of apoptosis genes promotes human T cell survival. *eLife* 11.

145. Bernard, A., R. Boidot, and F. Végran. 2022. Alternative Splicing in Cancer and Immune Cells. *Cancers* 14: 1726.

146. Sahoo, A., and S. H. Im. 2010. Interleukin and interleukin receptor diversity: role of alternative splicing. *Int Rev Immunol* 29: 77-109.

147. Carey, B. S., K. V. Poulton, and A. Poles. 2019. Factors affecting HLA expression: A review. *International Journal of Immunogenetics* 46: 307-320.

148. Weber, R., U. Ghoshdastider, D. Spies, C. Duré, F. Valdivia-Francia, M. Forny, M. Ormiston, P. F. Renz, D. Taborsky, M. Yigit, M. Bernasconi, H. Yamahachi, and A. Sendoel. 2023. Monitoring the 5′UTR landscape reveals isoform switches to drive translational efficiencies in cancer. *Oncogene* 42: 638-650.

149. Surani, A. A., K. A. Spriggs, C. Ufer, C. Polytarchou, and C. Montiel-Duarte. 2022. Implications of differential transcription start site selection on chronic myeloid leukemia and prostate cancer cell protein expression. *iScience* 25: 105519.

150. Oliveira, E. H., A. F. Assis, C. A. Speck-Hernandez, M. J. Duarte, and G. A. Passos. 2020. Aire Gene Influences the Length of the 3' UTR of mRNAs in Medullary Thymic Epithelial Cells. *Front Immunol* 11: 1039.

151. Mayr, C. 2019. What Are 3′ UTRs Doing? *Cold Spring Harbor Perspectives in Biology* 11: a034728.

152. Chen, C.-Y. A., and A.-B. Shyu. 1995. AU-rich elements: characterization and importance in mRNA degradation. *Trends in Biochemical Sciences* 20: 465-470.

153. Vlasova-St. Louis, I., and P. R. Bohjanen. 2014. Post-Transcriptional Regulation of Cytokine Signaling by AU-Rich and GU-Rich Elements. *Journal of Interferon & Cytokine Research* 34: 233-241.

154. Mayya, V. K., and T. F. Duchaine. 2019. Ciphers and Executioners: How 3'-Untranslated Regions Determine the Fate of Messenger RNAs. *Front Genet* 10: 6.

155. Rothamel, K., S. Arcos, B. Kim, C. Reasoner, S. Lisy, N. Mukherjee, and M. Ascano. 2021. ELAVL1 primarily couples mRNA stability with the 3’ UTRs of interferon-stimulated genes. *Cell Reports* 35.

156. Díaz-Muñoz, M. D., and M. Turner. 2018. Uncovering the Role of RNA-Binding Proteins in Gene Expression in the Immune System. *Front Immunol* 9: 1094.

157. Sun, L., G. Stoecklin, S. Van Way, V. Hinkovska-Galcheva, R.-F. Guo, P. Anderson, and T. P. Shanley. 2007. Tristetraprolin (TTP)-14-3-3 Complex

Formation Protects TTP from Dephosphorylation by Protein Phosphatase 2a and Stabilizes Tumor Necrosis Factor-α mRNA. *Journal of Biological Chemistry* 282: 3766-3777.

158. Ronkina, N., M. B. Menon, J. Schwermann, C. Tiedje, E. Hitti, A. Kotlyarov, and M. Gaestel. 2010. MAPKAP kinases MK2 and MK3 in inflammation: complex regulation of TNF biosynthesis via expression and phosphorylation of tristetraprolin. *Biochem Pharmacol* 80: 1915-1920.

159. Tiedje, C., M. D. Diaz-Muñoz, P. Trulley, H. Ahlfors, K. Laaß, P. J. Blackshear, M. Turner, and M. Gaestel. 2016. The RNA-binding protein TTP is a global post-transcriptional regulator of feedback control in inflammation. *Nucleic Acids Res* 44: 7418-7440.

160. Ye, J., M. Xu, X. Tian, S. Cai, and S. Zeng. 2019. Research advances in the detection of miRNA. *Journal of Pharmaceutical Analysis* 9: 217-226.

161. Iwakawa, H.-o., and Y. Tomari. 2022. Life of RISC: Formation, action, and degradation of RNA-induced silencing complex. *Molecular Cell* 82: 30-43.

162. Wang, J., B.-S. Shin, C. Alvarado, J.-R. Kim, J. Bohlen, T. E. Dever, and J. D. Puglisi. 2022. Rapid 40S scanning and its regulation by mRNA structure during eukaryotic translation initiation. *Cell* 185: 4474-4487.e4417.

163. Leppek, K., R. Das, and M. Barna. 2018. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology* 19: 158-174.

164. Hentze, M. W., S. W. Caughman, T. A. Rouault, J. G. Barriocanal, A. Dancis, J. B. Harford, and R. D. Klausner. 1987. Identification of the Iron-Responsive

Element for the Translational Regulation of Human Ferritin mRNA. *Science* 238: 1570-1573.

165. Gray, N. K., and M. W. Hentze. 1994. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *The EMBO Journal* 13: 3882-3891.

166. Galbán, S., Y. Kuwano, R. Pullmann, J. L. Martindale, H. H. Kim, A. Lal, K. Abdelmohsen, X. Yang, Y. Dang, J. O. Liu, S. M. Lewis, M. Holcik, and M. Gorospe. 2008. RNA-Binding Proteins HuR and PTB Promote the Translation of Hypoxia-Inducible Factor 1α. *Molecular and Cellular Biology* 28: 93-107.

167. Babendure, J. R., J. L. Babendure, J.-H. Ding, and R. Y. Tsien. 2006. Control of mammalian translation by mRNA structure near caps. *RNA* 12: 851-861.

168. Ben-Asouli, Y., Y. Banai, Y. Pel-Or, A. Shir, and R. Kaempfer. 2002. Human Interferon-γ mRNA Autoregulates Its Translation through a Pseudoknot that Activates the Interferon-Inducible Protein Kinase PKR. *Cell* 108: 221-232.

169. Liu, Q., X. Peng, M. Shen, Q. Qian, J. Xing, C. Li, and Richard I. Gregory. 2022. Ribo-uORF: a comprehensive data resource of upstream open reading frames (uORFs) based on ribosome profiling. *Nucleic Acids Research* 51: D248-D261.

170. Zhang, H., Y. Wang, X. Wu, X. Tang, C. Wu, and J. Lu. 2021. Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. *Nature Communications* 12.

171. Moro, S. G., C. Hermans, J. Ruiz-Orera, and M. M. Albà. 2021. Impact of uORFs in mediating regulation of translation in stress conditions. *BMC Molecular and Cell Biology* 22.

172. Bottorff, T. A., H. Park, A. P. Geballe, and A. R. Subramaniam. 2022. Translational buffering by ribosome stalling in upstream open reading frames. *PLOS Genetics* 18: e1010460.

173. Calvo, S. E., D. J. Pagliarini, and V. K. Mootha. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* 106: 7507-7512.

174. Wang, X. Q., and J. A. Rothnagel. 2004. 5'-untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. *Nucleic Acids Res* 32: 1382-1391.

175. Russell, P. J., J. A. Slivka, E. P. Boyle, A. H. M. Burghes, and M. G. Kearse. 2023. Translation re-initiation after uORFs does not fully protect mRNAs from nonsense-mediated decay. *RNA*: rna.079525.079122.

176. Chen, H.-H., and W.-Y. Tarn. 2019. uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. *RNA Biology* 16: 1327-1338.

177. Jayaram, D. R., S. Frost, C. Argov, V. B. Liju, N. P. Anto, A. Muraleedharan, A. Ben-Ari, R. Sinay, I. Smoly, O. Novoplansky, N. Isakov, D. Toiber, C. Keasar, M. Elkabets, E. Yeger-Lotem, and E. Livneh. 2021. Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs. *Proceedings of the National Academy of Sciences* 118: e2018899118.

178. Yabas, M., H. Elliott, and G. Hoyne. 2015. The Role of Alternative Splicing in the Control of Immune Homeostasis and Cellular Differentiation. *International Journal of Molecular Sciences* 17: 3.

179. Zhang, Z., N. Sauerwald, A. Cappuccio, I. Ramos, V. D. Nair, G. Nudelman, E. Zaslavsky, Y. Ge, A. Gaitas, H. Ren, J. Brockman, J. Geis, N. Ramalingam, D. King, M. T. McClain, C. W. Woods, R. Henao, T. W. Burke, E. L. Tsalik, C. W. Goforth, R. A. Lizewski, S. E. Lizewski, D. L. Weir, A. G. Letizia, S. C. Sealfon, and O. G. Troyanskaya. 2023. Blood RNA alternative splicing events as diagnostic biomarkers for infectious disease. *Cell Reports Methods* 3: 100395.

180. Hitomi, Y., Y. Aiba, K. Ueno, N. Nishida, Y. Kawai, M. Kawashima, M. Tsuiji, C. Iwabuchi, S. Takada, N. Miyake, M. Nagasaki, K. Tokunaga, and M. Nakamura. 2022. rs2013278 in the multiple immunological-trait susceptibility locus CD28 regulates the production of non-functional splicing isoforms. *Human Genomics* 16.

181. Peng, Q., Y. Zhou, L. Oyang, N. Wu, Y. Tang, M. Su, X. Luo, Y. Wang, X. Sheng, J. Ma, and Q. Liao. 2021. Impacts and mechanisms of alternative mRNA splicing in cancer metabolism, immune response, and therapeutics. *Molecular Therapy*.

182. Gruber, A. R., G. Martin, P. Müller, A. Schmidt, A. J. Gruber, R. Gumienny, N. Mittal, R. Jayachandran, J. Pieters, W. Keller, E. Van Nimwegen, and M. Zavolan. 2014. Global 3′ UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nature Communications* 5: 5465.

183. Chatterjee, S., and J. K. Pal. 2009. Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. *Biology of the Cell* 101: 251-262.

184. Amin, P. H., K. R. Carlson, and R. C. Wek. 2023. An RNA stem-loop functions in conjunction with an upstream open reading frame to direct preferential

translation in the integrated stress response. *Journal of Biological Chemistry* 299: 102864.

185. Pan, L., H. Q. Dinh, Y. Pawitan, and T. N. Vu. 2021. Isoform-level quantification for single-cell RNA sequencing. *Bioinformatics* 38: 1287-1294.

186. Volden, R., and C. Vollmers. 2022. Single-cell isoform analysis in human immune cells. *Genome Biology* 23.

187. Lähnemann, D., J. Köster, E. Szczurek, D. J. Mccarthy, S. C. Hicks, M. D. Robinson, C. A. Vallejos, K. R. Campbell, N. Beerenwinkel, A. Mahfouz, L. Pinello, P. Skums, A. Stamatakis, C. S.-O. Attolini, S. Aparicio, J. Baaijens, M. Balvert, B. D. Barbanson, A. Cappuccio, G. Corleone, B. E. Dutilh, M. Florescu, V. Guryev, R. Holmer, K. Jahn, T. J. Lobo, E. M. Keizer, I. Khatri, S. M. Kielbasa, J. O. Korbel, A. M. Kozlov, T.-H. Kuo, B. P. F. Lelieveldt, I. I. Mandoiu, J. C. Marioni, T. Marschall, F. Mölder, A. Niknejad, L. Raczkowski, M. Reinders, J. D. Ridder, A.-E. Saliba, A. Somarakis, O. Stegle, F. J. Theis, H. Yang, A. Zelikovsky, A. C. Mchardy, B. J. Raphael, S. P. Shah, and A. Schönhuth. 2020. Eleven grand challenges in single-cell data science. *Genome Biology* 21.

188. Gupta, I., P. G. Collier, B. Haase, A. Mahfouz, A. Joglekar, T. Floyd, F. Koopmans, B. Barres, A. B. Smit, S. A. Sloan, W. Luo, O. Fedrigo, M. E. Ross, and H. U. Tilgner. 2018. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nature Biotechnology* 36: 1197-1202.

189. Al'Khafaji, A. M., J. T. Smith, K. V. Garimella, M. Babadi, M. Sade-Feldman, M. Gatzen, S. Sarkizova, M. A. Schwartz, V. Popic, E. M. Blaum, A. Day, M.

Costello, T. Bowers, S. Gabriel, E. Banks, A. A. Philippakis, G. M. Boland, P. C. Blainey, and N. Hacohen. 2021. High-throughput RNA isoform sequencing using programmable cDNA concatenation. Cold Spring Harbor Laboratory.

190. Magrini, V., X. Gao, B. A. Rosa, S. Mcgrath, X. Zhang, K. Hallsworth-Pepin, J. Martin, J. Hawdon, R. K. Wilson, and M. Mitreva. 2018. Improving eukaryotic genome annotation using single molecule mRNA sequencing. *BMC Genomics* 19.

191. Deng, J., M. Gu, Y. Miao, S. Yao, M. Zhu, P. Fang, X. Yu, P. Li, Y. Su, J. Huang, J. Zhang, J. Yu, F. Li, J. Bai, W. Sun, Y. Huang, Y. Yuan, D. Hong, and Z. Wang. 2019. Long-read sequencing identified repeat expansions in the 5′UTR of the NOTCH2NLC gene from Chinese patients with neuronal intranuclear inclusion disease. *Journal of Medical Genetics* 56: 758-764.

192. Tardaguila, M., L. De La Fuente, C. Marti, C. Pereira, F. J. Pardo-Palacios, H. Del Risco, M. Ferrell, M. Mellado, M. Macchietto, K. Verheggen, M. Edelmann, I. Ezkurdia, J. Vazquez, M. Tress, A. Mortazavi, L. Martens, S. Rodriguez-Navarro, V. Moreno-Manzano, and A. Conesa. 2018. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Research* 28: 396-411.

193. Frankish, A., M. Diekhans, I. Jungreis, J. Lagarde, Jane E. Loveland, J. M. Mudge, C. Sisu, J. C. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, C. Boix, S. Carbonell Sala, F. Cunningham, T. Di Domenico, S. Donaldson, Ian T. Fiddes, C. García Girón, J. M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, K. L. Howe, T. Hunt, O. G. Izuogu, R. Johnson, F. J. Martin, L. Martínez, S. Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, F. C. Riera, M. Ruffier, B.

M. Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, M. Y. Wolf, J. Xu, Yucheng T. Yang, A. Yates, D. Zerbino, Y. Zhang, Jyoti S. Choudhary, M. Gerstein, R. Guigó, T. J. P. Hubbard, M. Kellis, B. Paten, M. L. Tress, and P. Flicek. 2020. GENCODE 2021. *Nucleic Acids Research* 49: D916-D923.

194. Au, K. F., V. Sebastiano, P. T. Afshar, J. D. Durruthy, L. Lee, B. A. Williams, H. Van Bakel, E. E. Schadt, R. A. Reijo-Pera, J. G. Underwood, and W. H. Wong. 2013. Characterization of the human ESC transcriptome by hybrid sequencing. *Proceedings of the National Academy of Sciences* 110: E4821-E4830.

195. Kuo, R. I., Y. Cheng, R. Zhang, J. W. S. Brown, J. Smith, A. L. Archibald, and D. W. Burt. 2020. Illuminating the dark side of the human transcriptome with long read transcript sequencing. *BMC Genomics* 21.

196. Edgar, R., M. Domrachev, and A. Lash. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* 30: 207-210.

197. Brett, D., H. Pospisil, J. Valcárcel, J. Reich, and P. Bork. 2002. Alternative splicing and genome complexity. *Nature Genetics* 30: 29-30.

198. Cock, P. J. A., C. J. Fields, N. Goto, M. L. Heuer, and P. M. Rice. 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research* 38: 1767-1771.

199. Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094-3100.

200. Andrews, S. 2015. FastQC: A Quality Control Tool for High Throughput Sequence Data., http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

201. Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21.

202. Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and A. D. Haussler. 2002. The Human Genome Browser at UCSC. *Genome Research* 12: 996-1006.

203. Kolberg, L., U. Raudvere, I. Kuzmin, J. Vilo, and H. Peterson. 2020. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* 9: 709.

204. Ip, J. Y., A. Tong, Q. Pan, J. D. Topp, B. J. Blencowe, and K. W. Lynch. 2007. Global analysis of alternative splicing during T-cell activation. *RNA* 13: 563-572.

205. Nguyen Quang, N., S. Goudey, E. Ségéral, A. Mohammad, S. Lemoine, C. Blugeon, M. Versapuech, J.-C. Paillart, C. Berlioz-Torrent, S. Emiliani, and S. Gallois-Montbrun. 2020. Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection. *Retrovirology* 17.

206. Singh, M., G. Al-Eryani, S. Carswell, J. M. Ferguson, J. Blackburn, K. Barton, D. Roden, F. Luciani, T. Giang Phan, S. Junankar, K. Jackson, C. C. Goodnow, M. A. Smith, and A. Swarbrick. 2019. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nature Communications* 10.

207. Chan, J. J., B. Zhang, X. H. Chew, A. Salhi, Z. H. Kwok, C. Y. Lim, N. Desi, N. Subramaniam, A. Siemens, T. Kinanti, S. Ong, A. Sanchez-Mejias, P. T. Ly, O. An, R. Sundar, X. Fan, S. Wang, B. E. Siew, K. C. Lee, C. S. Chong, B. Lieske, W.-K. Cheong, Y. Goh, W. N. Fam, M. G. Ooi, B. T. H. Koh, S. G. Iyer, W. H. Ling, J. Chen, B.-K. Yoong, R. Chanwat, G. K. Bonney, B. K. P. Goh, W. Zhai, M. J. Fullwood, W. Wang, K.-K. Tan, W. J. Chng, Y. Y. Dan, J. J. Pitt, X. Roca, E. Guccione, L. A. Vardy, L. Chen, X. Gao, P. K. H. Chow, H. Yang, and Y. Tay. 2022. Pan-cancer pervasive upregulation of 3′ UTR splicing drives tumourigenesis. *Nature Cell Biology* 24: 928-939.

208. Hong, D., and S. Jeong. 2023. 3'UTR Diversity: Expanding Repertoire of RNA Alterations in Human mRNAs. *Molecules and Cells* 46: 48-56.

209. Morales, J., S. Pujar, J. E. Loveland, A. Astashyn, R. Bennett, A. Berry, E. Cox, C. Davidson, O. Ermolaeva, C. M. Farrell, R. Fatima, L. Gil, T. Goldfarb, J. M. Gonzalez, D. Haddad, M. Hardy, T. Hunt, J. Jackson, V. S. Joardar, M. Kay, V. K. Kodali, K. M. Mcgarvey, A. Mcmahon, J. M. Mudge, D. N. Murphy, M. R. Murphy, B. Rajput, S. H. Rangwala, L. D. Riddick, F. Thibaud-Nissen, G. Threadgold, A. R. Vatsan, C. Wallin, D. Webb, P. Flicek, E. Birney, K. D. Pruitt, A. Frankish, F. Cunningham, and T. D. Murphy. 2022. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* 604: 310-315.

210. Crotty, S. 2014. T Follicular Helper Cell Differentiation, Function, and Roles in Disease. *Immunity* 41: 529-542.

211. Locci, M., J. E. Wu, F. Arumemi, Z. Mikulski, C. Dahlberg, A. T. Miller, and S. Crotty. 2016. Activin A programs the differentiation of human TFH cells. *Nat Immunol* 17: 976-984.

212. Dreos, R., G. Ambrosini, R. Groux, R. Cavin Périer, and P. Bucher. 2016. The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Research* 45: D51-D55.

213. Lugowski, A., B. Nicholson, and O. S. Rissland. 2018. Determining mRNA half-lives on a transcriptome-wide scale. *Methods* 137: 90-98.

214. Fuge, G., A.-P. Zeng, and U. Jandt. 2017. Weak cell cycle dependency but strong distortive effects of transfection with Lipofectamine 2000 in near-physiologically synchronized cell culture. *Engineering in Life Sciences* 17: 348-356.

215. McLenachan, S., J. P. Sarsero, and P. A. Ioannou. 2007. Flow-cytometric analysis of mouse embryonic stem cell lipofection using small and large DNA constructs. *Genomics* 89: 708-720.

216. Ratnadiwakara, M., and M.-L. Änkö. 2018. mRNA Stability Assay Using Transcription Inhibition by Actinomycin D in Mouse Pluripotent Stem Cells. *BIO-PROTOCOL* 8.

217. Woolley, C. R., J. H. Chariker, E. C. Rouchka, E. E. Ford, E. A. Hudson, S. J. Waigel, M. L. Smith, and T. C. Mitchell. 2022. Reference long-read isoform-aware transcriptomes of 4 human peripheral blood lymphocyte subsets. *G3 Genes|Genomes|Genetics* 12.

218. Putri, G. H., S. Anders, P. T. Pyl, J. E. Pimanda, and F. Zanini. 2022. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* 38: 2943-2945.

219. Giles, J. R., S. Manne, E. Freilich, D. A. Oldridge, A. E. Baxter, S. George, Z. Chen, H. Huang, L. Chilukuri, M. Carberry, L. Giles, N. P. Weng, R. M. Young, C. H. June, L. M. Schuchter, R. K. Amaravadi, X. Xu, G. C. Karakousis, T. C. Mitchell, A. C. Huang, J. Shi, and E. J. Wherry. 2022. Human epigenetic and transcriptional T cell differentiation atlas for identifying functional T cell-specific enhancers. *Immunity* 55: 557-574 e557.

220. Weinstein, J. S., K. Lezon-Geyda, Y. Maksimova, S. Craft, Y. Zhang, M. Su, V. P. Schulz, J. Craft, and P. G. Gallagher. 2014. Global transcriptome analysis and enhancer landscape of human primary T follicular helper and T effector lymphocytes. *Blood* 124: 3719-3729.

221. Bauer, L., L. J. Müller, S. M. Volkers, F. Heinrich, M. F. Mashreghi, C. Ruppert, L. E. Sander, and A. Hutloff. 2021. Follicular Helper-like T Cells in the Lung Highlight a Novel Role of B Cells in Sarcoidosis. *Am J Respir Crit Care Med* 204: 1403-1417.

222. Hinrichs, A. C., S. L. M. Blokland, A. P. Lopes, C. G. K. Wichers, A. A. Kruize, A. Pandit, T. Radstake, and J. A. G. van Roon. 2021. Transcriptome Analysis of CCR9+ T Helper Cells From Primary Sjögren's Syndrome Patients Identifies CCL5 as a Novel Effector Molecule. *Front Immunol* 12: 702733.

223. Herati, R. S., L. V. Silva, L. A. Vella, A. Muselman, C. Alanio, B. Bengsch, R. K. Kurupati, S. Kannan, S. Manne, A. V. Kossenkov, D. H. Canaday, S. A. Doyle,

H. C. J. Ertl, K. E. Schmader, and E. J. Wherry. 2021. Vaccine-induced ICOS(+)CD38(+) circulating Tfh are sensitive biosensors of age-related changes in inflammatory pathways. *Cell Rep Med* 2: 100262.

224. Herati, R. S., D. A. Knorr, L. A. Vella, L. V. Silva, L. Chilukuri, S. A. Apostolidis, A. C. Huang, A. Muselman, S. Manne, O. Kuthuru, R. P. Staupe, S. A. Adamski, S. Kannan, R. K. Kurupati, H. C. J. Ertl, J. L. Wong, S. Bournazos, S. Mcgettigan, L. M. Schuchter, R. R. Kotecha, S. A. Funt, M. H. Voss, R. J. Motzer, C.-H. Lee, D. F. Bajorin, T. C. Mitchell, J. V. Ravetch, and E. J. Wherry. 2022. PD-1 directed immunotherapy alters Tfh and humoral immune responses to seasonal influenza vaccine. *Nature Immunology* 23: 1183-1192.

225. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74.

226. Calderon, D., M. L. T. Nguyen, A. Mezger, A. Kathiria, F. Müller, V. Nguyen, N. Lescano, B. Wu, J. Trombetta, J. V. Ribado, D. A. Knowles, Z. Gao, F. Blaeschke, A. V. Parent, T. D. Burt, M. S. Anderson, L. A. Criswell, W. J. Greenleaf, A. Marson, and J. K. Pritchard. 2019. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nature Genetics* 51: 1494-1505.

227. Martino, D., M. Neeland, T. Dang, J. Cobb, J. Ellis, A. Barnett, M. Tang, P. Vuillermin, K. Allen, and R. Saffery. 2018. Epigenetic dysregulation of naive CD4+ T-cell activation genes in childhood food allergy. *Nature Communications* 9.

228. Hill, D. L., W. Pierson, D. J. Bolland, C. Mkindi, E. J. Carr, J. Wang, S. Houard, S. W. Wingett, R. Audran, E. F. Wallin, S. A. Jongo, K. Kamaka, M. Zand, F. Spertini, C. Daubenberger, A. E. Corcoran, and M. A. Linterman. 2019. The adjuvant GLA-SE promotes human Tfh cell expansion and emergence of public TCRβ clonotypes. *Journal of Experimental Medicine* 216: 1857-1873.

229. Gao, X., K. Luo, D. Wang, Y. Wei, Y. Yao, J. Deng, Y. Yang, Q. Zeng, X. Dong, L. Xiong, D. Gong, L. Lin, K. Pohl, S. Liu, Y. Liu, L. Liu, T. H. Nguyen, L. F. Allen, K. Kedzierska, Y. Jin, M.-R. Du, W. Chen, L. Lu, N. Shen, Z. Liu, I. A. Cockburn, W. Luo, and D. Yu. 2023. T follicular helper 17 (Tfh17) cells are superior for immunological memory maintenance. *eLife* 12.

230. Cook, L., M. Stahl, X. Han, A. Nazli, K. N. Macdonald, M. Q. Wong, K. Tsai, S. Dizzell, K. Jacobson, B. Bressler, C. Kaushic, B. A. Vallance, T. S. Steiner, and M. K. Levings. 2019. Suppressive and Gut-Reparative Functions of Human Type 1 T Regulatory Cells. *Gastroenterology* 157: 1584-1598.

231. Schmidt, A., F. Marabita, N. A. Kiani, C. C. Gross, H. J. Johansson, S. Éliás, S. Rautio, M. Eriksson, S. J. Fernandes, G. Silberberg, U. Ullah, U. Bhatia, H. Lähdesmäki, J. Lehtiö, D. Gomez-Cabrero, H. Wiendl, R. Lahesmaa, and J. Tegnér. 2018. Time-resolved transcriptome and proteome landscape of human regulatory T cell (Treg) differentiation reveals novel regulators of FOXP3. *BMC Biology* 16.

232. Hertweck, A., M. Evans, Catherine, M. Eskandarpour, C. H. Lau, Jonathan, K. Oleinika, I. Jackson, A. Kelly, J. Ambrose, P. Adamson, J. Cousins, David, P. Lavender, L. Calder, Virginia, M. Lord, Graham, and G. Jenner, Richard. 2016.

T-bet Activates Th1 Genes through Mediator and the Super Elongation Complex. *Cell Reports* 15: 2756-2770.

233. Bediaga, N. G., H. D. Coughlan, T. M. Johanson, A. L. Garnham, G. Naselli, J. Schröder, L. G. Fearnley, E. Bandala-Sanchez, R. S. Allan, G. K. Smyth, and L. C. Harrison. 2021. Multi-level remodelling of chromatin underlying activation of human T cells. *Scientific Reports* 11.

234. Vella, L. A., M. Buggert, S. Manne, R. S. Herati, I. Sayin, L. Kuri-Cervantes, I. Bukh Brody, K. C. O'Boyle, H. Kaprielian, J. R. Giles, S. Nguyen, A. Muselman, J. P. Antel, A. Bar-Or, M. E. Johnson, D. H. Canaday, A. Naji, V. V. Ganusov, T. M. Laufer, A. D. Wells, Y. Dori, M. G. Itkin, M. R. Betts, and E. J. Wherry. 2019. T follicular helper cells in human efferent lymph retain lymphoid characteristics. *Journal of Clinical Investigation* 129: 3185-3200.

235. Leinonen, R., H. Sugawara, and M. Shumway. 2011. The sequence read archive. *Nucleic Acids Res* 39: D19-21.

236. Goujon, M., H. Mcwilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, and R. Lopez. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research* 38: W695-W699.

237. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. Mcwilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7: 539.

238. McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, and R. Lopez. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research* 41: W597-W600.

239. Sayers, E. W., E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, and S. T. Sherry. 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50: D20-d26.

240. Wheeler, D. L. 2003. Database resources of the National Center for Biotechnology. *Nucleic Acids Research* 31: 28-33.

241. Uribe-Querol, E., and C. Rosales. 2020. Phagocytosis: Our Current Understanding of a Universal Biological Process. *Frontiers in Immunology* 11.

242. Pincetic, A., S. Bournazos, D. J. Dilillo, J. Maamary, T. T. Wang, R. Dahan, B.-M. Fiebiger, and J. V. Ravetch. 2014. Type I and type II Fc receptors regulate innate and adaptive immunity. *Nature Immunology* 15: 707-716.

243. Mazgaeen, L., and P. Gurung. 2020. Recent Advances in Lipopolysaccharide Recognition Systems. *International Journal of Molecular Sciences* 21: 379.

244. Kanneganti, T. D. 2020. Intracellular innate immune receptors: Life inside the cell. *Immunological Reviews* 297: 5-12.

245. Li, D., and M. Wu. 2021. Pattern recognition receptors in health and diseases. *Signal Transduction and Targeted Therapy* 6.

246. Stögerer, T., and S. Stäger. 2020. Innate Immune Sensing by Cells of the Adaptive Immune System. *Frontiers in Immunology* 11.

247. Ten Hacken, E., R. Guièze, and C. J. Wu. 2017. SnapShot: Chronic Lymphocytic Leukemia. *Cancer Cell* 32: 716-716.e711.

248. Blake, D., and K. W. Lynch. 2021. The three as: Alternative splicing, alternative polyadenylation and their impact on apoptosis in immune function. *Immunological Reviews* 304: 30-50.

249. Andersson, R. 2015. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays* 37: 314-323.

250. Tippens, N. D., A. Vihervaara, and J. T. Lis. 2018. Enhancer transcription: what, where, when, and why? *Genes & Development* 32: 1-3.

251. Schmidl, C., L. Hansmann, T. Lassmann, P. J. Balwierz, H. Kawaji, M. Itoh, J. Kawai, S. Nagao-Sato, H. Suzuki, R. Andreesen, Y. Hayashizaki, A. R. R. Forrest, P. Carninci, P. Hoffmann, M. Edinger, and M. Rehli. 2014. The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations. *Blood* 123: e68-e78.

252. Morganti, S., P. Tarantino, E. Ferraro, P. D'Amico, G. Viale, D. Trapani, B. A. Duso, and G. Curigliano. 2019. Complexity of genome sequencing and reporting: Next generation sequencing (NGS) technologies and implementation of precision medicine in real life. *Critical Reviews in Oncology/Hematology* 133: 171-182.

253. Stempsey, W. E. 2006. The Geneticization of Diagnostics. *Medicine, Health Care and Philosophy* 9: 193-200.

254. Veiga, D. F. T., A. Nesta, Y. Zhao, A. Deslattes Mays, R. Huynh, R. Rossi, T. C. Wu, K. Palucka, O. Anczukow, C. R. Beck, and J. Banchereau. 2022. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. *Sci Adv* 8: eabg6711.

255. Lim, Y., S. Arora, S. L. Schuster, L. Corey, M. Fitzgibbon, C. L. Wladyka, X. Wu, I. M. Coleman, J. J. Delrow, E. Corey, L. D. True, P. S. Nelson, G. Ha, and A. C. Hsieh. 2021. Multiplexed functional genomic analysis of 5' untranslated region mutations across the spectrum of prostate cancer. *Nature Communications* 12.

256. Huang, K. K., J. Huang, J. K. L. Wu, M. Lee, S. T. Tay, V. Kumar, K. Ramnarayanan, N. Padmanabhan, C. Xu, A. L. K. Tan, C. Chan, D. Kappei, J. Göke, and P. Tan. 2021. Long-read transcriptome sequencing reveals abundant promoter diversity in distinct molecular subtypes of gastric cancer. *Genome Biology* 22.

257. Freen-Van Heeren, J. J. 2021. Post-transcriptional control of T-cell cytokine production: Implications for cancer therapy. *Immunology* 164: 57-72.

258. Boyle, E. A., Y. I. Li, and J. K. Pritchard. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169: 1177-1186.

259. Chen, J., and W. Tian. 2016. Explaining the disease phenotype of intergenic SNP through predicted long range regulation. *Nucleic Acids Research* 44: 8641-8654.

260. Bartonicek, N., M. B. Clark, X. C. Quek, J. R. Torpy, A. L. Pritchard, J. L. V. Maag, B. S. Gloss, J. Crawford, R. J. Taft, N. K. Hayward, G. W. Montgomery, J.

S. Mattick, T. R. Mercer, and M. E. Dinger. 2017. Intergenic disease-associated regions are abundant in novel transcripts. *Genome Biology* 18.

261. Takada, S., I. Pico-Knijnenburg, M. Pac, A. Warris, and M. Van Der Burg. 2022. A Pitfall of Whole Exome Sequencing: Variants in the 5′UTR Splice Site of BTK Causing XLA. *Journal of Clinical Immunology* 42: 709-712.

262. Cao, J., E. M. Novoa, Z. Zhang, W. C. W. Chen, D. Liu, G. C. G. Choi, A. S. L. Wong, C. Wehrspaun, M. Kellis, and T. K. Lu. 2021. High-throughput 5′ UTR engineering for enhanced protein production in non-viral gene therapies. *Nature Communications* 12.

263. Sample, P. J., B. Wang, D. W. Reid, V. Presnyak, I. J. Mcfadyen, D. R. Morris, and G. Seelig. 2019. Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. *Nature Biotechnology* 37: 803-809.

264. Morse, L. J., S. M. Payton, G. D. Cuny, and J. T. Rogers. 2004. FDA-Preapproved Drugs Targeted to the Translational Regulation and Processing of the Amyloid Precursor Protein. *Journal of Molecular Neuroscience* 24: 129-136.

265. Pickering, B. M., and A. E. Willis. 2005. The implications of structured 5′ untranslated regions on translation and disease. *Seminars in Cell & Developmental Biology* 16: 39-47.

266. Nelde, A., L. Flötotto, L. Jürgens, L. Szymik, E. Hubert, J. Bauer, C. Schliemann, T. Kessler, G. Lenz, H.-G. Rammensee, J. S. Walz, and K. Wethmar. 2022. Upstream open reading frames regulate translation of cancer-associated transcripts and encode HLA-presented immunogenic tumor antigens. *Cellular and Molecular Life Sciences* 79.

CURRICULUM VITA

NAME	Cassandra Woolley
CONTACT	University of Louisville School of Medicine Clinical and Translational Research Building 505 S. Hancock St. Suite 327 E Louisville, KY 40204 606-922-5882 | cassandra.woolley01@gmail.com

EDUCATION & TRAINING

2018 – Present	M.D., University of Louisville School of Medicine, Louisville, KY
2020 – 2023	Ph.D., Microbiology and Immunology, University of Louisville School of Medicine, Louisville, KY
2016 – 2018	B.S., Biology, University of Kentucky, Lexington, KY
2015 – 2016	Vanderbilt University

HONORS, AWARDS, AND SCHOLARSHIPS

2023	Kentucky Science Advocate, Kentucky Academy of Science
2020 – 2021	AAAS/Science Program for Excellence in Science, American Association for the Advancement of Science
2018	Summa Cum Lade, University of Kentucky
2016 – 2018	Lewis Honors College, University of Kentucky
2016 – 2018	Coca-Cola Commonwealth Scholarship, University of Kentucky
2017	Alumni Association Scholarship, University of Kentucky

2016	Transfer Excellence Scholarship, University of Kentucky
2015 – 2016	College Scholars Program, Vanderbilt University
2015	Cornelius Vanderbilt Scholarship, Vanderbilt University

PROFESSIONAL MEMBERSHIPS

2022 – Present	Kentucky Academy of Science
2021 – Present	American Society for Clinical Pathology
2021 – Present	College of American Pathologists, Medical Student Forum
2020 – Present	American Association for the Advancement of Science
2018 – Present	American Medical Association
2018 – Present	American Physician Scientist Association
2018 – Present	Greater Louisville Medical Society

PUBLICATIONS (*co-first authorship)

Woolley CR, Brinkman NC, Cash ED, Chandran SK, Mitchell TC. 2022. An Unexpected Role for Cell Density Rather than IgM in Cell-Surface Display of the Fc Receptor for IgM on Human Lymphocytes. *ImmunoHorizons* 6:47-63. doi: 10.4049/immunohorizons.2100094.

Woolley CR*, Chariker JH*, Rouchka EC, Ford EE, Hudson EA, Waigel SJ, Smith ML, Mitchell TC. 2022. Reference long-read isoform-aware transcriptomes of 4 human peripheral blood lymphocyte subsets. *G3 Genes|Genomes|Genetics* Volume 12, Issue 11. doi: 10.1093/g3journal/jkac253

Ford EE, Tieri D, Rodriguez O, Francoeur N, Soto J, Kos J, Peres A, Gibson W, Silver CA, Deikus G, Hudson E, **Woolley CR**, Beckmann N, Charney A, Mitchell TC, Yaari G, Sebra RP, Watson CT, Smith ML. 2022. FLAIRR-seq: A novel method for single molecule resolution of near full-length immunoglobulin heavy chain repertoires doi:10.1101/2022.09.24.509352. Cold Spring Harbor Laboratory.

Cash E, Sephton S, **Woolley C**, Elbehi AM, R I A, Ekine-Afolabi B, Kok VC. The role of the circadian clock in cancer hallmark acquisition and immune-based cancer therapeutics. J Exp *Clin Cancer Res*. 2021 Apr 1;40(1):119. doi: 10.1186/s13046-021-01919-5. PMID: 33794967; PMCID: PMC8017624.

SenGupta S, Rane MJ, Uriarte SM, **Woolley C**, Mitchell TC. Human neutrophils depend on extrinsic factors produced by monocytes for their survival response to TLR4 stimulation. *Innate Immun*. 2019;25(8):473-486. doi:10.1177/1753425919871994

AlSiraj Y, **Woolley C**, Thatcher SE, Cassis LA. Chapter 11 - Sex Differences and the Role of the Renin-Angiotensin System in Atherosclerosis and Abdominal Aortic Aneurysms. In: LaMarca B, Alexander BT, eds. *Sex Differences in Cardiovascular Physiology and Pathophysiology*. Academic Press; 2019:167-184. doi:10.1016/B978-0-12-813197-8.00011-7

RESEARCH PRESENTATIONS

C. Woolley, J. Chariker, E. Rouchka, E. Ford, E. Hudson, S. Waigel, C. Casella, M. Smith, T. Mitchell, "Full-length mRNA sequencing resolves 5' UTR regions and suggests novel promoter usage in T cells." Autumn Immunology Conference, Chicago, IL. November 2022.

C. Whitley, **C. Woolley**, C. Casella, T. Mitchell, "Culture systems to model human inflammatory reactions to bacterial endotoxin." Autumn Immunology Conference, Chicago, IL. November 2022.

C. Woolley, J. Chariker, E. Rouchka, E. Ford, E. Hudson, S. Waigel, C. Casella, M. Smith, T. Mitchell. "Full-length mRNA sequencing resolves 5' UTR regions and suggests novel promoter usage in T cells." Research!Louisvillle, Louisville, KY. October 2022.

K. Weston, **C. Woolley**, T. Mitchell, "Satellite Cells: The Pathways and Proteins involved in Muscle Regeneration." APSA Virtual Summer Research Program Symposium, virtual. August 2022.
J. Olabode, **C. Woolley**, T. Mitchell, "Investigating Isoform Novelties in CD62L and Possible Manifestations in Autoimmune Diseases." APSA Virtual Summer Research Program Symposium, virtual. August 2022.

C. Woolley, N. Brinkman, E. Cash, S. Chandran, T. Mitchell. "Regulation of cell-surface expression of the Fc Receptor for IgM (FcMR) in human lymphocytes." National MD-PhD Student Conference, Copper Mountain, CO. July 2022.

C. Woolley, N. Brinkman, E. Cash, S. Chandran, T. Mitchell. “Regulation of cell-surface expression of the Fc Receptor for IgM (FcMR) in human T cells.” HSC Colloquium on Inflammation and Pathogenesis, Louisville, KY. May 2022.

C. Woolley, N. Brinkman, E. Cash, S. Chandran, T. Mitchell. “Regulation of cell-surface expression of the Fc Receptor for IgM (FcMR) in human T cells.” APSA Southeastern Medical Scientist Symposium, virtual. January 2022.

J. Carpenter, **C. Woolley**, T. Mitchell. “Type I Diabetes and Enteroviruses: Clues to Pathogenesis.” APSA Southeastern Medical Scientist Symposium, virtual. January 2022.

J. Carpenter, **C. Woolley**, T. Mitchell. “Type I Diabetes and Enteroviruses: Clues to Pathogenesis.” National Pre-Health Conference, virtual. August 2021.

C. Woolley, S. SenGupta, M. Rane, S. Uriarte, T. Mitchell. “Human neutrophils show intrinsic survival to LTB4 but not LPS,” Southeastern Medical Scientist Symposium, Birmingham, AL. October 2019.

C. Woolley, S. SenGupta, M. Rane, S. Uriarte, T. Mitchell. “Human neutrophils show intrinsic survival to LTB4 but not LPS,” Research!Louisville, Louisville, KY. October 2019.

N. Brinkman, **C. Woolley**, T. Mitchell. "Donor variability in monocyte subsets correlates to cytokine production." Research!Louisville, Louisville, KY. Fall 2019.

C. Woolley, S. SenGupta, M. Rane, S. Uriarte, T. Mitchell. “Human neutrophils depend on accessory cells for their survival response to LPS,” Southeastern Medical Scientist Symposium, Nashville, TN. November 2018.

C. Woolley, S. SenGupta, M. Rane, S. Uriarte, T. Mitchell. “Human neutrophils depend on accessory cells for their survival response to LPS,” Research!Louisville, Louisville, KY. October 2018

University of Kentucky Student Activities Board. “Leaving a Legacy: How to Create a Strong Strategic Plan,” National Association for Campus Activities National Convention, Boston, MA. February 2018.

C. Woolley, Y. AlSiraj, S.E. Thatcher, L.A. Cassis. “The XY sex chromosome complement augments Ang-II induced aortic arch aneurysms in female LDLr-/- mice,” University of Kentucky Cardiovascular Research Day, Lexington, KY. November 2017.

RESEARCH FELLOWSHIPS

2018, 2019	Summer Research Scholars Program, University of Louisville School of Medicine

2017	Undergraduate Summer Training in Cardiovascular Research, University of Kentucky
2016	SyBBURE Searle Undergraduate Research Program, Vanderbilt University

LEADERSHIP

2022 – Present	Louisville Regional Science and Engineering Fair Board Member
2022 – 2023	Co-Chair of the Outreach Committee, Louisville Regional Science and Engineering Fair
2022 – 2023	Director, Science Policy and Outreach Group, University of Louisville School of Medicine
2021 – 2022	Secretary and Marketing Chair, Science Policy and Outreach
Group,	University of Louisville School of Medicine
2019 – 2020	Co-president Innovation in Medicine, University of Louisville School of Medicine
2019 – 2020	Pharmacology Course Representative, University of Louisville School of Medicine
Spring 2019	Immunology Course Representative, University of Louisville School of Medicine
Spring 2019	Orientation Committee, University of Louisville School of
Medicine	
2019 – 2018	Student Activities Board Vice President of Promotions, University of Kentucky
2017 – 2018	Homecoming Coordinator, University of Kentucky
Spring 2017	Property Manager, Alpha Delta Pi, University of Kentucky
2016 – 2017	Philanthropy Assistant, Alpha Delta Pi, University of Kentucky
Spring 2016	Secretary, Relay for Life, Vanderbilt University
2015 – 2016	Special Events Committee and Team Captain, Relay for Life, Vanderbilt University
2015 – 2016	Sutherland House Service Commissioner

2015 – 2016	International Genetic Engineering Machine Team Captain, Vanderbilt University

MENTORSHIP AND ACADEMIC EXPERIENCE

2020 – Present	Mentor, American Physician Scientist Association Undergraduate Mentorship Program
2020 – Present	Research Mentor, Louisville Regional Science and Engineering Fair
2022	Mentor, Kentucky Academy of Science
2021, 2022	Research Mentor and Journal Club Lead, Virtual Summer Research Program, American Physician Scientists Association
2021	Research Mentor, Louisville Science Pathways Research Program
2019 – 2021	Tutor, University of Louisville School of Medicine
2019	Scribe, Disability Resource Center
2017	Teaching Assistant, University of Kentucky ChemExcel
2017	Chemistry and Algebra Tutor, University of Kentucky Center for Academic Resources and Enrichment Services

COMMUNITY SERVICE AND VOLUNTEERISM

2021 – Present	Volunteer, Louisville Regional Science and Engineering Fair
2019 – Present	Volunteer, Hosparus Health
2022, 2023	Volunteer, Kentucky Science Center Celebrations
Fall 2022 Invitational	Event Supervisor, Brown High School Science Olympiad
Summer 2021	Volunteer, Special Olympics Kentucky
2019 – 2020	Volunteer, Family Health Center Longitudinal Clinic
2018 – 2019	Volunteer, Snuggle Squad, Norton Children's Hospital Neonatal Intensive Care Unit

Summer 2018	Volunteer, Veteran's Wheelchair Games
Spring 2018	Volunteer, Surgery on Sunday 5K
2016 – 2017	Volunteer Tutor, Lexington Public Library
2016 – 2017	Volunteer, Ronald McDonald House Charities
Spring 2016	Alternative Spring Break, Medical and Elementary School Service, Jellico TN
2015 – 2016	Volunteer, Vanderbilt Relay for Life