2019

Investigations of the DNA-binding activity and gene regulatory properties of IRF3, IRF5, and IRF7 homodimers

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

INVESTIGATIONS OF THE DNA-BINDING ACTIVITY AND GENE REGULATORY PROPERTIES OF IRF3, IRF5, AND IRF7 HOMODIMERS

by

KELLEN K. ANDRILENAS

BS, Biology, University of Washington, 2011
BS, Psychology, University of Washington, 2011
MA, Biology, Boston University, 2016

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2019

Approved by

First Reader

Trevor Siggers, Ph.D.
Associate Professor of Biology

Second Reader

Juan Fuxman-Bass, Ph.D.
Assistant Professor of Biology

ACKNOWLEDGMENTS

First, I would like to acknowledge my husband Ethan for his support and encouragement across my graduate career. Thank you for listening to me go on about transcriptional regulation and the latest paper I was excited about. I love you and couldn't have done this without your partnership.

I also acknowledge my family and friends for their support and for providing an environment outside of graduate school to unwind. I also couldn't have done this without the compassionate ear of my mother and step-dad; thank you for listening to me even when the science got complicated. I am grateful for friends that have been a well-spring of encouragement and understanding. Sanda Zolj, Kimberly Nguyen, Millan AbiNader, J. Eliot DeGolia, Barkha Shah, Gregg Harbaugh, Nick Lopez and many more – Thank you for everything.

If it takes a village to raise a child, it takes a community to raise a PhD student. I would like to acknowledge the community of scientists and staff that make this department function and have enriched my time at Boston University. Dennis Batista, Christina Honeycutt, Peter Buston, Todd Blute, Peter Castellano, Maddy Davis, Charlie Kieswetter, Kathryn Spilios, Ilda Freitas, Tom Symancyk, Andrea

Voehringer and so many more. You are the substrate of this institution and I appreciate what you do and have done.

I want to acknowledge my colleagues and members of the Siggers Lab for their camaraderie and thoughtful scientific conversations. My time in the Siggers Lab wouldn’t have been the same without you. Ashley Penvose, Jessica Keenan, Nima Mohaghegh, David Bray, Heather Hook, VIjendra Ramlal – Thank you! I also want to acknowledge the undergraduate students that have helped with this research: Brandon Leung and Amanda Chaplin – thank you for your dedication, enthusiasm and willingness to learn.

I would like to acknowledge my advisor Trevor Siggers – Thank you for taking me in as a young evolutionary biology/invertebrate zoology/psychology student. I have greatly enjoyed our scientific conversations and I would never have realized the stunning complexity and beauty of the human immune system without your help.

Last but not least, I thank you (the reader) and my committee members past and present for reading this thesis and providing valuable feedback over the course of my PhD: Tom Gilmore, David Waxman, Juan Fuxman-Bass, Ulla Hansen, Chip Celenza, Frank Naya – Thank you for your insight, support and time.

THE DNA-BINDING AND GENE-REGULATORY SPECIFICITY OF IRF3, IRF5 AND IRF7 HOMODIMERS

KELLEN K. ANDRILENAS

Boston University Graduate School of Arts and Sciences, 2019

Major Professor: Trevor Siggers, Associate Professor of Biology

ABSTRACT

The innate immune response is an essential component of the mammalian immune system that responds rapidly to pathogens. This response to pathogens is initiated by the detection of pathogen associated molecular patterns (PAMPs) by pathogen recognition receptors (PRRs). PRR signaling activates antipathogen gene programs via transcription factors (TFs) such as the interferon regulatory factors (IRFs). IRF3, IRF5, and IRF7 (IRF3/5/7) are key signal-dependent TFs that have overlapping, yet distinct, roles in the mammalian response to pathogens. To examine the role that DNA-binding specificity plays in delineating IRF3/5/7-specific gene regulation, we used protein-binding microarrays (PBMs) to characterize the DNA binding of IRF3/5/7 homodimers. We identified both common and dimer-specific DNA binding sites, and show that DNA-binding differences can translate into dimer-specific gene regulation. Central to the antiviral response, IRF3/5/7 regulate type I interferon (IFN) genes. We show that IRF3 and IRF7 bind to many interferon-stimulated response element (ISRE)-type sites in the virus-response elements (VREs) of IFN promoters. However, strikingly, IRF5 does not bind the VREs, suggesting evolutionary selection against IRF5 homodimer binding. Mutational analysis identified a a critical specificity-determining residue that inhibits IRF5 binding to the ISRE-variants present in the IFN gene promoters. Integrating PBM and reporter gene data we find that both DNA-binding affinity and affinity-independent mechanisms determine the transcriptional activation ability of DNA-bound IRF dimers, suggesting that DNA-based allostery plays a role in IRF binding site function. To assay the sequence determinants of IRF-dependent transcriptional regulation, we propose using a modified massively parallel reporter assay (MPRA). The proposed MPRA leverages unique molecular identifiers to improve the accuracy of reporter gene quantitation. This work provides new insights into the role and limitations of DNA-binding affinity in delineating IRF3/5/7-specific gene expression and lays groundwork for further understanding the complexities of IRF-dependent transcriptional regulation of innate immune genes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

°C	degrees Celsius
aa	amino acid(s)
bp	base pair(s)
BSA	Bovine Serum Albumin
BU	Boston University
C-term	carboxy terminal
cDNA	complementary DNA
c	carboxyl
CBP	CREB-binding protein
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
DBD	DNA-binding domain
dH_2O	deionized water
DMEM	Dulbecco’s modified Eagle’s medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphate
DTT	dithiothreitol
E. coli	Escherichia coli
EDTA	ethylenediamine tetraacetic acid disodium salt

FBS	fetal bovine serum
g	gram(s)
GST	glutathione-S-transferase
eGFP	enhanced Green Fluorescent Protein
h	hour(s)
IFN	Interferon
IKK	IκB kinase
IPTG	isopropyl-β-D-1-thiogalactopyranoside
IRF	interferon regulatory factor
IRF3/5/7	IRF3, IRF5 and IRF7
kDa	kilodalton(s)
LB	Luria broth
LPS	lipopolysaccharide
M	molar
mA	milliamp(s)
MCS	multiple cloning site
mg	milligram(s)
min	minute(s)
ml	milliliter(s)
mM	millimolar
MPRA	Massively Parallel Reporter Assay

mRNA	messenger RNA
N	amino
NEB	New England Biolabs
NES	nuclear export signal
NF-κB	nuclear factor-κB
OD	optical density
p	phosphor
PBM	Protein Binding Microarray
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEI	polyethylenimine
PMSF	phenylmethylsulfonyl fluoride
rcf	relative centrifugal force
RLR	RIG-I like Recptor
rpm	revolutions per minute
RT	room temperature
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
Sec	second(s)
SNV	Single Nucleotide Variant
TB	terrific broth

TBK1	TANK binding kinase 1
TBS	Tris-buffered saline
TBS-Tween	TBS with Tween 20
TE	Tris buffer with EDTA
TEMED	N, N, N’, N’-tetramethylethylenediamine
TF	transcription factor
TLR	Toll-like receptor
Tris	1-[bis(2,3-dibromopropoxy)phosphinoyloxy]-2,3-dibromopropane
Tween-20	polyxyethelene sorbitan monolaurate
UV	ultraviolet
v	volume
V	volts
w	weight
WT	wild-type
µg	micrograms(s)
µl	microliter(s)
µM	micromolar

1 CHAPTER ONE - INTRODUCTION

1.1 The innate immune response

The innate immune response is an essential component of the mammalian immune system that responds rapidly to pathogens. The swiftness of the innate immune system depends on the detection of Pathogen Associated Molecular Patterns (PAMPs), which are conserved molecular features shared across many classes of pathogen (Kumar et al., 2011). Lipopolysaccharide (LPS; present on the exterior of many bacteria) and viral genome nucleic acids (ssDNA, dsRNA) are PAMPs that elicit strong, pathogen-specific innate-immune responses. Detection of these PAMPs is mediated by Pathogen Recognition Receptors (PRRs) present on both the exterior and interior of most cell types (Lee and Kim, 2007; Takeuchi and Akira, 2010). When a PAMP binds to its corresponding PRR, a rapid signaling cascade activates anti-pathogen gene programs necessary for fighting the detected pathogen and broadcasting the alarm to surrounding cells (Kawai and Akira, 2006). Two well-known PRR families are the Toll-Like receptors (TLRs) and the RIG-I-Like receptors (RLR) (Creagh and O'Neill, 2006, 2006; Vogel et al., 2003). TLRs respond to a broad range of PAMPs associated with both bacteria and viruses (Kawai and Akira, 2011; Medzhitov, 2001), while RLRs respond to viral nucleic acids (Loo, 2011).

1.1.1 Innate immune response to viruses

Viruses infect a wide range of cell types in the human body. After the detection of viral PAMPs by PRRs, three key innate immune responses attempt to halt the progress of a viral infection: **(1)** Infected cells up regulate antiviral effectors and may initiate apoptosis to prevent viral compromise of cellular machinery (Chattopadhyay et al., 2013; Liu et al., 2013; Zhou et al., 2013); **(2)** Infected cells produce inflammatory cytokines that recruit immune cells to the site of infection (Hiscott, 2007; Kimura et al., 2013; Lazear et al., 2015); **(3)** Activated immune cells, like dendritic cells and macrophages, produce interferons and cytokines that activate antiviral responses in many cell types and help stimulate the adaptive immune response (Barnes et al., 2004; Matta and Barnes, 2019; Tailor et al., 2006; Taniguchi et al., 2001). These innate immune functions are coordinated by multiple transcription factor (TF) families activated by PRR signaling (Hiscott, 2007).

1.2 Transcriptional regulation of the innate immune response

The innate immune system is dependent on complex signaling cascades which unfold at the proteomic level; however, an essential endpoint of these mechanisms is a change in cell state orchestrated by transcriptional regulation. The activation of immune cells in response to pathogens and the production of anti-viral effectors is driven by signal dependent TFs in conjunction with cell type-specific TFs (Mancino et al., 2015; Medzhitov and Horng, 2009). This integration of environment (signal dependent TFs) and cellular context (cell-type-specific TFs) allows the human immune system to mount pathogen-specific and immune cell-type-specific responses (Pope and Medzhitov, 2018). Here we focus on the Interferon Regulatory Factors, which play a central role in the innate immune response to viruses and bacteria.

1.2.1 The Interferon Regulatory Factors (IRFs)

The Interferon Regulatory Factors (IRFs) were first described approximately 30 years ago — identified by their ability to bind the Interferon beta promoter (Enoch et al., 1986; Miyamoto et al., 1988; Zinn and Maniatis, 1986). The IRFs are named for their role in regulating the family of immune signaling peptides known as interferons, which coordinate the anti-viral immune response. Since the discovery of the first IRF (IRF1), nine human IRFs have been characterized with diverse roles in immune regulation and cell development (Table 1.1). The IRFs can be divided into two general categories: signal-dependent (IRF1, IRF3, IRF5, IRF7) (Honda, Taniguchi, 2006), and cell-type-dependent IRFs (IRF4, 8) (Tamura et al., 2008) — notably excluding IRF2, which is an IRF1-induced repressor (Honda, Taniguchi, 2006), and IRF6, which is involved in craniofacial development (Ingraham et al., 2006).

Despite their diverse functions, all IRFs contain a highly conserved N-terminal DNA-binding domain (DBD) that forms a helix-turn-helix structure (Figure 1.1 and Figure 1.2). The IRF core binding motif is a short 5’-GAAA-3’ sequence which is recognized by contacts along the α3 helix of the IRF DBD which inserts into the major groove of DNA (Figure 1.2). The signal-dependent IRF3, IRF5, and IRF7 bind DNA as dimers and recognize a consensus interferon-stimulated response element (ISRE) motif, 5’-GAAANNGAAA-3’. ISREs are found in the promoters of key immune genes such as the type-I-interferons (IFNα, IFNβ) (Panne et al., 2007), type-III-interferons (IFNλ) (Gad et al., 2009; Hillyer et al., 2012; Iversen and Paludan, 2010), cytokines (CXCL10) (Brownell et al., 2014), and important antiviral effectors (IFIT1-3) (Diamond and Farzan, 2012; Zhou et al., 2013).

IRF3, IRF5 and IRF7 (IRF3/5/7) are constitutively expressed at low levels in multiple cell types and their signal-dependent activation is an essential part of the innate immune response (Honda and Taniguchi, 2006; Tamura et al., 2008). Constitutive expression and inducible activation allow IRF3/5/7 to rapidly drive anti-pathogen gene programs (Chen and Royer, 2010; Honda and Taniguchi, 2006). IRF3/5/7 are activated by PRR signaling and bind DNA as phosphorylation induced dimers (Chen and Royer, 2010) (Figure 1.3). Dimerization is mediated through a C-terminal IRF-association domain (IAD), which is regulated by an auto-inhibitory domain (AID) (Chen and Royer, 2010) (Figure 1.3). In the inhibiting state, the AID forms a hydrophobic structure that occludes the IAD, preventing dimerization (Chen, Royer, 2010). A hinge region in the AID is targeted by PRR-activated kinases (TBK1, IKKe, IKKb) (Chen et al., 2008; Marié et al., 2000; Ren et al., 2014; Ryzhakov et al., 2015; Takahasi et al., 2010), which phosphorylate serine and threonine residues in the hinge, causing the opening of the hydrophobic AID (Chen, Royer, 2010) (Figure 1.3).

Considerable cross-talk between PRR signaling pathways contributes to a robust innate immune response (Czerkies et al., 2018; Kawai and Akira, 2011), but can make mechanistic interpretations of PRR signaling in stimulated cells challenging. Phosphomimetic IRF constructs have been used extensively in the field to study IRF dimer-dependent gene regulation and biochemistry (Caillaud et al., 2005; Chen et al., 2008; Chen and Royer, 2010; Cheng et al., 2006; Clement et al., 2008; Dragan et al., 2007; Foreman et al., 2012; Marié et al., 2000; Mori et al., 2004; Prakash and Levy, 2006; Qin et al., 2003) . Serine to aspartic acid (S→D) amino acid substitutions in the AID generate constitutively dimeric IRF3/5/7 proteins. We use them throughout the research in this dissertation as an effective molecular tool for exploring IRF3/5/7 DNA binding and transcriptional regulation (see chapter 3).

1.2.2 The interferon regulatory factors 3, 5 and 7 have overlapping yet distinct roles

IRF3, IRF5 and IRF7 have essential roles in the innate immune response to viruses and bacteria. A central output of the IRF-dependent immune response is the rapid production of Interferon beta (IFNβ) followed by the expression of IFN alpha (IFNα) genes. The secretion of IFNβ by virus-infected cells triggers anti-viral responses initiated by IFNβ receptor signaling. The direct role of IRF3 and IRF7 in driving IFNβ expression has been extensively studied.

The canonical model of IFNβ regulation involves the cooperative binding of multiple transcription factors (cJun, ATF2, NFκB, IRF3, IRF7) that form an 'enhanceosome' required to initiate transcription in a tightly regulated manner (Figure 1.4) (Panne et al., 2007). IRF3 is ubiquitously expressed in most cells and is thought to initially drive IFNβ production (Taniguchi and Takaoka, 2001) (Figure 1.5a). Next, IFNβ receptor signaling up-regulates the expression of IRF7, which then continues to drive IFNβ production as an IRF7 homodimer, or IRF3/7 heterodimer (Figure 1.5b). The level of basal IRF7 expression varies across immune cell types and cell environment, which can alter innate immune gene expression profiles and the role of IRF3 or IRF7 in IFNβ production. The role of IRF5 in IFNβ production is less clear; however, immune cells from IRF3(-)/IRF7(-) knock-out mice are still capable of producing IFNβ and this activity is lost in IRF3/5/7(-) triple knockout mice (Lazear et al., 2013). Recently, a distinct role of IRF5 in driving IFNβ production was shown in human plasmacytoid dendritic cells (pDCs) (Chow et al., 2018). In these cells, IRF5 was found to drive IFNβ expression after endosomal TLR signaling, which would allow pDCs to produce IFNβ after sampling their environment via endocytosis as part of their surveillance role in the immune system. Chow et al. (2018) illustrated the complex and overlapping roles of IRF3/5/7 at the same regulatory locus (i.e. IFNβ) that are fundamental to understanding IRF-dependent gene regulation.

Another important set of IRF gene targets are the interferon-alpha family of signaling peptides. Humans have 13 IFNα genes, which are differentially regulated by IRF3/5/7 (Genin et al., 2009). All IFNα gene promoters feature a conserved Viral Response Element (VRE) which contains multiple ISREs or ISRE-like binding sites (Figure 1.6). Sequence variations in the IFNα VREs likely contribute to the differential regulation of these genes by the IRFs. IRF3/5/7 regulation of the IFNα genes is complex, with evidence that IRF3 and IRF7 function as transcriptional activators or repressors depending on the relative levels of IRF3/7 expression (Genin et al., 2009). IRF5 drives the expression of a subset of IFNα genes (Barnes et al., 2001), but has been described as a master regulator of inflammatory cytokines (Takaoka et al., 2005). Despite their central role in the response to pathogens, little is known about the mechanisms by which IRF3/5/7 target both overlapping and distinct gene programs.

1.3 Methods for systematically understanding IRF dependent gene regulation

1.3.1 Protein Binding Microarrays (PBMs)

PBMs are a well-established microarray-based technique to study the in vitro binding of proteins to DNA (Berger et al., 2006; Bulyk et al., 1999; Field et al., 2006; Linnell et al., 2004; Mukherjee et al., 2004) . PBM experiments involve applying protein to a double-stranded DNA microarray and then quantifying the amount of DNA-bound protein using a fluorescently labeled antibody (Figure 1.7). Available high-density, multi-chambered microarray platforms allow the DNA binding of multiple protein samples to be tested in parallel to tens of thousands of DNA sequences (Badis et al., 2009; Berger et al., 2008, 2006). This high-throughput (HT) platform has facilitated numerous DNA-binding studies on large groups of TFs (Badis et al., 2009, 2008; Berger et al., 2008; Franco-Zorrilla et al., 2014; Gordân et al., 2011; Grove et al., 2009; Jolma et al., 2013; Wei et al., 2010; Zhu et al., 2009) that have provided rich data sets for comparative analysis of TF-DNA-binding specificity. Furthermore, PBM experiments can be used to measure protein–DNA-binding interactions spanning several orders of magnitude in affinity, and resolve binding affinities that differ by less than 1.5-fold (Siggers et al., 2011). Therefore, comparison of PBM-binding profiles provides a detailed and sensitive approach to compare the DNA-binding specificity of TFs.

PBMs can be used to assay binding to synthetic (Badis et al., 2009; Berger et al., 2008, 2006; Linnell et al., 2004; Siggers et al., 2012) or genome-derived (Bolotin et al., 2010; Gordân et al., 2013; Mukherjee et al., 2004; Siggers et al., 2011) DNA sequences. Studies focused on specific TFs, including those presented in this dissertation, have selected microarray oligos based on prior knowledge about TF specificity (Andrilenas et al., 2018; Bolotin et al., 2010; Fang et al., 2012; Udalova et al., 2002; Wong et al., 2011) . PBMs using genome-derived DNA sequences have also been used to analyze TF specificity and have been instrumental in identifying features such as the role of flanking DNA (i.e. genomic context of a DNA-binding sites) (Gordân et al., 2013) and coregulatory motifs for multi-protein complexes (Siggers et al., 2011). In summary, the PBM is a robust, HT methodology that provides a sensitive and flexible platform with which to examine TF–DNA-binding specificity.

1.3.2 Massively Parallel Reporter Assays

Luciferase reporter assays have been used to study transcriptional regulation since the characterization of firefly luciferase in 1985 (Wet et al., 1985). Over the past decade, decreases in the cost of high-throughput sequencing (HTS) and DNA synthesis have enabled greater use of HTS technologies in biological research (Bonetta, 2010; Sboner et al., 2011). Massively parallel reporter assays (MPRAs) are a recently developed technique that uses HTS to measure the transcriptional activity of many thousands of cis-regulatory elements (CREs) at once (Kwasnieski et al., 2012; Melnikov et al., 2012; Tewhey et al., 2016). MPRAs achieve this scale by quantitating mRNA abundance transcribed from a complex pool of reporter plasmids after transfection into cells. Every CRE present in the MPRA plasmid pool is identified by a sequence barcode embedded in the plasmid's 3'-untranslated region (3'-UTR) (Figure 1.8a). This transcribed barcode is a key feature of MPRAs and results in reporter mRNAs that identify which CRE induced their expression. After mRNA extraction, CRE-associated barcodes are quantified using RNA-seq and normalized to the CRE plasmid abundances (Figure 1.8c). By using high-throughput sequencing, systematic comparisons of many CREs can be performed at a scale that would be prohibitive using low-throughput luciferase assays.

MPRAs are an excellent experimental paradigm for measuring the relationship between TF DNA-binding (PBMs) and transcriptional regulation. The highly-detailed protein:DNA-binding affinity data generated by PBMs pairs well with MPRA experiments. For example, single-nucleotide variation (SNV) alteration of transcription factor binding sites, and thus the TF dependent regulation, of gene expression is a primary mechanistic model for understanding disease-associated expression quantitative trait loci (eQTLs) (Brown et al., 2013; Kasowski et al., 2010; Majewski and Pastinen, 2011). By integrating PBM-derived TF DNA-binding affinities and MPRA data, greater mechanistic insights into human variation may be achieved. For example, genome derived and mutational MPRAs have been used to functionally assess enhancer sequences and explore the impacts of single-nucleotide variants on transcriptional regulation (Kwasnieski et al., 2012; Patwardhan et al., 2009; van Arensbergen et al., 2019).

1.4 Thesis rationale

The innate immune response is an essential component of host defense. Immune responses to pathogens are tightly regulated, allowing a robust response to pathogens without damaging the host. Defects in innate immune signaling can result in chronic bacterial infections and other diseases due to a compromised immune response (Ku et al., 2005; van de Vosse et al., 2009), while overactive immune responses are a hallmark of autoimmune disease (Ghodke-Puranik and Niewold, 2015; Matta and Barnes, 2019). Understanding the regulation of normal immune responses at a mechanistic level may provide insights valuable for human health.

IRF3, IRF5 and IRF7 play a key role in regulating the innate immune response to viruses through complex regulatory mechanisms. Considerable progress has been made toward understanding the regulation and function of the Interferon Regulatory Factors since their discovery 30 years ago (Enoch et al., 1986; Miyamoto et al., 1988; Zinn and Maniatis, 1986); however, key questions still remain. For example, the molecular determinants of IRF3/5/7's simultaneously overlapping yet distinct gene regulatory activity are still not understood at a mechanistic level.

The research presented in this dissertation makes inroads to understanding how IRF3, IRF5 and IRF7 induce both common and factor-specific gene targets. We systematically characterize the DNA-binding landscapes of homodimeric IRF3/5/7 proteins, using PBMs, and examine the impact of small variations in IRF binding sites on gene expression (Chapter 3). Additionally, we propose a modified MPRA experimental design to elucidate the relationship between IRF DNA-binding affinity and transcriptional regulation which leverages massively parallel reporter assays (chapter 4).

Table 1.1 – The IRFs have diverse roles

Refer to source paper for table references. **Adapted from Tamura et al.** (2008)

	Expression	Role in immune cell function (target genes)	Role in development of immune cells and other cells	Role in cell growth	References
IRF1	• Constitutive and IFN–inducible in various cell types • Inducible by DNA damageat transcriptional and posttranslational levels • Mainly in the nucleus and partially in the cytoplasm • Modified by TLR signalingto efficiently translocate to the nucleus	• Stimulates expression of IFN–inducible genes (GBP, iNOS, Caspase–1, Cox–2, CIITA, TAP1, and LMP2) • Binds to MyD88 and enhances TLR–dependent gene induction in IFN–γ–treated cells (IFN–β, iNOS, IL–12p35, and IL–12p40)	• Required for NK cell development (IL–15 in bone marrow stromal cells) • Required for differentiation of $CD8^+$ T cells • Promotes Th1 differentiation through NK cells (IL–15), M s/DCs (IL–12), and a T cell–intrinsic mechanism • Suppresses Th2 differentiation (represses IL–4)	• Suppresses oncogene–induced transformation (Lysyl oxidase) • Required for DNA damage–induced growth arrest (p21/WAF1/CIP1) • Required for DNA damage–induced apoptosis	2, 53, 134, 220–222, 231–233, 249, 250, 256, 260, 271, 272, 278, 280, 294, 295, 349
IRF2	• Constitutive and IFN–inducible in various cell types	• Attenuatestype I IFN responses by antagonizing IRF1and IRF9 • In some casescooperates with IRF1 to activate transcription (IL–12p40 and Cox–2)	• Required for differentiation of $CD4^+$ DCs • Required for NK cell development • Suppresses basophil expansion • Promotes Th1 differentiation (IL–12 in M s) • Suppresses Th2 differentiation (represses IL–4)		48, 53, 179, 180, 220, 224, 228, 234, 252, 260, 271, 336
IRF3	• Constitutive in various cell types • Mainly in the cytoplasm • Phosphorylated upon virus infection, TRIF–dependent signaling, cytosolic PRR signaling, and DNA damage, and then translocates to the nucleus	• Inducestype I IFNs (IFN–α4 and IFN–β) and chemokines (CXCL10) upon virus infection, TLR stimulation, and DNA stimulation	Unknown	• Stimulates apoptosis in M s upon bacterial infection • May promote DNA damage–induced apoptosis	36, 39, 51, 54, 56, 59, 61, 99, 105, 350
IRF4	• Constitutive in B cells, M s, and $CD11b^+$ DCs and inducible by antigen stimulation in T cells and by TLR signaling in M s • Mainly in the nucleus and partially in the cytoplasm	• Binds to MyD88 and negatively regulates TLR–dependent induction of proinflammatory cytokine genes	• Required for differentiation of $CD4^+$ DCs • Supports B cell development (Ig light chains) • Required for plasmacell differentiation (Blimp–1 and AID) and GC formation • Required for Th2 differentiation (IL–4)	• May possessoncogenic potential	133, 135, 137, 138, 238, 240, 244, 246, 253, 261, 262, 266

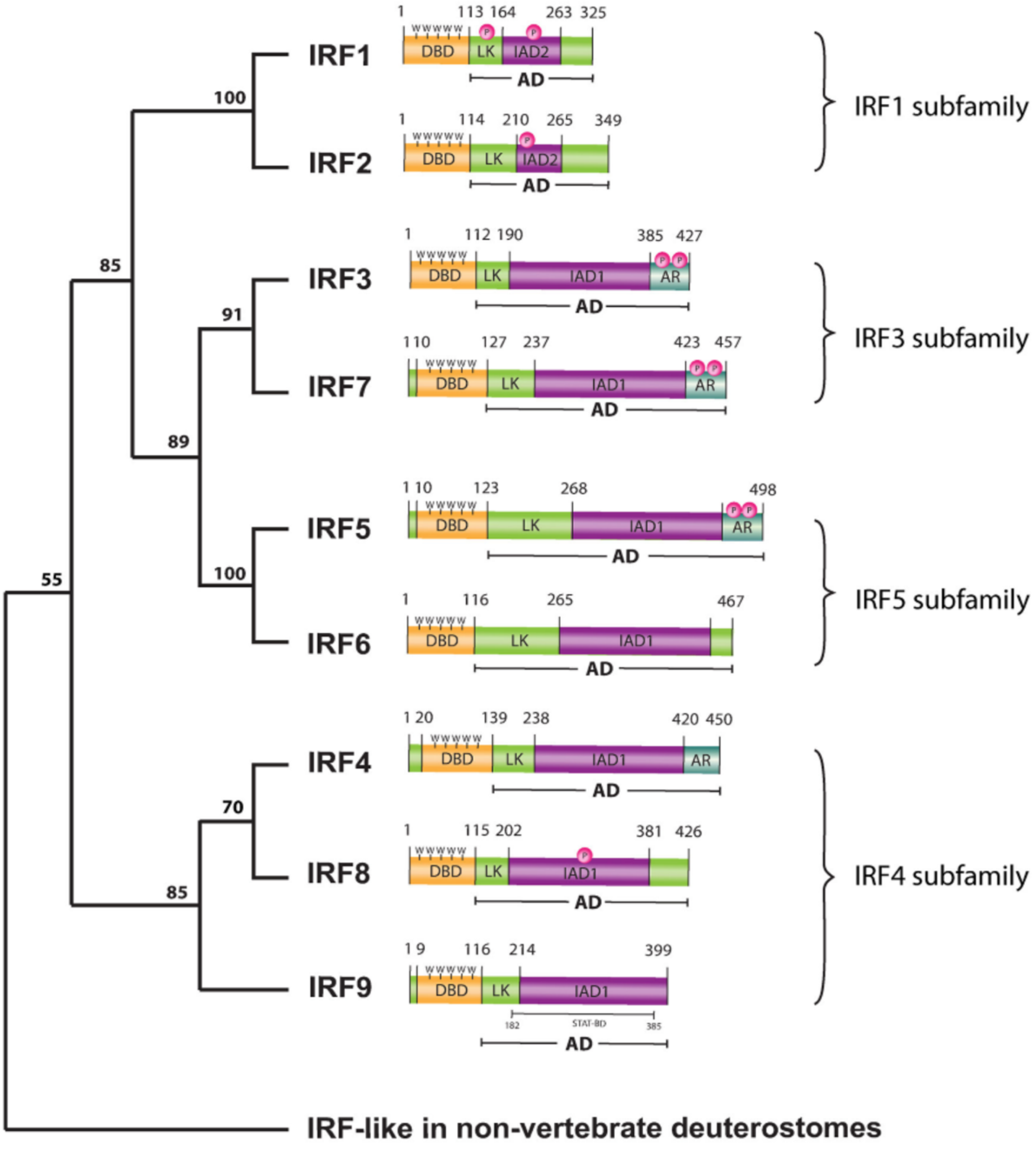
Tamura, T., Yanai, H., Savitsky, D. and Taniguchi, T. (2008) 'The IRF Family Transcription Factors in Immunity and Oncogenesis', Annual Review of Immunology, 26(1), pp. 535–584. doi: 10.1146/annurev.immunol.26.021607.090400.

	Expression	Role in immune cell function (target genes)	Role in development of immune cells and other cells	Role in cell growth	References
IRF5	• Constitutive in B cells and DCs, and inducible by type I IFNs and TLR signaling • Mainly in the cytoplasm • Phosphorylated upon virus infection, TLR-dependent signaling, and DNA damage, and then translocates to the nucleus	• Binds to MyD88 and positively regulates TLR-dependent induction of proinflammatory cytokine genes (IL-12p40, IL-6, and TNF-α) • Induces type I IFNs and proinflammatory cytokines upon virus infection (type I IFNs, IL-6, and TNF-α)	Unknown	• Suppresses oncogene-induced transformation • Required for DNA damage-induced apoptosis	46, 65, 130, 314, 315
IRF6	• Constitutive in skin	Unknown	• Required for keratinocyte differentiation	Unknown	267, 268
IRF7	• Constitutive in B cells, pDCs, and monocytes and inducible by type I IFNs in various cell types • Mainly in the cytoplasm • Phosphorylated upon virus infection and TLR-dependent signaling, and then translocates to the nucleus	• Binds to MyD88 and positively regulates TLR-dependent induction of type I IFNs (IFN-α/β)	Unknown	Unknown	44, 45, 49, 50, 64, 115, 123, 124, 128, 351
IRF8	• Constitutive in B cells, M s, and CD11b$^-$ DCs and further inducible by IFN-γ in M s and by antigen stimulation in T cells • Mainly in the nucleus and partially in the cytoplasm	• Binds to TRAF6 and is required for TLR9 signaling in DCs and M s • Promotes type I IFN production in DCs (IFN-α/β) • Stimulates IFN-γ- and PAMP-inducible genes (IL-12p40, iNOS, FcγRI, PML, and others)	• Required for differentiation of CD8α$^+$ DCs and pDCs • Stimulates M differentiation (Blimp-1, METS, and lysosomal/endosomal enzyme-related genes; represses disabled-2) • Supports B cell development (Ig light chains) • Stimulates the GC program (BCL6 and AID) • Promotes Th1 differentiation through M s/DCs (IL-12)	• Inhibits myeloid cell growth (Blimp-1, METS, and p15/INK4B) • Promotes apoptosis in myeloid cells • Its absence leads to CML-like disease	52, 140, 142, 173, 175, 187, 189, 190, 193, 199, 202, 214, 215, 238, 243, 352
IRF9	• Constitutive and inducible by IFN-γ in various cell types • Mainly in the nucleus	• Binds to STAT1 and STAT2 to form ISGF3 and stimulates type I IFN-inducible genes (OAS, PKR, IRF7, and many others)	Unknown	• Mediates type I IFN induction of p53 (p53)	4, 331, 353, 354

Tamura, T., Yanai, H., Savitsky, D. and Taniguchi, T. (2008) ‘The IRF Family Transcription Factors in Immunity and Oncogenesis’, Annual Review of Immunology, 26(1), pp. 535–584. doi: 10.1146/annurev.immunol.26.021607.090400.

Figure 1.1 – Phylogenetic tree of the Human IRF proteins and their structural domains

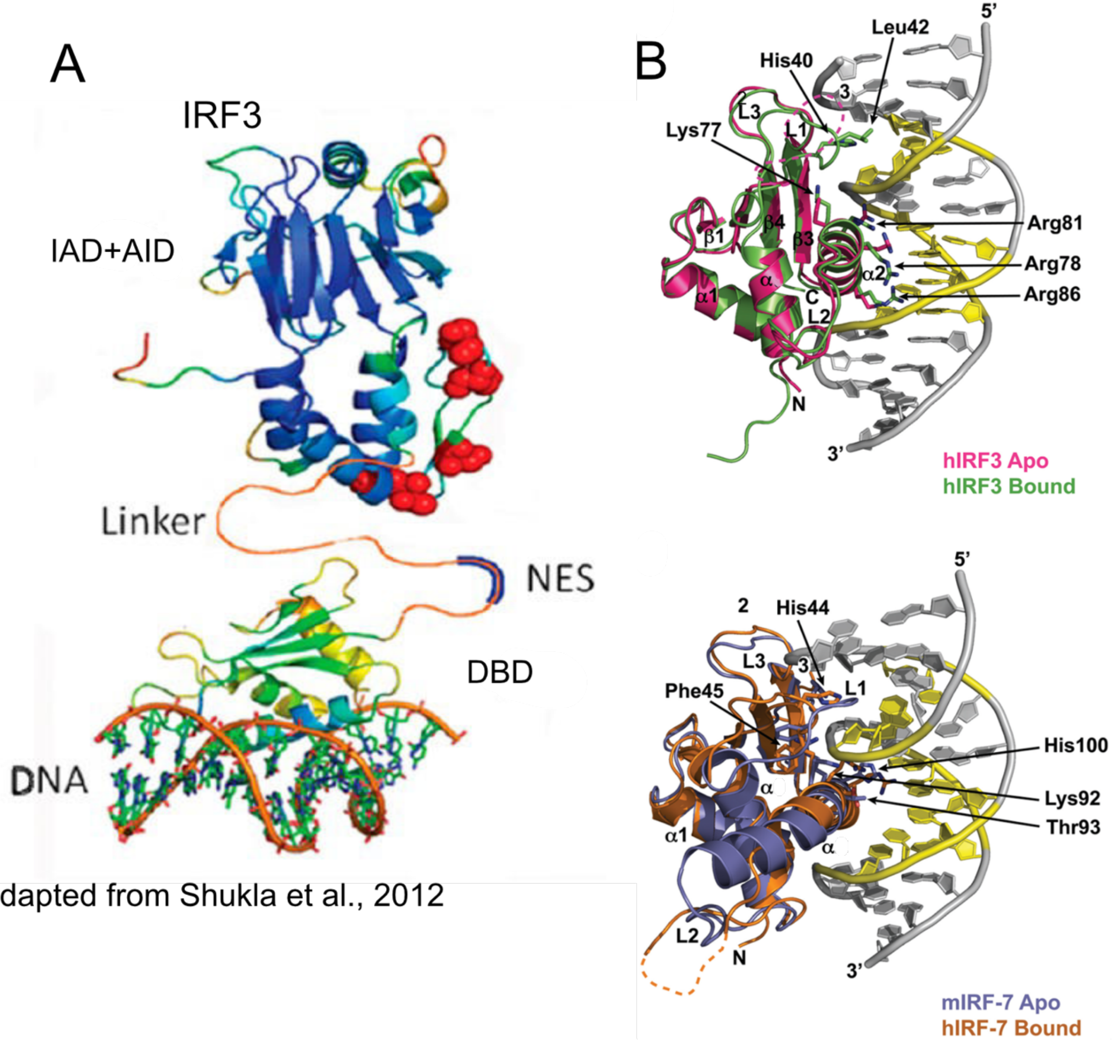
The human IRFs are structurally similar. IRF3, 7, 5, and 6 form an IRF structural subfamily. Labels – DNA-binding Domain (DBD), linker (LK), IRF Association Domain (IAD), Auto-inhibitory domain (AR). Phosphorylation sites are marked in pink. **Adapted from Antonczyk et al (2019).** (Antonczyk et al., 2019)



From Antonczyk et al., 2019

Figure 1.2 – Structural feature of the IRFs (Composite full-length and DNA Binding Domain)

A) A composite 3D structure showing the IRF DBD, and C-terminal IRF-association domain and inhibitory domain (IAD+AID). **B)** Crystal structure of the IRF3 [top] and IRF7 [bottom] DBD interacting with DNA. An alpha helix inserts into the major groove to make contacts with DNA. **Adapted from De Ioannes et al., 2011 and Shukla et al., 2012.**
(De Ioannes et al., 2011; Shukla et al., 2012)

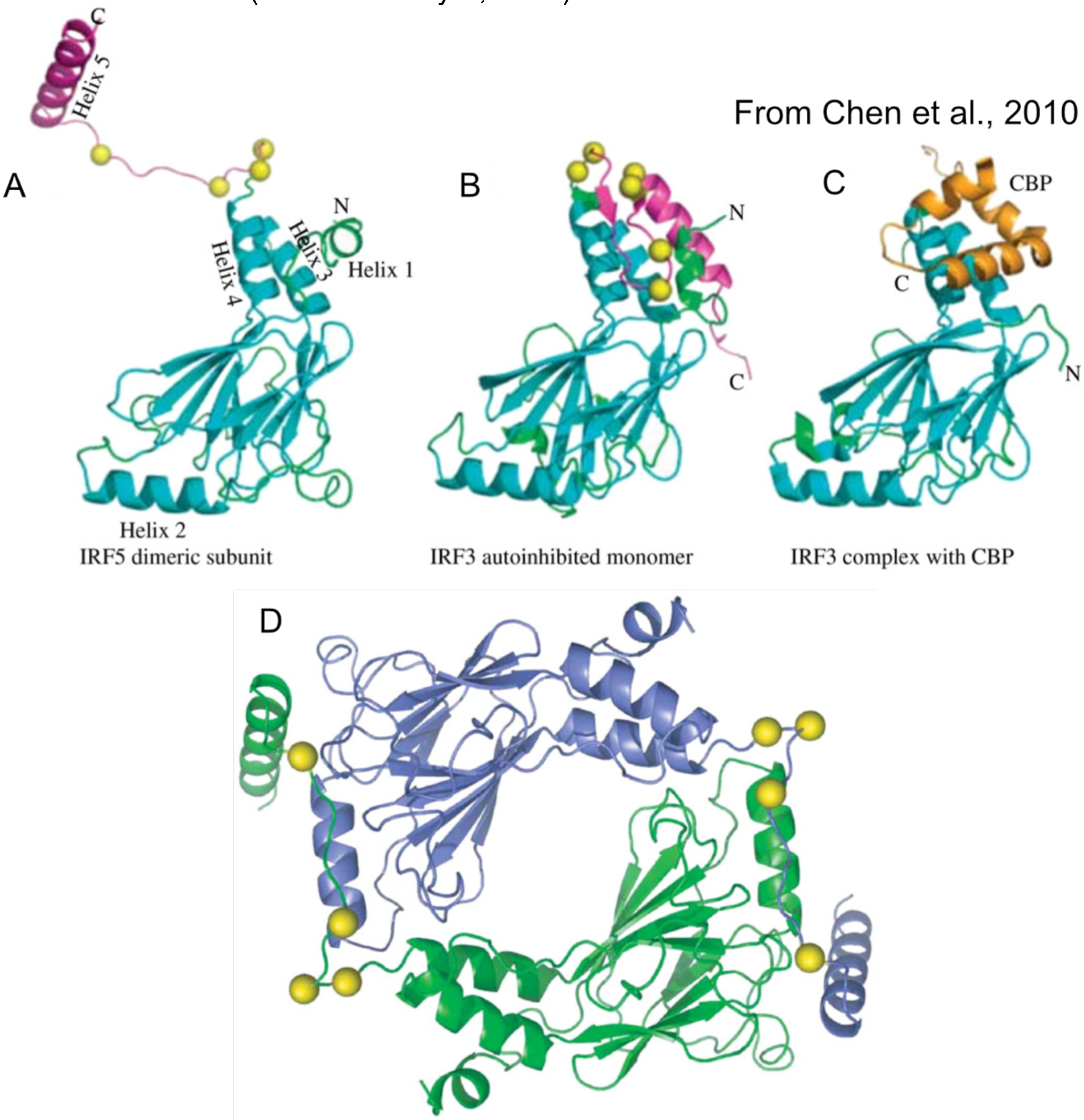


Adapted from Shukla et al., 2012

Adapted from De Ioannes et al., 2011

Figure 1.3 – IRF3/5/7 form dimers after activating dimerization

IRF3/5/7 are activated by phosphorylation. **A)** Crystal structure of the IRF5 C-terminal domain (AID+IAD) in an extended, phosphorylated state (yellow – phosphates). The phosphorylation opens the hydrophobic helix bundle shown in **(B). C)** Structure showing how the co-activator CBP interfaces with IRF3 without an (IAD). **D)** Structure of the dimerized IRF5 c-terminal domain. **Adapted from Chen et al. 2010** (Chen and Royer, 2010)



From Chen et al., 2010

IRF5 dimeric subunit

IRF3 autoinhibited monomer

IRF3 complex with CBP

Figure 1.4 - The structure of the interferon beta enhanceosome

A) Crystal structure model of the IFNβ enhanceosome complex. **B)** Model of the enhanceosome complex interacting with CBP co-activators and transcriptional machinery after successful assembly. **Adapted from Honda et al., 2006 and Panne et al., 2007** (Honda et al., 2006; Panne et al., 2007)

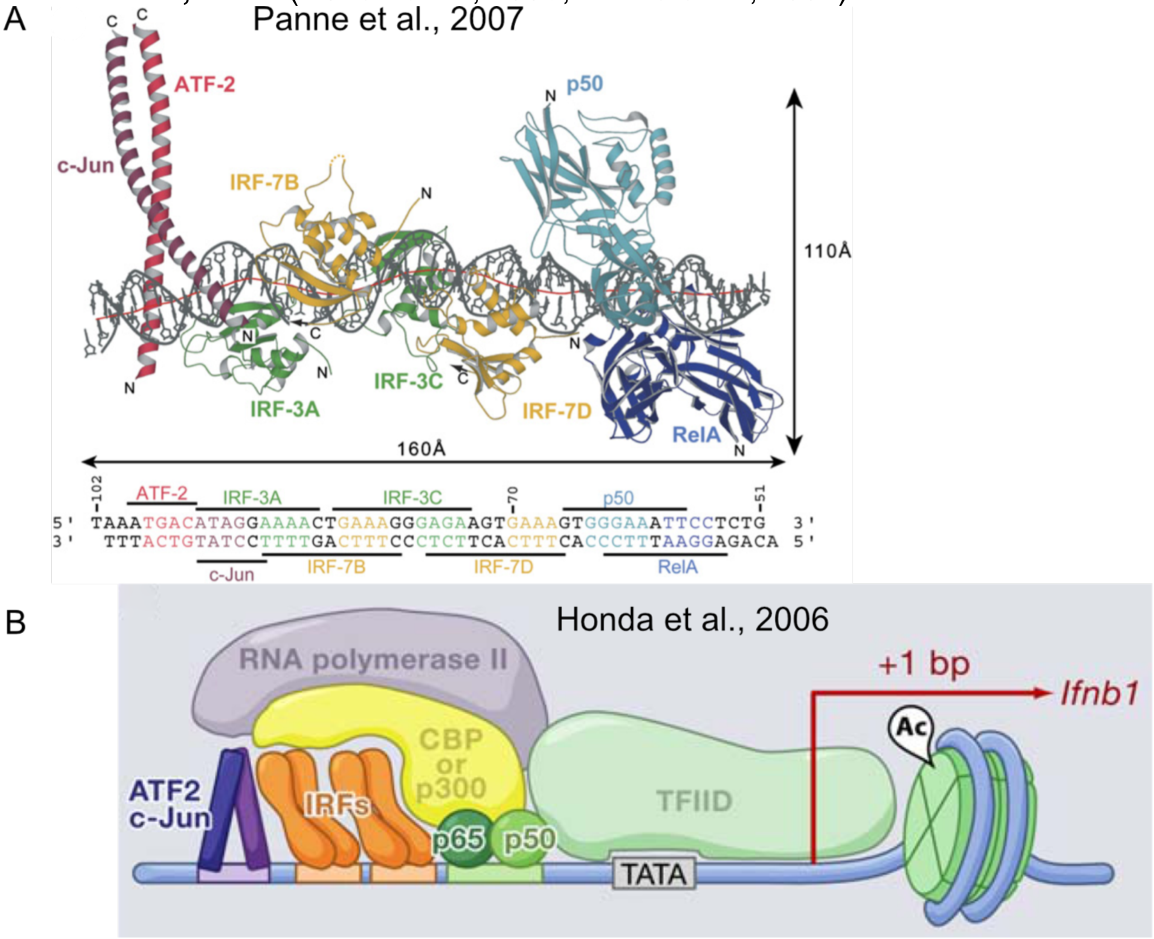
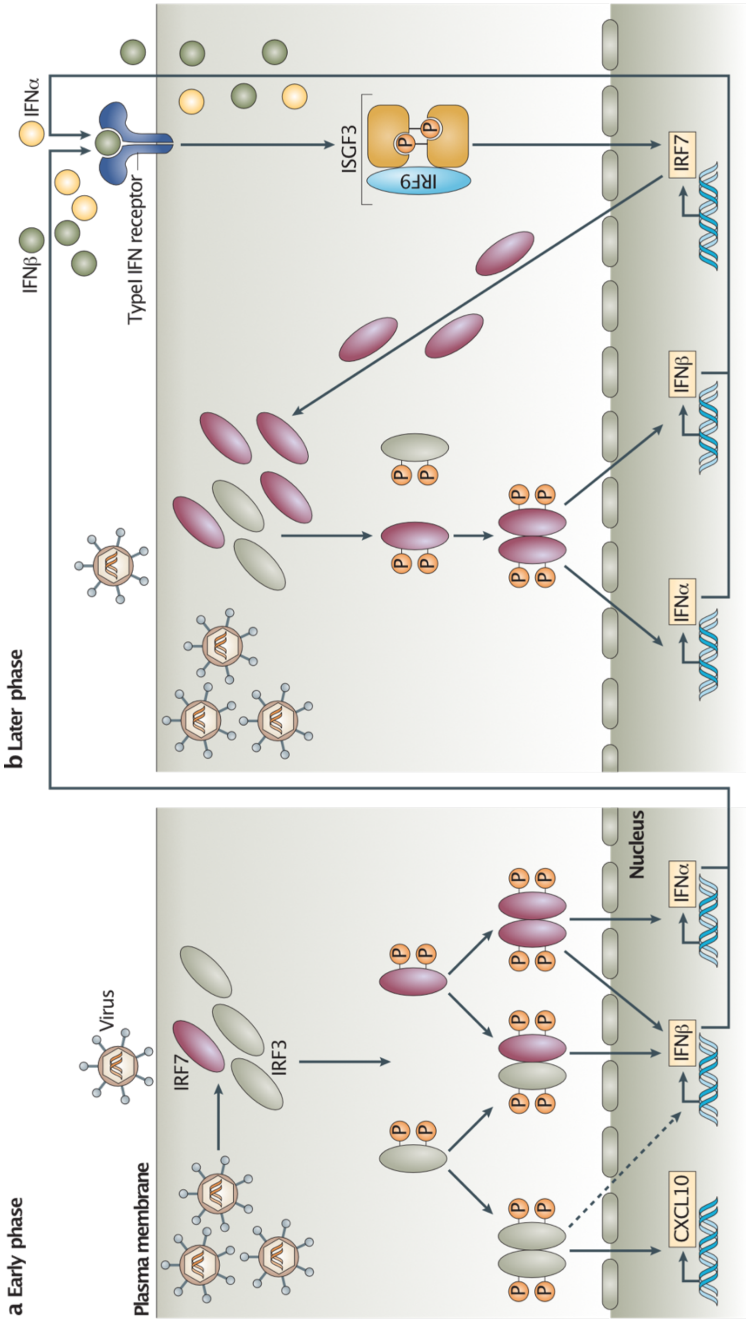


Figure 1.5 – Model of the IRF dependent IFN response

The IRF-dependent anti-viral response is described in two phases: **A)** The early phase is dominantly driven by activated IRF3. IFNβ production in this phase activates the late phase via autocrine signaling **B)** The later phase is driven by IRF7 and IRF3/7 heterodimers and drives IRF7 production and the IFNα genes. **Adapted from Honda and Taniguchi 2006** (Honda and Taniguchi, 2006)



From Honda et al., 2006

Figure 1.6 – The IFNα promoter contain conserved Viral Response Elements (VREs)

The IFNα Viral Response Elements (VRE) are a conserved cis-regulatory module that feature multiple IRF binding sites (B, C, D). **Adapted from Civas et al. 2006** (Civas et al., 2006)

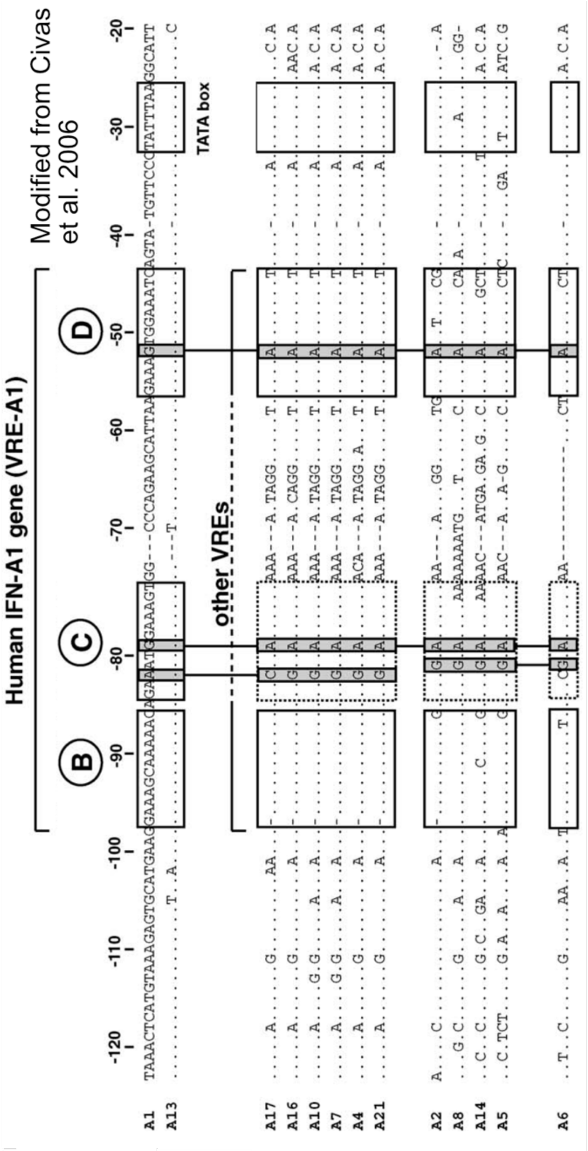

Modified from Civas et al. 2006
Human IFN-A1 gene (VRE-A1)
B
C
D
other VREs
TATA box
-120
-110
-100
-90
-80
-70
-60
-50
-40
-30
-20
A1
A13
A17
A16
A10
A7
A4
A21
A2
A8
A14
A5
A6
TAAACTCATGTAAAGAGTGCATGAAGGAAAGCAAAAACAGAAATGGAAAGTGG--CCCAGAAGCATTAAGAAAGTGGAAATCAGTA-TGTTCCCTATTTAAGGCATT

Figure 1.7 – Protein Binding Microarray (PBM) methodology

Single-stranded DNA oligo arrays (**A**) are made double-stranded using polymerase extension (**B**) Purified transcription factors tagged with GST are applied to the dsDNA PBM (**C**) After washing excess protein away, bound TFs are detected using fluorescently labeled antibodies (**D**) using a microarray scanner. **Adapted from Berger et al. 2009** (Berger and Bulyk, 2009)

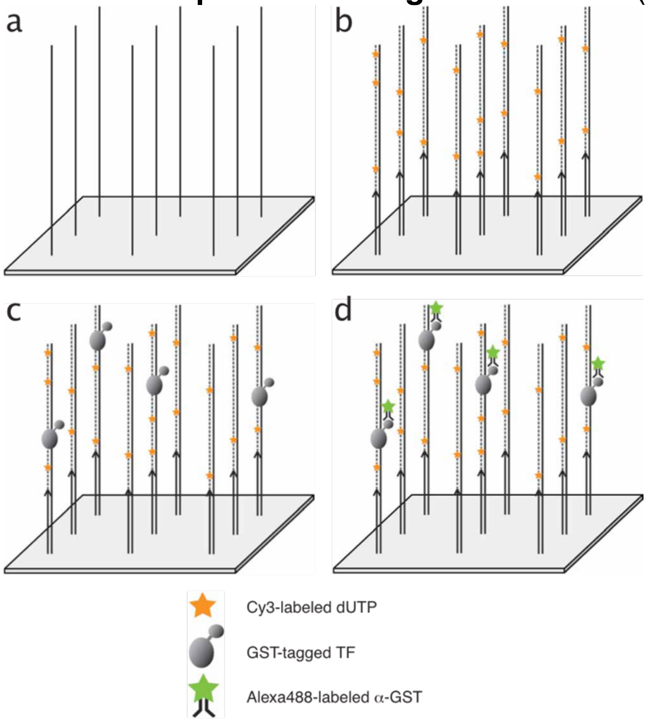
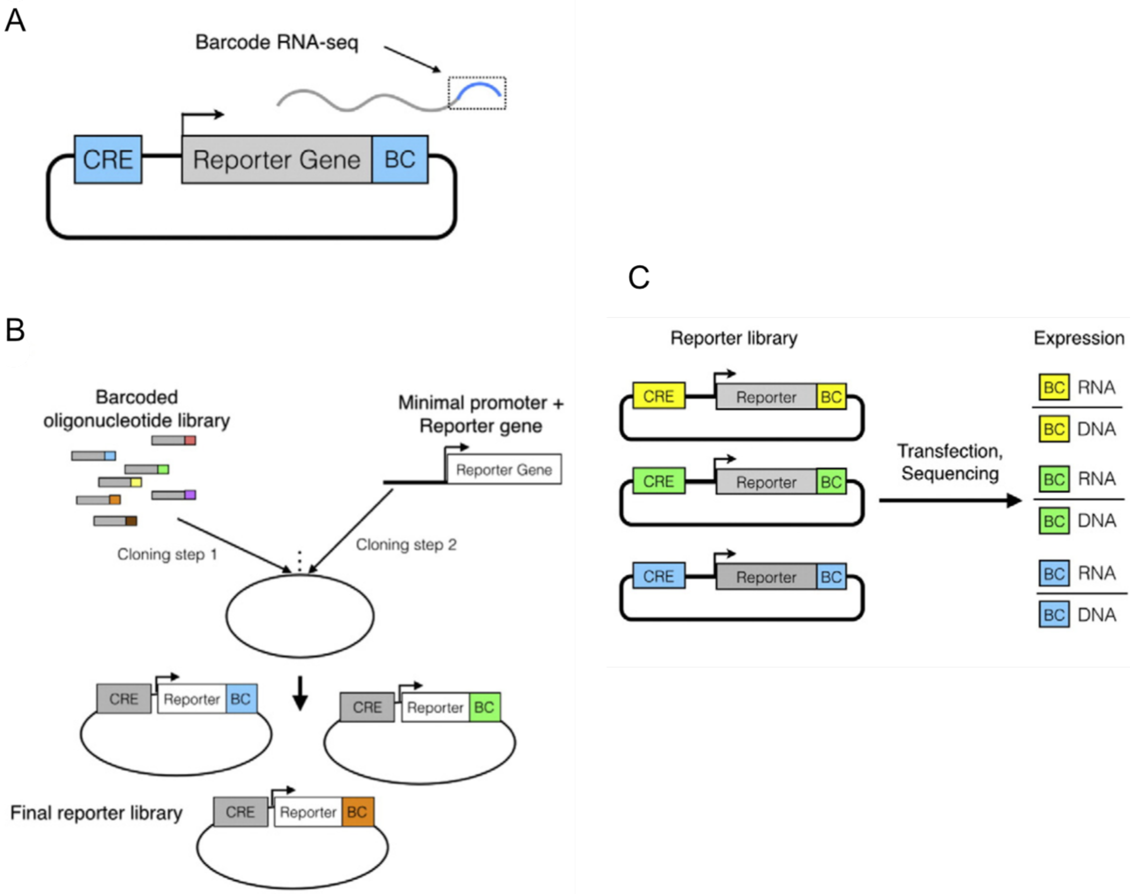


Figure 1.8 – Massively Parallel Reporter Assay (MPRA) methodology

(**A**) MPRAs use barcoded mRNA produced from barcode tagged CRE reporters. (**B**) MPRA libraries are synthesized as barcoded oligo pools and are cloned into an expression plasmid. (**C**) To quantify CRE activity, MPRAs use barcode counts from sequenced mRNA and normalize them by the plasmid library barcode counts. **Adapted from White, 2015** (White, 2015)



From White et al., 2015

2 CHAPTER TWO - MATERIALS & METHODS

2.1 Protein Binding Microarray (PBM) methods (Chapter 3)

2.1.1 Molecular Cloning and site-directed mutagenesis

2.1.1.1 Phosphomimetic IRFs

Constructs used to generate the phosphomimetic IRFs used in this work were graciously gifted to us by multiple labs. A constitutively active IRF3-5D phosphomimetic plasmid was gifted to us by Dr. Rongtuan Lin. Wild-type IRF5 plasmids were gifted by Dr. Betsy Barnes and Dr. Nancy Reich. A plasmid containing wild-type IRF7 isoform A was also gifted to us by the lab of Dr. Rongtuan Lin. All starting IRF3/5/7 constructs were subcloned into the Gateway system pDONR221 vector (ThermoFisher).

Phosphomimetic constructs for IRF5 and IRF7 (IRF5(4D) (Cheng et al., 2006); IRF7(8D) (Caillaud et al., 2005)) were made by site-directed mutagenesis using the QuikChange Lightning (Agilent) and NEB Q5 (New England Biolabs) site-directed mutagenesis kits following the manufacturer's instructions. The phosphomimetic IRF3(6D) construct (Chen et al., 2008; Lin et al., 1999) was codon optimized for *E. coli* expression and synthesized as an IDT-gBlock (Integrated DNA Technologies) with Gateway AttB sites then subsequently cloned into the Gateway vector system (ThermoFisher).

2.1.1.2 IRF DNA binding Domain Mutants

Using the phosphomimetic constructs described above, IRF5(K96S) and IRF7(S101K) DNA binding domain (DBD) mutations were made with In-vivo Assembly (IVA) site-directed mutagenesis (García-Nafría et al., 2016).

2.1.2 Protein Expression and Purification

2.1.2.1 IRF3/5/7 protein overexpression using *Escherichia coli*

Interferon regulatory factor phosphomimetic proteins were overexpressed using *E. coli.* Expression conditions were optimized on a per-protein basis to produce the greatest quantity of full-length protein and minimize truncated protein products as indicated by anti-GST western blot.

GST-IRF5(4D) (pDEST15) and GST-IRF7(8D) (pDEST15) were expressed using the OverExpress C41(DE3) *E. coli* strain (Lucigen) co-transformed with the pRare tRNA plasmid (Novagen). Transformed bacteria were propagated on Terrific Broth (TB) + 1% glucose + antibiotic plates. Protein expression was carried out in TB + 1% glucose + antibiotic with an initial outgrowth at 37°C up to OD 0.6 followed by 0.5 mM IPTG induction and expression at room temperature (~20°C) for 5 h. Addition of glucose to media (to suppress leaky protein expression) and co-transformation with pRare plasmid (to enhance translation) led to the highest yield of full-length protein.

Codon optimized GST-IRF3(6D)-6xHis(pDEST15) protein was expressed using BL21(DE3) E. coli co-transformed with the pLysS plasmid. Transformed bacteria were propagated on TB + antibiotic plates and protein expression was carried out in TB + antibiotic media with an initial outgrowth at 37°C up to an O/D of 0.6

followed by 0.2 mM IPTG induction and expression at room temperature for 5 h. Bacterial cultures were pelleted and stored at -80°C until lysis and purification.

2.1.2.2 FPLC affinity chromatography purification

IRF5 and IRF7 were purified with Glutathione-S-Transferase (GST) affinity chromatography using GSTrapFF columns (GE Healthcare) and an ÄKTApurifier-10 Fast Protein Liquid Chromatography (FPLC) device (GE Healthcare). The binding and elution buffers recommended in the GSTrapFF manual were used and supplemented with 1 mM PMSF serine protease inhibitor (Binding: PBS, pH 7.3 [140 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.3] + 1 mM PMSF; Elution: 50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0 + 1 mM PMSF). Sample was buffer exchanged into binding buffer + 20% glycerol using Amicon 30k MWCO filtration spin units (EMD-Millipore) then snap frozen and stored at -80°C. Protein concentration was determined using the Coomassie Plus Bradford assay (Pierce).

IRF3 was tandem affinity purified first using immobilized metal ion affinity chromatography (IMAC) (C-terminal-6xHis tag) followed by GST affinity chromatography (N-term-GST tag). Frozen cell pellets were resuspended in IMAC binding buffer supplemented with protease inhibitors (Sigma P8340), powdered lysozyme and Benzonase endonuclease (EMD Millipore). The resuspension was then lysed using an EmulsiFlex-C3 (Avestin, Inc.) homogenizer that was prechilled using recirculated ice-cold binding buffer. Multiple passes and a large buffer volume (100 - 200 ml) were required to reduce sample viscosity enough to allow thorough lysis by the EmulsiFlex-C3 and subsequent loading onto the chromatography columns. Lysed samples were clarified by centrifugation at 4°C at 15,000 RCF for 60 min and the supernatant was filtered using a 0.45 µm PES syringe filter (Fisher Scientific) before loading into the ÄKTApurifier-10 FPLC.

HIS purification was carried out using HisTrapFF columns (GE Healthcare) and the buffer conditions recommended by the manufacturer. (Binding buffer: 20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4; Elution buffer: 20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4) Buffers were supplemented with 1mM PMSF serine protease inhibitor. To prepare samples for affinity chromatography and prevent precipitation of IMAC purification eluates, samples were slowly buffer exchanged using Slide-a-Lyzer 3.5K MWCO dialysis cassettes (Thermo Scientific). Dialysis was carried out at 4°C first using IMAC binding buffer to slowly reduce the imidazole concentration of the sample. The sample was then dialyzed with GST-binding buffer overnight using a flow regulated beaker system allowing for the dropwise addition of target buffer over many hours. The next day the GST-binding buffer was refreshed and dialysis continued while preparing for GST affinity chromatography. The sample was then diluted in freshly made GST-binding buffer for automated FPLC sample loading. GST-tag purification was carried out as described for IRF5 and IRF7.

IRF5(K96S) and IRF7(S101K) DNA binding domain (DBD) mutations were made with In-vivo Assembly (IVA) site-directed mutagenesis (García-Nafría et al., 2016). IRF-DBD mutant proteins were expressed and purified as described above.

2.1.3 Protein Binding Microarrays (PBM)

2.1.3.1 PBM Design

Our IRF-specific PBM design included both synthetic and genome-derived IRF binding sites. *Synthetic probes:* Microarray probes with synthetic sites were based on 108 seed IRF binding sites that were each 20-bp long (for 2-bp half-site spacers), and 54 seed sites that were 21-bp (for 3-bp half-site spacer). Seed IRF sequences were within constant flanking DNA sequence (Figure 1A). For each seed sequence, we included all single-nucleotide variants (SNVs) across the 20-bp or 21-bp long site, for a total collection of 10,044 synthetic IRF sites. For each unique IRF site, 6 replicate probes were included in each orientation (12 replicates per unique site). The 162 seed sequences were chosen based on IRF binding sites from the literature, available HT-SELEX datasets (Jolma et al., 2013), and preliminary PBM experiments. Seed sequences were selected to capture a range of binding affinity, and to include alternate core sequences (i.e., alternates to the canonical 5’-GAAA-3’ core). This seed+SNV design allows us to directly compare IRF-specific DNA binding preferences across the IRF-binding sites in a straightforward manner, and to generate position weight matrices (PWMs) with relatively small number of sequences (see below).

Genome-derived probes: Putative IRF binding sites from the promoters of type I IFN genes, and from published cis-regulatory elements for other genes (e.g., CXCL10) were extracted from the genomes as 21-bp genomic fragments and centered on the microarray probes in an identical manner to the synthetic probes. The proximal promoters of the type-1 interferons and other cytokines were scanned using BioPython (v1.68; bio.motifs, bio.SeqUtils.lcc modules; (Cock et al., 2009; Mangalam, 2002)) and degenerate IRF consensus motifs with a 2-bp or 3-bp spacer length (figure 2.1). Sequences with a low stringency PWM score above zero and a sequence complexity score above 0.5 were included in the array. These selection criteria were empirically selected to remove long poly-A runs present in the IFN promoters while still capturing potential IRF binding sites. To reduce redundant genomic probes, sites within 1 bp of each other were removed retaining probes with the PWM hit closest to the end of the PBM probe. To control for potential IRF binding site shifts, trimmed probes were generated by replacing the flanking regions of the IRF-PWM with a low/moderate affinity constant sequence as determined in preliminary array designs. Genomic loci and full PBM probe sets are available upon request and published online (Andrilenas et al., 2018) https://doi.org/10.1093/nar/gky002.

2.1.3.2 PBM Experiments

PBM experiments were performed using custom-designed microarrays (Agilent Technologies Inc. AMADID 084215, 4x180K format). Microarrays were double-stranded as previously described (Berger et al., 2006; Berger and Bulyk, 2009). Wash steps were carried out in coplin jars on an orbital shaker at 125 rpm. Double stranded microarrays were first pre-wetted in PBS containing 0.01% Triton X-100 for five min, rinsed in a PBS bath, and then blocked with 2% milk in PBS for 1 hour. Following the blocking step, arrays were washed in PBS containing 0.1% Tween-20 for 5 min, then in PBS containing 0.01% Triton X-100 for 2 min and finally briefly rinsed in a PBS bath. Arrays were then incubated with the protein sample(s) for one hour in a binding reaction containing: 2% milk with 10 mM Tris, pH 7.5; 50 mM NaCl; 2 mM DTT; 0.2 mg/ml BSA; 0.02% Triton X-100; and 0.4 mg/ml salmon testes DNA (Sigma D7656). See Table 2.1 for PBM protein concentrations and conditions. After protein incubation, microarrays were washed with PBS containing 0.5% Tween-20 for 3 min, then in PBS containing 0.01% Triton X-100 for 2 min followed by a brief PBS rinse. Microarrays were then incubated with 20 µg/ml of Alexa Fluor-488 conjugated Anti-GST antibody (LifeTech, Cat# A-11131) in 2% milk in PBS for 20 min. Excess antibody was removed by washing with PBS containing 0.05% Tween-20 for 3 min, then PBS for 2 min.

2.1.3.3 PBM data acquisition and analysis

Microarrays were scanned with a GenePix 4400A scanner and fluorescence was quantified using GenePix Pro 7.2. Exported data were normalized using MicroArray LINEar Regression (Berger et al., 2006). Microarray probe sequences and fluorescence values from each experiment are available upon request and published online (Andrilenas et al, 2018). IRF dimers exhibit an orientation-specific bias in our PBM experiments; therefore, data from probes in a single orientation (i.e., '_o2' probes; available upon request and online https://doi.org/10.1093/nar/gky002 (Andrilenas et al., 2018)) was used in our final analysis. However, all results were observed for probes in both orientations.

PBM experiments for the IRF3/5/7 phosphomimetic dimers were performed at four protein concentrations for each IRF dimer (Table 2.1). A saturation-binding curve was fit independently to the fluorescence values for each probe sequence:

$$F = \frac{F_{max} * [P]}{K_d + [P]} \tag{2.1}$$

F is probe fluorescence, **F$_{max}$** is max fluorescence, **[P]** is applied protein concentration, **K$_d$** is dissociation constant. We previously showed that this approach can accurately estimate relative binding affinities over a wide affinity range (Siggers et al., 2011). Curve fitting was performed in the statistical package R using the *optim* function (method - "Brent", **F$_{max}$** – highest fluorescence value on PBM at highest protein concentration) with the cost function:

$$Cost = \sum_{i=1}^{N} \left(1 - \frac{log\left(\frac{F_{max} * [P]^{i}}{[P]^{i} + Kd^{i}}\right)}{log\left(F_{obs}^{i}\right)} \right)^{2} \qquad (2.2)$$

K$_d$ values were determined for each PBM probe, the median **K$_d$** across the six replicate probes was then reported for each unique DNA sequence on the PBM. We refer to these resulting binding constants as **K$_{PBM}$** to highlight that these are PBM-derived estimates of relative binding constants. K$_{PBM}$ values were virtually identical (Pearson correlation R = 0.98) when a single value was determined by fitting simultaneously on fluorescence measurements from all replicate probes (i.e., 4x6=24 fluorescence values). Mutant IRF experiments were performed at a single concentration (Table 2.1). For each DNA sequence, the median fluorescence intensity, over 6 replicate probe measurements, is used to quantify the binding of the protein to the DNA. We found that K$_{PBM}$ values determined from PBMs done at multiple concentrations (as described above) was approximated well by PBM experiments performed at a concentration ~100-200 nM. For all PBM experiments, z-scores were determined for the log(F) or log(K$_{PBM}$) values using the mean (μ) and variance (σ2) of the values for the randomly selected background probes: $z_i = (\log(F_i) - \mu)/\sigma$. Z-scores provide an internally consistent way to quantify the specificity above background for measurements in each experiment.

2.1.3.4 SNV method for constructing position frequency matrices (PFMs)

PFMs can be constructed using the PBM z-score values of a single seed sequence (i.e., starting sequence), and the 3x21 associated SNV sequences. In the low-protein limit (i.e., when $[P] << K_d$), z-scores are related to $\log(F) \sim \log(K_d) \sim E$, where *E* is the binding free energy. Therefore, starting from an individual seed sequence, we compute the relative base preferences for base *k* at position *i* as a probability based on the Boltzmann distribution:

$$P_{ik} = \frac{e^{\beta z_{ik}}}{\sum_{k=1}^{4} e^{\beta z_{ik}}} \qquad (2.3)$$

Z_{ik} – z-score for this particular seed sequence with SNV *k* at position *i*. **β** - a normalization factor chosen to maximize the correlation between the PFM scores and the PBM-measured $\log(K_{PBM})$ values. We found that **β**=1 worked well for all our experiments. Using this approach, we can generate a PWM from a single seed and its complement of SNV probes. To generate a representative PFM for a PBM experiment we determine individual PFMs using the 15 top-scoring seed sequences and average the fifteen individual p_{ik} values to determine an average PFM for the experiment. Logos were then generated using the ENOLOGOS webserver (Workman et al., 2005), with background frequencies set to equally probable. To identify base positions that are discriminatory for a single dimer, we systematically analyze our PBM data to find single-base changes that abrogate binding of one IRF dimer but not another. Specifically, to identify base positions that discriminate between two dimers, we analyze the binding to all pairs of DNA sequences that differ by a single base (i.e., SNV pairs). SNV pairs are identified where one IRF dimer is bound with high-affinity to both sequences (z-score > 8.0 for both probes, and z-score difference between probes < 3.0) while the other IRF dimer is bound with high-affinity to one probe (z-score 8.0) but with low affinity to the other probe (z-score < 5.0, and z-score difference between probes > 5.0). Identifying all such SNVs reveals base positions and variants that provide strong discrimination between IRF dimers.

2.1.4 Electro-mobility shift assays (EMSAs)

2.1.4.1 EMSA probe generation

IRF binding mode EMSAs (Chapter 3 - Figure 3.5C) used complementary oligo annealing to form double-stranded EMSA probes. Briefly, single stranded DNA oligos were annealed at 100 μM concentration in TE buffer + 50 mM NaCl (10mM Tris, 1mM EDTA, 50mM NaCl, pH 8.0). Probes were denatured at 95°C for 2 min in a dry-heating block which was allowed to cool to room temperature. See Table 2.2 for EMSA oligo sequences.

IRF PBM probe EMSAs (Chapter 3 - Figure 3.3) used oligo extension to generate 60-bp dsDNA probes. Briefly, probes were double-stranded using BST polymerase (New England Biolabs). Probe and primer were slowly annealed from 95°C to 63°C using a thermocycler with a 0.1°C/sec cooling rate in buffer containing 8 μM probe ssDNA oligo, 8 μM extension primer, 1x ThermoPol buffer (New England Biolabs), 1.6 mM dNTPs (New England Biolabs). Four units of BST polymerase diluted in 1x ThermoPol buffer were pre-heated to 63°C and quickly added to the annealed mixture. The isothermal double stranding reaction was held at 63°C for 1.5 h, then double-stranded probes were purified using a MinElute PCR purification kit according to manufacturer's instructions (Qiagen).

2.1.4.2 DNA probe radiolabeling

Purified dsDNA probes were radiolabeled with [γ-^{32}P]-adenosine triphosphate (PerkinElmer) using T4 Polynucleotide Kinase (New England Biolabs) following manufacturer's protocol. Radiolabeled probes were purified using a QIAquick nucleotide removal kit (Qiagen) according to the manufacturer's instructions. Radiolabeling and purification yield were assumed to be 100 percent efficient for EMSA probe concentration calculations.

2.1.4.3 EMSA binding reactions, electrophoresis and imaging

Experiments were carried out as described (Hellman, Fried, 2007). Binding reactions were carried out in 20 *μ*l volumes containing: 1 nM P-32 labeled DNA probe and PBM buffer with nonspecific DNA competitor (10 mM Tris pH 7.5; 50 mM NaCl; 2 mM DTT; 0.08% Triton-X100; 50 ng/*μ*l poly(dI:dC) (LI-COR Biosciences); 0.005 ug/*μ*l salmon sperm DNA (LI-COR Biosciences)). To prevent adsorption to tubes and pipette tips, protein samples were diluted to appropriate concentrations in reaction buffer + 1 mg/ml BSA (New England Biolabs). Reactions were incubated at room temperature for 45 min. Before loading into gels, 1 *μ*l of 50% glycerol and 1 *μ*l Orange Loading Dye were added to each reaction (LI-COR Biosciences). Reactions were resolved in 6% polyacrylamide gels (29:1 cross-linking) with 0.5x TBE running buffer at 10 V/cm in a 4°C water bath until the loading dye front reached the bottom of the gel (~2 - 3 h depending on gel length). Before drying under vacuum, EMSA gels were fixed for 30 min using a 20% Methanol, 10% acetic acid solution then rinsed in ddH2O for 15 min and transferred to Whatman filter paper. Autoradiography was performed using a BAS storage phosphor screen (GE Healthcare). After overnight exposure, the phosphor screen was scanned using a Typhoon-Trio scanner (GE Healthcare). Typhoon square root space *.GEL* files were linearized using the ImageJ Linearize GelData plugin. To improve band visibility, the brightness of the linearized TIFF files was decreased by 25 units using Adobe Photoshop CS6.

2.1.5 Reporter Assays

IRF3(6D), IRF5(4D) and IRF7(8D) were cloned into the N-terminal His-tagged protein mammalian expression plasmids (pDest26) (LifeTech). HEK293T cells were cultured in DMEM (Gibco 11965092) + 10% FBS (Gibco 26140079). Cells were plated in tissue culture treated 96-well plates seeded at 12 000 cells per well and allowed to adhere overnight. PEI:DNA complexation reactions were carried out in 500 *μ*l of serum free media and cells were transfected using polyethylenimine (PEI) (Polysciences, Inc.) at a ratio of 2:1 (PEI:DNA). Each 96-well plate well received 10 *μ*l of transfection mixture containing: 12.5 ng of Tk-Luciferase transfection normalization plasmid (pGL4.54); 10 ng of E1α-eGFP carrier DNA; 12.5 ng of reporter plasmid (pNL3.1); and 1.25 ng of His-tagged protein expression plasmid (pDest26) or GFP plasmid (E1α-eGFP) in background-expression controls. The key difference between an experimental condition and a GFP control condition, is whether a phosphomimetic IRF is being over expressed versus GFP. Every IRF-binding site reporter plasmid (pNL3.1: i3, i5, i7, C-2, C-3, i7-2, empty) was tested with exogenous protein expression (IRF3, IRF5, IRF7, GFP), with the GFP conditions as controls. Additionally, every well received the luciferase transfection control plasmid (pGL4.54). Cells were incubated with transfection reagent overnight and then cell culture media was changed. Cells were lysed and assayed 24 h after transfection using the Nano-Glo Dual Luciferase reporter assay system (Promega). Dual-luciferase signal was quantified using a VICTOR-3 plate reader (PerkinElmer). Reporter assay conditions had at least three biological replicates and at least three technical replicates per biological replicate

2.1.5.1 Normalization of reporter assay data

Nano Luciferase reporter plasmid signal was normalized to the constitutive luciferase signal (i.e. signal from pGL4.54 plasmid, transfection normalization) for each transfected well:

For each well: $$Norm_i = \frac{Experimental\ Reporter_i\ [NanoLuc, pNL3.1]}{Constitutive\ transfection\ control_i\ [Luc,\ pGL4.54]}$$

Where $Norm_i$ is the transfection normalized value for a given well.

Fold-induction values for each Protein X Reporter combination were calculated relative to the background-expression signal for each reporter plasmid condition:

For each condition well: $$gfp\_norm_i = \frac{Norm_i[Reporter\ X\ IRF]}{mean(Norm_{i \to x}[Reporter\ X\ GFP])}$$

Where $Norm_{i \to x}$ represents the transfection normalized data from all biological and technical replicates for a given Reporter x GFP condition.

Descriptive statistics (mean, standard deviation, etc.) and plots of all intermediate steps for all experimental conditions are provided in Chapter 3: luciferase (Figure 3.10; Table 3.1), nano luciferase (Figure 3.11, Table 3.2), transfection normalized (Norm; Figure 3.12, Table 3.3), GFP normalized (gfp_norm; Figure 3.13, Table 3.4).

We chose to normalize reporter assay data to the GFP condition that has no IRF over expression to account for potential variation in background activity of the reporter plasmids. We also used GFP over expression to account for the impact of transcriptional and translational burden incurred by cells over expressing the IRF3/5/7 constructs. When examining the raw transfection normalized data, we find that our data still suggests affinity independent transcriptional regulation by IRF3/5/7.

Computations were performed using the Python programming language and the Pandas module (see below). Bar plots of reporter assay data in Chapter 3 (Figure 3.9) are scaled to the i7-2 reporter for each protein condition. Scaling to the i7-2 reporter values was an aesthetic choice that allowed for the use of a single set of axes while not altering the relative relationships between data points in a given protein condition. All intermediate plots and data are in Chapter 3 (Figures 3.10 – 3.14; Table 3.1 – 3.5)

Stepping through the normalization process:

1. Transfection Normalization: Divide each nano-luciferase data point by its corresponding luciferase transfection control. (*Norm*)

2. GFP Normalization: For each transfection normalized Norm data point, divide each Norm value by the mean of the corresponding Norm[reporter x GFP] data. (*gfp_norm*)

3. I7-2 reporter scaling: Divide the GFP normalized values (*gfp_norm*) by the mean of the I7-2 data for a given protein condition.

Python Code:

```
## %% Setup ##
import pandas as pd

## %% Data Import ##
data_file = 'reporter_data_export_sep23.csv'
df = pd.read_csv(data_file)

## %% Normalize ##
# Divide each nanoluc value by its luciferase value
df['norm'] = df['nanoluc'] / df['luc']


# Group data by experiment date, then reporter condition.
Divide each 'norm' data point
# by the mean of the GFP control for that respective date
and reporter condition.
df['gfp_norm'] = df.groupby(['Date', 'Reporter']).apply(
    lambda x: x[['norm']] / x.loc[x['Protein'] == 'GFP',
['norm']].mean())

## %% Scale Data ##
# Group data by protein. Divide gfp_norm values from each
protein group
# by the mean of the respective i7-2 gfp_norm values
df['i72_scale'] = df.groupby(['Protein']).apply(
    lambda x: x[['gfp_norm']] / x.loc[x['Reporter'] == 'i7-
2', ['gfp_norm']].mean())
```

2.1.6 Western Blots

2.1.6.1 Phosphomimetic IRF overexpression western blots

IRF overexpression conditions from reporter assays were scaled up to generate whole cell lysates from 10cm culture plates. Plates were PEI transfected as described above with 7.2 *μ*g of plasmid DNA consisting of: pGL4.54 TK-Luciferase [2.48 *μ*g]; E1α-eGFP [1.99 *μ*g]; pDEST26 IRF3/5/7 or additional E1α-eGFP [0.25 *μ*g]; pGEM3zf(-) carrier DNA [2.48 *μ*g], these quantities are scaled up from 96-well plates assuming a 20 *μ*g typical 10cm plate transfection. Carrier DNA was used in place of any IRF-specific pNL3.1 NanoLuc reporter plasmids to avoid the effect of differential transcriptional output from the reporter plasmids. Cells were harvested 24-h after transfection and lysed on ice in RIPA buffer + 1:100 protease inhibitor cocktail (50 mM Tris HCl, pH 8.0; 150 mM NaCl; 1% Tergitol (NP-40); 0.5% sodium deoxycholate; 0.1% SDS; Sigma P8340). Released genomic DNA was digested with 0.5 *μ*L Benzonase nuclease (Sigma E1014). Total protein content was measured using the Coomassie Plus (Bradford) Assay kit (Pierce) with assay compatible dilutions of RIPA buffer in BSA standards (1:50) and samples. SDS-polyacrylamide gels were run with 30 *μ*g of total protein per lane as well as 25 ng of purified GST tagged IRF protein as a positive control. Electrophoresed protein was then transferred to Immobilon-FL PVDF membranes (Millipore-Sigma). Membranes were blocked in Tris Buffered Saline + 1% Tween-20 (TBST; 20 mM Tris, 150 mM NaCl) with 5% w/v non-fat milk. Primary antibodies were incubated with membranes overnight at 4°C in TBST without milk at the following dilutions: 1:5,000 mouse-anti-IRF-3 (Santa Cruz Bio, sc-33641); 1:5,000 mouse-anti-IRF-5 (Santa Cruz Bio., sc-390364); 1:4,000 mouse-anti-IRF7 (Santa Cruz Bio., sc-74472); 1:5,000 rabbit-anti-PARP1 (Santa Cruz Bio., sc-7150). Membranes were washed four times in TBST and incubated with 1:10,000 fluorescent secondary antibodies in TBST+5% milk for one hour (Invitrogen AlexaFluor 488-goat-anti-mouse, A11001; LifeTech. Cy5-goat-anti-rabbit, A10523). Membranes were washed again and imaged on a Typhoon Trio imager (GE Healthcare). Dual-color composite images were created from linearized GEL files (see EMSA methods above) and brightness and contrast were adjusted to increase signal-to-background ratio in ImageJ.

2.2 MPRA Methods

Massively Parallel Reporter Assay experimental design and procedures were based on a protocol and personal communications with the lab of Dr. Barack Cohen at Washington University in St. Louis, Missouri.

2.2.1.1 MPRA Library design

A complex oligo pool was ordered from Agilent. 15,000 sequences were custom ordered with sequences divided amongst three MPRA library experiments. The Interferon Regulatory Factor MPRA library consists of 6,131 unique barcodes with 10 barcodes per regulatory element. Ten barcodes were considered sufficient for cells with transfection efficiencies of 10 - 40% (personal communications with Cohen lab).

The IRF MPRA library was designed to include multiple experiments derived from IRF PBM data. IRF MPRA regulatory elements were designed using the type-I interferon Viral Response Element structure (VRE) where two IRF dimer sites are adjacent to each other with a short spacer (See Chapter 4 - for more information on experimental features). Redundant regulatory elements were removed before barcode assignment to reduce the sequence space required by the library. MPRA cloning features were designed in coordination with Jessica Keenan and Ashley Penvose.

2.2.1.2 MPRA vector engineering

MPRA libraries were cloned into a modified pNL3.1 (Promega) vector, hereafter referred to as pNL-MPRA. The pNL3.1 vector was modified for use with a GoldenGate cloning strategy that utilizes two Type-IIs restriction enzymes frequently used in complex GoldenGate assembly: BsaI and BbsI (Andreou and Nakayama, 2018; Engler and Marillonnet, 2013). The ccdB and CamR cassette from the Gateway cloning vector pDEST26 was added to the pNL3.1 vector to reduce background colonies that lack the MPRA library insert. The vector modification process required mutating BsaI and BbsI sites present in the pNL3.1 vector backbone and ccdB gene. Creating the pNL-MPRA vector also involved removing the NanoLuc gene and minimal promoter and adapting the multicloning site to use BsaI sites for library insertion. The IRF-MPRA design uses alternative restriction sites (Acc65I and XbaI) in place of the BsaI sites given the similarity of the BsaI site (NNNNINGAGACC) to an alternate IRF binding site present in the MPRA oligo pool (GAGACCGAGA).

2.2.1.3 MPRA molecular cloning

To prepare for MPRA library cloning, pNL-MPRA-ccdB vector was digested with the restriction enzymes Acc65I and XbaI (New England Bio) using NEB 3.1 restriction buffer which allowed for the highest activity of both enzymes (Acc65I: 100%; XbaI 75%). Vector DNA was simultaneously dephosphorylated using recombinant Shrimp Alkaline Phosphatase (rSAP; New England Bio) to prevent re-ligation of the vector. Digests were purified using a PCR purification column (Epoch Life Science).

The 10 pmol Agilent OLS ssDNA library was resuspended in 100 μl of nuclease free TE buffer pH 8.0 (Invitrogen) at an end concentration of 0.1μM. The IRF-MPRA library was amplified from the total Agilent OLS pool using IRF-MPRA specific primers in a low cycle PCR reaction (Table 2.3A) using primers in Table 2.4. Library PCRs resulted in poor DNA yield after PCR DNA cleanup and multiple reactions were pooled and purified to reach sufficient DNA concentration for downstream handling. Purified dsDNA MPRA library was digested using the restriction enzymes Acc65I and XbaI and then enzymes were heat inactivated for 20min at 65°C. Small scale regulatory element insert ligations were performed using unpurified restriction reaction and purified phosphatase treated vector. Molar ratios from 1:3 to 1:5 (Vector:Insert) were tested in small 20μl ligations. Ligation reactions were performed using T4 DNA ligase in NEB CutSmart buffer supplemented with 1mM ATP (New England Bio) and incubated for 1 hour at 16°C and then 10 min at 23°C followed by heat inactivation at 65°C for 20 min. Small scale *E. coli* transformations were performed using 10μl of NEB DH5α chemically competent cells with 2μl of ligation reaction. Transformation reactions were suspended in 100μl of SOC media and shaken at 37°C for 15min and then 100μl was plated on Nunc OmniTray (ThermoFisher) microplate format agar plates with carbenicillin. Colony counts were estimated using the OpenCFU (v3.9.0 for Windows) software and plate images acquired using an Epson flatbed scanner (Geissmann, 2013). The Epson scanner was used because of its front illumination imaging sensor. A ligation vector:insert molar ratio of 1:4 had the greatest colony count per plate.

The number of colonies required to be 99% confident that all sequences in the IRF-MPRA library were captured was estimated using the geometric distribution in equation 2.4, where p is the probability of picking a single oligo from the IRF-MPRA pool, assuming a 50% error rate from Agilent, and n is the number of colonies to collect (eq. 2.4).

$$P(picking\ all\ correct\ colonies) = 1 - (1 - p)^{(n+1)} \tag{2.4}$$

$$p = {}^{1}\!/\!_{(2 * number\ of\ barcodes\ in\ library)} \tag{2.5}$$

Equation 2.4 can then be rearranged to solve for the number of colonies needed to have a 99% probability that all barcodes are represented in the culture:

$$Number\ of\ colonies\ = \frac{log(1 - \ 0.99)}{log(1 - p)\ - \ 1} \tag{2.6}$$

For the IRF-MPRA library with 3,161 unique barcodes we must select at least 29,110 colonies to have a 99% probability that all barcodes are represented in the library. At a colony count of 2,000 colonies per plate, approximately 15 plates will be required to have sufficient colony counts.

Ligation and transformation reactions were scaled up to multiple 50μl transformations using the estimated colony counts found during ligation tests. A vector only, no insert control reaction was also performed. Transformations were performed with a 30 min. Incubation on ice followed by a thermocycler transformation program (Pre-chill at 4°C; 30 sec at 42°C; return to 4°C; then move to ice for 5 min). 500 μl of 37°C SOC media was added to each 50 μl reaction and cultures were shaken for 15 min at 37°C, after recovery, 100 μl of culture was spread across each OmniTray plate with a bar spreader. Plates were incubated at 37°C for 12 - 18 h. Four randomly selected plates were counted as described above, using OpenCFU, to estimate colony counts. Colonies were scraped from plates using an extra-large metal dry-reagent spatula (Fisher Scientific 14-373-25). This tool was used as opposed to a plastic cell scraper or glass slide, as it could be flamed between plates and easily manipulated with gloved hands. Scraped colonies were resuspended in a small number of 50 ml conical tubes, each containing 15 mL of Luria Broth (LB). Colony suspensions were pooled in a 2 L baffled flask and additional media was added to bring the final culture volume to 20 ml of LB per plate. An appropriate volume of 1000x carbenicillin antibiotic was added to the large culture as well as 5 μl of Antifoam 204 (Sigma Aldrich). Initial small-scale pilot liquid cultures at 37°C did not produce turbid cultures after 24 h. We found that reducing the culture temperature to 30°C resulted in expected growth. The MPRA library culture was shaken at 250 RPM at 30°C for 5 h. The culture optical density at 595 nm reached approximately 1.3 units. The cell culture was then pelleted in 50 ml conical tubes at ~1500 RCF for 15 min and frozen at -80°C. Plasmid DNA from the large-scale MPRA-library cell pellets was isolated using a Plasmid Plus Maxi Kit (Qiagen) following the manufacturer's high-yield purification protocol. Qiagen only recommends a maximum of 100 ml of overnight culture, given the decreased culture time and culture temperature, a pellet from 400 ml of media did not overload the column.

2.2.1.4 MPRA library cloning quality assurance

MPRA library step 1 was sequenced for quality assurance using the Massachusetts General Hospital Center for Computational & Integrative Biology DNA core (MGH CCIB) CRISPR amplicon sequencing service. A 200 bp amplicon was generated for sequencing using PCR with primers flanking the MPRA library insert (see primers in Table 2.3). Sequencing data was analyzed using custom python scripts (Python version 3.5, Biopython 1.73, Pandas 0.24.2), as well as BWA (Li, Durbin, 2010) and BBtools(Bushnell et al., 2017)

Figure 2.1- Degenerate Position Frequency Matrices.

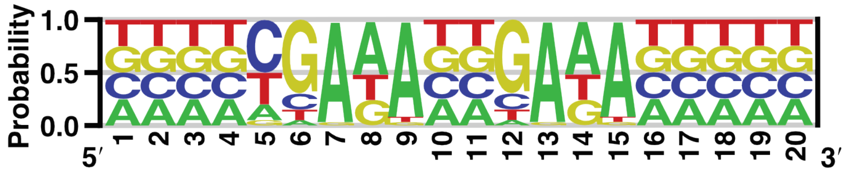
Degenerate Position Frequency Matrices used to extract potential IRF3/5/7 sites from the human type-1 interferon promoters for inclusion in the IRF protein binding microarray design. PFMs are a count-based model of DNA sequence motifs; here, each PFM has been visualized as a sequence logo. The 2-bp PFM (top) has a two base pair spacer between the core IRF GAAA sites, the 3bp (bottom) has a 3 base pair spacer.

Figure 2.1

Degenerate PFM (2-bp)

PO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	25	25	25	25	15	5	95	50	90	25	25	5	95	50	90	25	25	25	25	25
C	25	25	25	25	50	15	1	2	2	25	25	15	1	2	2	25	25	25	25	25
G	25	25	25	25	5	70	3	24	3	25	25	70	3	24	3	25	25	25	25	25
T	25	25	25	25	30	10	1	24	5	25	25	10	1	24	5	25	25	25	25	25

Degenerate PFM (2bp)



Degenerate PFM (3-bp)

PO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	25	25	25	25	15	5	95	50	90	25	25	25	5	95	50	90	25	25	25	25	25
C	25	25	25	25	50	15	1	2	2	25	25	25	15	1	2	2	25	25	25	25	25
G	25	25	25	25	5	70	3	24	3	25	25	25	70	3	24	3	25	25	25	25	25
T	25	25	25	25	30	10	1	24	5	25	25	25	10	1	24	5	25	25	25	25	25

Degenerate PFM (3bp)

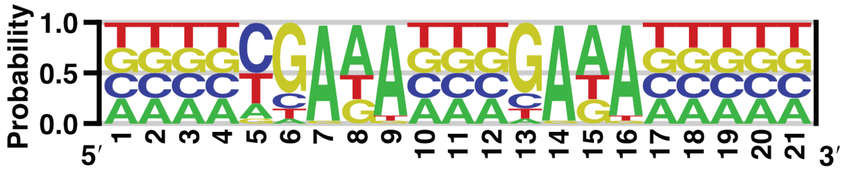


Table 2.1 - Protein Binding Microarray experimental conditions used in PBM experiments.

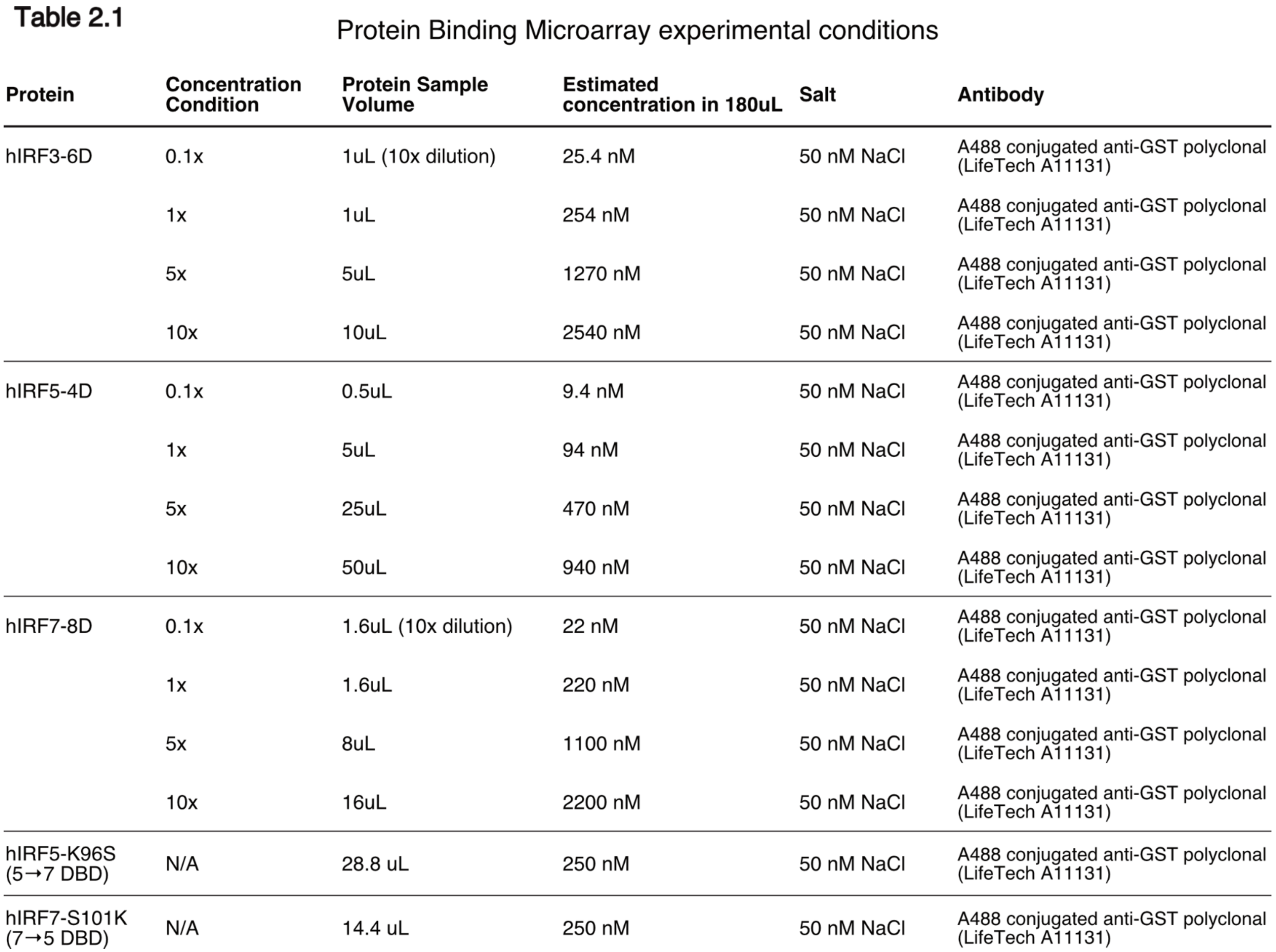


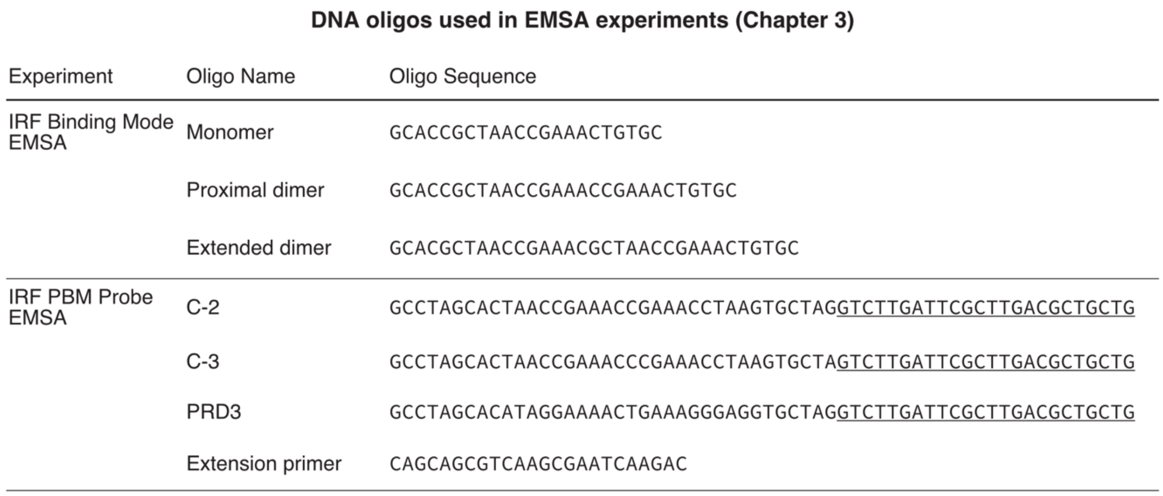
Table 2.1

Protein Binding Microarray experimental conditions

Protein	Concentration Condition	Protein Sample Volume	Estimated concentration in 180uL	Salt	Antibody
hIRF3-6D	0.1x	1uL (10x dilution)	25.4 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	1x	1uL	254 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	5x	5uL	1270 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	10x	10uL	2540 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
hIRF5-4D	0.1x	0.5uL	9.4 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	1x	5uL	94 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	5x	25uL	470 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	10x	50uL	940 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
hIRF7-8D	0.1x	1.6uL (10x dilution)	22 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	1x	1.6uL	220 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	5x	8uL	1100 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
	10x	16uL	2200 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
hIRF5-K96S (5→7 DBD)	N/A	28.8 uL	250 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)
hIRF7-S101K (7→5 DBD)	N/A	14.4 uL	250 nM	50 nM NaCl	A488 conjugated anti-GST polyclonal (LifeTech A11131)

Table 2.2 - Single stranded DNA oligos used to generate EMSA probes.

The IRF binding mode EMSAs used complementary oligo annealing to generate double-stranded probes. The complementary sequences are not listed. The IRF PBM Probe EMSAs used polymerase extension to form dsDNA products. The annealing site for the complementary extension primer is underlined. Refer to Chapter 3, Figure



DNA oligos used in EMSA experiments (Chapter 3)

Experiment	Oligo Name	Oligo Sequence
IRF Binding Mode EMSA	Monomer	GCACCGCTAACCGAAACTGTGC
	Proximal dimer	GCACCGCTAACCGAAACCGAAACTGTGC
	Extended dimer	GCACGCTAACCGAAACGCTAACCGAAACTGTGC
IRF PBM Probe EMSA	C-2	GCCTAGCACTAACCGAAACCGAAACCTAAGTGCTAGGTCTTGATTCGCTTGACGCTGCTG
	C-3	GCCTAGCACTAACCGAAACCCGAAACCTAAGTGCTAGTCTTGATTCGCTTGACGCTGCTG
	PRD3	GCCTAGCACATAGGAAAACTGAAAGGGAGGTGCTAGGTCTTGATTCGCTTGACGCTGCTG
	Extension primer	CAGCAGCGTCAAGCGAATCAAGAC

Table 2.3 – MPRA Protocol PCR conditions

A) PCR reaction conditions and PCR program used in MPRA ssDNA OLS library amplification. **B)** PCR reaction conditions and PCR program used for amplicon sequencing of MPRA plasmid library after CRE insertion.

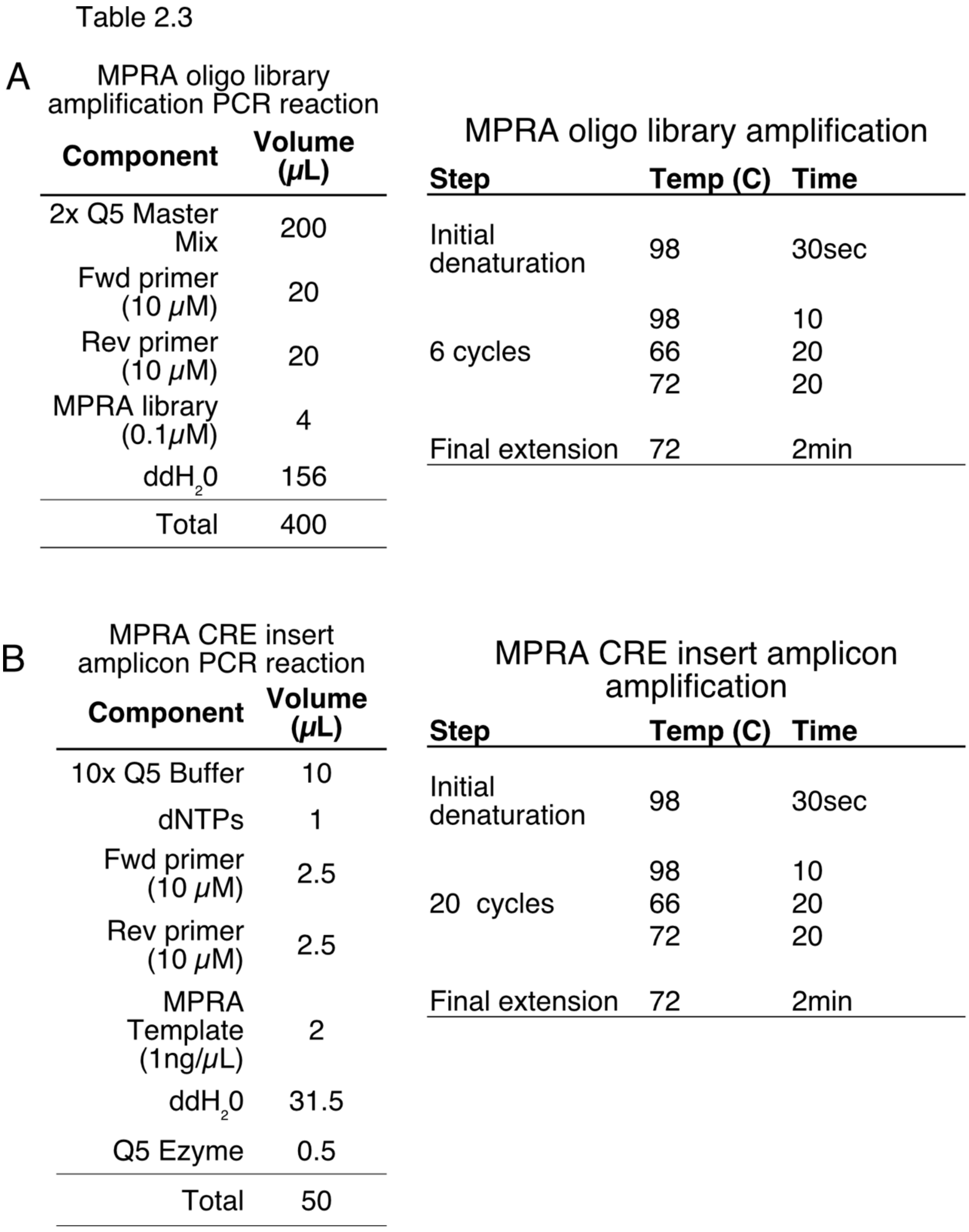


Table 2.3

A

MPRA oligo library amplification PCR reaction

Component	**Volume (μL)**
2x Q5 Master Mix	200
Fwd primer (10 μM)	20
Rev primer (10 μM)	20
MPRA library (0.1μM)	4
ddH$_2$0	156
Total	400

MPRA oligo library amplification

Step	**Temp (C)**	**Time**
Initial denaturation	98	30sec
6 cycles	98 66 72	10 20 20
Final extension	72	2min

B

MPRA CRE insert amplicon PCR reaction

Component	**Volume (μL)**
10x Q5 Buffer	10
dNTPs	1
Fwd primer (10 μM)	2.5
Rev primer (10 μM)	2.5
MPRA Template (1ng/μL)	2
ddH$_2$0	31.5
Q5 Ezyme	0.5
Total	50

MPRA CRE insert amplicon amplification

Step	**Temp (C)**	**Time**
Initial denaturation	98	30sec
20 cycles	98 66 72	10 20 20
Final extension	72	2min

Table 2.4 – MPRA PCR primer sequences

ssDNA oligos used for MPRA library PCR reactions. See table 2.3 for reaction conditions.

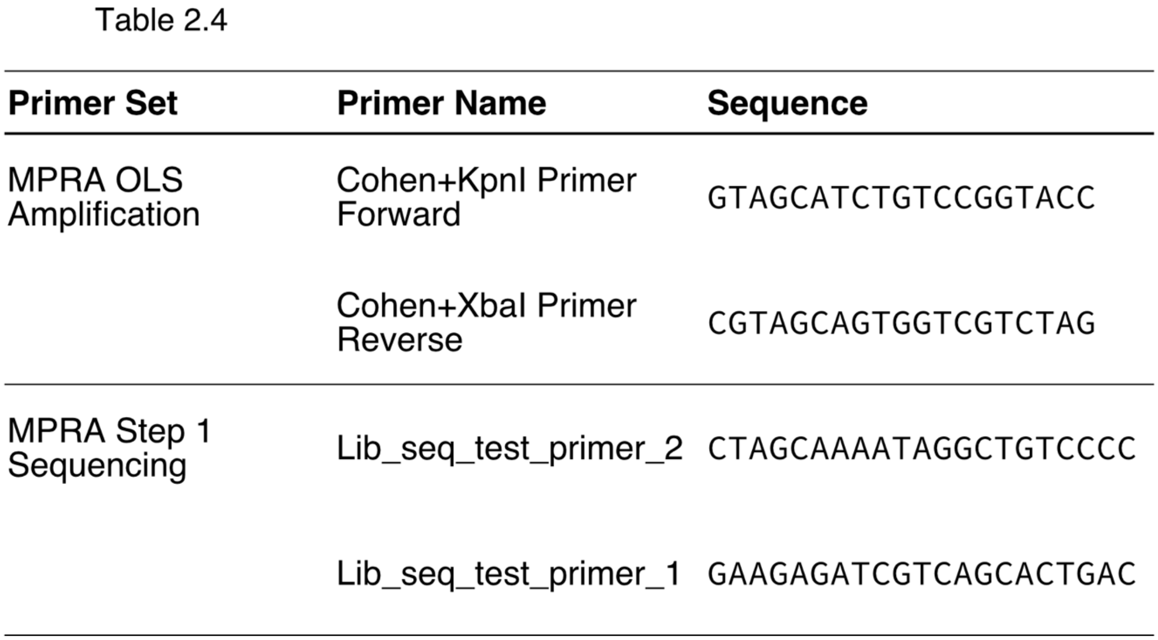


Table 2.4

Primer Set	Primer Name	Sequence
MPRA OLS Amplification	Cohen+KpnI Primer Forward	GTAGCATCTGTCCGGTACC
	Cohen+XbaI Primer Reverse	CGTAGCAGTGGTCGTCTAG
MPRA Step 1 Sequencing	Lib_seq_test_primer_2	CTAGCAAAATAGGCTGTCCCC
	Lib_seq_test_primer_1	GAAGAGATCGTCAGCACTGAC

3 CHAPTER 3

3.1 Introduction

Pathogen detection by pathogen recognition receptors (PRRs), such as the Toll-like receptors (TLRs) and the RIG-I-like receptors (RLRs), activate a network of transcription factors (TFs) that regulate host defense genes (Honda, Taniguchi, 2006). The TFs interferon regulatory factor 3 (IRF3), IRF5 and IRF7 (IRF3/5/7) are central to PRR signaling in response to viruses and intracellular pathogens with distinct, yet overlapping, roles in host defense (Honda and Taniguchi, 2006; Lazear et al., 2013; Stetson and Medzhitov, 2006a, 2006b). IRF3/5/7 reside predominantly in the cytoplasm and upon PRR-induced phosphorylation they dimerize and translocate to the nucleus to promote gene transcription (Barnes et al., 2002; Honda et al., 2006). Upon activation, IRF3/5/7 induce both common and factor-specific target genes (Barnes et al., 2002; Cheng et al., 2006; Honda et al., 2006; Tamura et al., 2008), but they can also function antagonistically (Barnes et al., 2002; Negishi et al., 2012). For example, IRF3 can repress (IFNA10, IFNA22) or enhance (IFNA1, IFNA7) IRF7-dependent gene activation (Barnes et al., 2002). Despite their central role in the response to pathogens, little is known about the mechanisms by which IRF3/5/7 target overlapping gene programs.

IRFs share a conserved N-terminal DNA-binding domain that recognizes a consensus 5'-AANNGAAA-3' DNA sequence found upstream of many virus- and IFN-inducible genes (Honda and Taniguchi, 2006). Activated IRF3/5/7 function as dimers that recognize the longer composite dimer site 5'-A/GNNGAAANNGAAA-3' (Figure 3.1A), referred to as the IFN-stimulated response element (ISRE) (Tamura et al., 2008). Despite a shared ability to bind to consensus binding elements, the inherent DNA-binding differences between IRF3/5/7 may partially account for regulatory differences (Cheng et al., 2006; Tamura et al., 2008; Yanai et al., 2007). In virus-infected BJAB lymphoma cells, expression of exogenous IRF5 or IRF7 induced 568 and 630 target genes, respectively (Barnes et al., 2004); however, only 371 (approx. 60%) of the genes were induced by both proteins, suggesting that DNA-binding differences led to alternate target genes. Similarly, studies examining the regulation of the type I IFNs by IRF3/5/7 have revealed inherent DNA-binding and regulatory differences for each factor, and attributed regulatory differences to dimer-specific binding to the viral-response elements (VREs) in the type I IFN gene promoters (Barnes et al., 2004; Civas et al., 2006; Génin et al., 2009; Yeow et al., 2000). The full extent of IRF3/5/7 DNA-binding differences and their impact on dimer-specific gene regulation remains unclear.

To better understand the scope of IRF3/5/7 DNA-binding differences and their role in defining dimer-specific target genes, we have used protein-binding microarrays (PBMs) to characterize the DNA-binding landscape of IRF3/5/7 dimers We used constitutively dimeric, phosphomimetic mutants of IRF3 (Chen et al., 2008a; Marié et al., 2000; Ren et al., 2014; Ryzhakov et al., 2015; Takahasi et al., 2010), IRF5 (Cheng et al., 2006) and IRF7 (Caillaud et al., 2005) (Figure 3.1B) to best characterize the active dimeric form of the IRF proteins. We examined the binding of IRF3/5/7 homodimers to thousands of ISRE-type elements to characterize both common and dimer-specific DNA binding features. To explore the role of binding affinity in the regulation of type I IFN genes, we characterized IRF3/5/7 binding to the VREs of all human and mouse type I IFN genes. Finally, integrating our PBM data with gene expression data, we relate DNA binding affinity to gene regulatory specificity. Our results provide new insights into the role and limitations of affinity as a distinguishing mechanism of IRF3/5/7 gene expression and function.

3.2 Results

3.2.1 Characterizing IRF3/5/7 dimer binding with PBMs

PBMs are double-stranded DNA microarrays that enable the *in vitro* measurement of protein binding to tens of thousands of unique DNA sequences (Berger et al., 2006; Siggers et al., 2011). To examine the DNA-binding specificity of IRF3/5/7 we used custom-designed PBMs. We designed a PBM that included 10,044 IRF-type binding sites spanning a range of affinities, half-site sequences, and half-site spacing (Figure 3.1). To query base preferences across the IRF binding site, we designed the PBM to contain 162 distinct IRF binding sites and all possible single-nucleotide variants (SNVs) across a 20-bp sequence centered on the IRF binding site (Figure 3.1A). This SNV-type PBM design allowed us to directly and comprehensively compare the base preferences of the IRF3/5/7 dimers across the IRF binding sites, and to construct position-weight matrix (PWM) descriptions of IRF3/5/7-DNA binding (Figure 3.1 D, Materials and Methods).

PBM and other experiments were performed using phosphomimetic variants of IRF3(6D) (Chen et al., 2008a; Lin et al., 1999), IRF5(4D) (Cheng et al., 2006) and IRF7(8D)(Caillaud et al., 2005) that have been previously shown to form the active, homodimeric form of each protein (Figure 3.1B). Previous studies examining IRF binding used monomeric DNA-binding domains (DBDs) (Badis et al, 2009), or overexpressed the proteins in HEK293T cells in conditions not known to promote dimerization (Jolma, et al., 2013). Hereafter, the phosphomimetic proteins are referred to simply as IRF3, IRF5 and IRF7 (unless otherwise noted). The PBM experiments were performed with purified GST-tagged IRF3/5/7 homodimers at four concentrations, and relative binding constants (KPBM) were determined for each sequence (Chapter 2). We have previously shown this approach of integrating PBMs at multiple concentrations can reliably define relative protein-DNA binding affinities (Siggers et al., 2011). Log(KPBM) values were transformed into z-scores based on the distribution of scores from 600 random DNA probes (i.e., a background set). We use z-scores to represent our PBM binding data.

To assess the quality of our PBM experiments, we examined the z-scores for known *in vivo* IRF target sites and our ability to capture known dimer-specific DNA-binding preferences. We find that known IRF binding sites are bound significantly better than the random background sequences (Figure 3.1C), demonstrating sensitivity in our assay. DNA-binding logos were constructed for the IRF dimers (Figure 3.1D; Chapter 2). While clear differences are evident for each IRF dimer, these logos agree with reported ISRE sites (Tamura et al., 2008) and logos generated from a high-throughput HT-SELEX assay (Jolma et al., 2013).

The IRF5 logo from our PBMs resembles a half-site logo with a single core 5'-GAAA-3' element. This logo matches the HT-SELEX 'monomer' logo, as opposed to the 'extended' logo that resembles the full-length logos determined for IRF3 and IRF7 that appear to contain two 5'-GAAA-3' elements (Figure 3.2) (Jolma et al., 2013). For several reasons, we believe that this shorter logo represents a true dimer site but results from an IRF5 preference to engage more strongly to one ISRE half-site. First, if we relax the β parameter in our logo generation procedure (Figure 3.2, Chapter 2 - Materials and Methods), we see the correct 5'-GAAA'-3' preferences appearing in the 5-prime half of the logo. This suggests that both half-sites are engaged by IRF5 proteins and that the base preferences are simply weaker for the 5-prime half-site. Second, with our SNV approach to logo generation, we can generate a binding logo for individual sites by altering the DNA base identities and monitoring the change in binding affinity. We find that while the majority of starting seed sequences (13/15) result in the shortened half-site logo (as in Figure 3.1), logos from a minority of seed sequences (2/15) appear longer, and match the extended-version of the IRF5 logo also identified by HT-SELEX. We find that these extended logos occur when the 5-prime half-site in the seed is a better match to the consensus 5'-GAAA-3' than for the 3-prime half-site (e.g., 5'-GAAANNGATA-3'). These observations suggest that IRF5 binds as a dimer with an asymmetric half-site preference.

As a further confirmation of the PBM data, we compared PBM z-scores to electro-mobility shift assays (EMSAs) for select high- and low-affinity binding sites (Figure 3.3). We find that EMSAs qualitatively recapitulate the differential binding indicated by our z-scores: IRF3 z-scores 11.4 to 6.3 correspond to >25-fold change in affinity and 11.4 to 3.6 correspond to >100-fold change in affinity; IRF5 z-score 16.4 to 2.2 correspond to >100-fold change in affinity; IRF7 z-scores 16.6 to 5.1 correspond to > 50-fold change in affinity. These results demonstrate that our PBM data accurately captures the DNA-binding landscape of IRF3/5/7 dimers over a wide range of binding site affinities.

3.2.2 Common and IRF-specific binding sites

To investigate the nature and extent of IRF3/5/7 binding differences we compared the PBM-determined binding profiles for the IRF dimers (Figure 3.4A). Consistent with binding logo differences, we observed IRF-specific binding preferences indicated by 'off-diagonal' data points that represent sequences bound much better by one IRF dimer than another. For comparison, pseudo replicate IRF experiments showed no such off-diagonal data points (Figure 3.4B). DNA sequence preferences for the IRF dimers can be queried by examining the distribution of specific sequence subsets. For example, examining the z-score distribution of all sequences that contain IRF sites with the alternate 5'-GATA core elements (i.e., that match the pattern 5'-GATANNGATA-3') we see clear dimer preferences (Figure 3.4C). Specifically, IRF7 cannot tolerate these alternate core elements and, therefore, all sequences in the IRF7 experiment have z-scores indistinguishable from background (i.e., z-score near zero). In contrast, both IRF3 and IRF5 can bind sequences with this alternate core sequence. For IRF5, some of these 5'-GATA sequence variants score among the highest in our dataset; however, this is not the case for IRF3 indicating that IRF5 is more tolerant of this alternate core element. Pairwise comparison of our PBM data revealed many differences between the IRF3/5/7 dimers.

In general, base preferences that distinguish the DNA binding of IRF3/5/7 dimers from each other are apparent in their respective DNA-binding logos (Figure 3.1D). For example, based on the logos, IRF3 appears more tolerant than IRF5 of base variants at most positions; however, this difference is most apparent at position 11 where IRF5 highly prefers a Cyt base. We can directly confirm this selectivity by analyzing our PBM data and identifying single-base changes (i.e., SNVs) that abrogate binding of IRF5 but not IRF3. We identified 8 such IRF5-abrogating SNVs in our PBM dataset, and all were C-to-G changes at position 11. For example, 5'-CCGAAACCGAAACC-3' was bound highly by both IRF3 (z-scores 10.9) and IRF5 (z-score 14.8), but the SNV 5’-CCGAAAC**G**GAAACC-3’ was bound well by IRF3 (z-score 8.2) but not by IRF5 (z-score 2.9). Base differences at position 11 largely explain the bifurcation seen in the PBM data (Figure 3.4C) in which IRF5 is bound poorly to a number of DNA sites that are bound with high affinity by IRF3. In a similar manner, we examined the selection against IRF7 binding observed in Figure 3C for the 5’-GATANNGATA-3’ sequences. Comparing the logos of IRF5 and IRF7, we see that IRF7 is primarily more base selective at positions 7 and 8, suggesting that SNVs at these positions differentially affect IRF5 and IRF7. We identified 50 sequence pairs in which a SNV abrogates IRF7 binding but does not perturb high-affinity IRF5 binding: 41/50 SNVs were A-to-C,G or T changes at position 8, which disrupts the IRF7-preferred Ade base, and 2/50 SNVs were A-to-T changes at position 7. For example, IRF5/7 both bound to 5’-CCGAAACCGAAACC-3’ with high affinity (z-scores 16.6 and 16.4, respectively), but the SNV 5’-CCGA**C**ACCGAAACC-3’ was bound well by IRF5 (z-score 13.6) but not by IRF7 (z-score 4.2). Unexpectedly, we also found 5 A-to-T SNVs at position 14 were selective against IRF7, but not readily expected from the logo comparisons. For example, IRF5/7 both bound to 5’-CCGATACCGAAACC-3’ with high affinity (z-scores 14.9 and 9.8, respectively), but the SNV 5’-CCGATACCGA**T**ACC-3’ abrogated IRF7 binding (z-score 2.2) while IRF5 was largely unaffected (z-score 12.0). This IRF5 (and IRF3) tolerance for a Thy at position 14 is indicated by a slightly weaker Ade selectivity in the logo. Pairwise comparison of IRF3/5/7 logos and our SNV binding data reveal features of IRF3/5/7 binding specificity that distinguish the separate dimers and highlight the impact that genomic SNPs may have on relative IRF3/5/7 binding and function.

Despite the binding differences observed for the IRF dimers, we observe common, high affinity binding sites shared by all three proteins. For example, the IRF high affinity probe (Figure 3.1C) is a top-scoring sequence for each IRF dimer and is consistent with the known canonical ISRE element. This shared specificity landscape for close paralogs, in which both common and dimer-specific binding sequences are observed, has been reported for a number of transcription factor paralog families (reviewed in (Andrilenas et al., 2015)). The existence of shared and dimer-specific sites provides a potential mechanism for the IRF dimers to regulate both common and dimer-specific genes.

3.2.3 Alternate ISRE half-site spacing

IRF3/5/7 binding has been reported for binding sites with both the canonical 2-bp spacer between each 5'-GAAA-3' half-site, as well as with a 3-bp spacer, hereafter referred to as 2- and 3-bp sites (Escalante et al., 2007; Panne et al., 2007). However, DNA-binding logos determined using our PBM dataset (Figure 3.1), other *in vitro* methods such as HT-SELEX (Jolma et al., 2013), or learned from ChIP-seq datasets (Freaney et al., 2013) support a dominant 2-bp site. The X-ray crystal structure of IRF3 and IRF7 bound to the IFNB gene promoter demonstrates the binding of alternate IRF3 and IRF7 DBDs that support binding to both a 2-bp site (Figure 3.5A, PRDIII) and an adjacent 3-bp site (Figure 3.5B, PRDI). Therefore, we sought to determine whether IRF dimers could bind to both 2-bp and 3-bp sites with similar affinities, and whether specific IRF dimers have individual preferences.

We first examined the binding of the IRF3/5/7 dimers to the variant ISRE elements found in the IFNB promoter for which there is structural evidence for binding to both 2- and 3-bp sites. The X-ray crystal structure supports a simple model of two adjacent IRF3-IRF7 heterodimers bound to the adjacent PRDIII (2-bp) and PRDI (3-bp) sites, P1 and P3, respectively (Figure 3.5A). However, there is a third possible ISRE element (P2) that spans these canonical elements (Figure 3.5A). We used PBMs to assay the binding of the IRF3/5/7 dimers to all three possible ISRE sites (Figure 3.5 A). We find that IRF3 and IRF7 bind well to both the proto-typical ISREs PRDIII and PRDI, and IRF5 binds well to PRDI. Notably, all three dimers bind most strongly to the 3-bp PRDI site. While we cannot determine the binding register from our PBM data (i.e., whether canonical residues contact the same positions in the 5'-GAAA-3'-type half-site), X-ray crystal structures (Escalante et al., 2007; Panne et al., 2007) suggest IRF dimers bind to the PRDI site in a 3-bp binding register. These results demonstrate that, in the context of the IFNB promoter, the 3-bp PRDI site is the highest affinity IRF binding site. Therefore, a description of IRF-DNA binding that does not include 3-bp sites will fail to capture functional binding sites.

To determine the impact of the spacer length on IRF binding, we examined the binding of IRF3/5/7 to 120 matched pairs of 2-bp/3-bp ISRE sites. Matched pairs differ by a single base in the spacer between half-sites (i.e., GAAANNGAAA -> GAAANXNGAAA, Figure 3.5 B). We find that altering the spacing from 2- to 3-bp lowers the binding affinity for all three IRF dimers by a similar amount (mean delta z-score ~ 5.0)(Figure 3.5B). These data suggest an explanation for the canonical 2-bp PWMs (i.e., binding models) described in the literature (Freaney et al., 2013; Jolma et al., 2013; Mathelier et al., 2016) — 2-bp sites are higher affinity. However, despite the lower binding affinity to 3-bp binding site variants, we find that many 3-bp sites in our dataset score well above background. For example, IRF7 binds to

559 of 3456 3-bp IRF site variants with z-scores > 4.0. Therefore, many 3-bp sites in the genome are likely of sufficient affinity to be functional (e.g., PRDI in the IFNB promoter).

Previous studies (Dragan et al., 2007; Panne et al., 2007) have also proposed an alternate binding mode in which IRF dimers are bound to half-sites 8-bp apart (e.g., 5'-GAAA(N)8GAAA-3'). The extended binding conformation would allow IRF proteins to reside on the same side of the DNA helix(Panne et al., 2007) (Figure 3.5C). We examined the binding of IRF3 to this extended site by electro mobility shift assay (EMSA)(Figure 3.5C). We found that IRF3 binding to an 8-bp extended-dimer site was at least 100-fold weaker than to a sequence-matched 2-bp site. These results reaffirm that the preferred binding mode for IRF dimers is to closely-spaced half-sites (i.e., 2-bp and 3-bp sites), and not to an extended dimer site (i.e., 8-bp site).

3.2.4 Differential dimer binding to type I IFN gene VREs

IRF3/5/7 regulate the type I IFN genes (Honda et al., 2006; Stetson and Medzhitov, 2006a; Tamura et al., 2008) that coordinate immunity to viruses and other intracellular pathogens (Honda et al., 2006; Stetson and Medzhitov, 2006a). The type I IFN genes consist of IFNB and multiple IFNA genes (Figure 3.6). IRF3 and IRF7 are the primary type I IFN regulators (Honda et al., 2006, 2005; McNab et al., 2015); however, IRF5 can also regulate select IFNs (Tamura, Yanai, Savitsky, Taniguchi, 2008). Induction of the type I IFNs in virus-infected cells is primarily a consequence of IRF regulatory input from the promoter VREs (Génin et al., 2009; Honda and Taniguchi, 2006). Despite their central role in IFN gene regulation, it remains unclear to what extent DNA-binding of IRF3/5/7 differs across the VREs.

To address the role of DNA binding in type I IFN gene regulation, we used PBMs to measure the binding of IRF3/5/7 to the promoters of all type I IFN genes from human and mouse. We measured dimer binding to all potential IRF binding sites found in the 250-bp immediately upstream of the IFN genes that encompasses the VREs, these regions were selected because of their density of IRF-binding sites (VREs) and the sequence space constraints of the PBM platform. Considering only significant binding sites (PBM z-score > 4.0), we found there was little binding to the regions outside of the VRE; therefore, we have focused our analysis on the VREs.

For both human and mouse, we observe binding for IRF3 and IRF7 to multiple sites within the VREs (Figures 3.6A and 3.6B). Studies have delineated sub-elements (A/B, C and D) within the IFN VREs that play critical but different roles in IFN regulation (Civas et al., 2006, 2002; Génin et al., 2009; Yeow et al., 2000) (Figures 3.6A and 3.6B). We find differences in IRF3 and IRF7 binding profiles across the various VREs sub-elements. IRF3 and IRF7 bind with similar, though not identical, patterns to the VRE-C and VRE-D elements. However, they differ strongly in their binding to VRE-A/B elements — IRF7 binds VRE-A/B in most VREs (20/27), whereas IRF3 binds in only three (3/27). It has been proposed, based on studies of mouse Ifna4 and Ifna2 genes (Civas et al., 2002), that late-phase expression of IFNA genes is controlled by Irf7 binding to VRE-A/B, whereas early-phase expression is controlled by Irf3 binding to VRE-C. We show that for many IFNs (63%, 17/27) it is only IRF7 that binds to VRE-A/B, suggesting that late-phase expression driven by IRF7-VRE-A/B binding may be a common regulatory feature for most IFNs. The sequence logo for IRF sites in the human and mouse VRE-A/B (Figure 3.7) does not indicate a simple explanation for the IRF7 binding preference over IRF3. In contrast, we find that the majority of VRE-C elements are bound by both IRF3 and IRF7, with only the VRE-C of the mouse Ifna4 and Ifnab genes showing IRF3-exclusive binding, suggesting that the regulatory logic of early-phase IFN expression from VRE-Cs is more complicated.

Unexpectedly, we observed a near complete lack of IRF5 binding to the IFNA VREs. IRF5 does not bind to any human VREs, and only binds three mouse VREs (Figures 3.6A and 3.6B). As a control, we see strong IRF5 binding to regulatory loci for other IRF target genes, such as IFNB (Barnes et al., 2002), IL10 (Krausgruber et al., 2011), and CXCL10 (Honda and Taniguchi, 2006) (Figure 3.6C). The lack of IRF5 binding to the VREs is broadly consistent with the less prominent role described for IRF5 in the regulation of the IFNA genes (Honda and Taniguchi, 2006; McNab et al., 2015; Tamura et al., 2008). However, given the large number of IRF binding site variants that *can be* bound strongly by all three IRF dimers (Figure 3.4A), it is striking that the IRF binding site variants from human and mouse VREs would maintain their inability to bind to IRF5. The sequence logo for IRF sites in the human and mouse VRE regions (Figure 3.7) highlights the sequence variability across individual sites and suggests a mechanism for the observed absence of IRF5 binding. There is a nearly complete absence of Cyt bases at positions 11 and 16 flanking the 5'-GAAA-3' core that are both highly preferred by IRF5 (Figure 3.1). The evolutionarily conserved absence of IRF5 binding suggests a selective pressure against IRF5 binding to these loci that is based on selective use of binding sites variants bound preferentially by IRF3/7 (addressed more below).

3.2.5 A single amino acid dictates the IRF binding selectivity to IFNA VREs

To investigate how IRF5 binding is selected against in the VREs, we examined the multiple protein sequence alignment for IRF3/5/7 to find potential specificity-altering residues. We reasoned that residues critical to distinguishing IRF3/5/7 binding specificity would (i) vary between all three IRFs (but would be conserved between human and mouse orthologs), and (ii) make base-specific contacts with DNA. Based on protein-DNA interactions defined in available IRF3 and IRF7 crystal structures (Escalante et al., 2007; Panne et al., 2007), we found three residues that fit these criteria (Figure 3.6E). We examined the IRF sites within the IFNA VREs and determined that selection against IRF5 binding was most likely due to DNA sequence features 3-prime to the canonical 5'-GAAANNGAAA-3' ISRE sequence; therefore, we chose to examine the residue position in alpha helix α3 shown to contact DNA bases in this region (Figure 3.6E, red highlight). Given that IRF7 and IRF5 exhibited the largest differences in their binding profiles across the IFNA VREs, we made mutants in which we swapped orthologous residues between IRF5 and IRF7 — IRF5(K96S) and IRF7(S101K).

We examined the DNA binding of the IRF5(K96S) and IRF7(S101K) mutants by PBM experiment. Binding logos generated for the mutants revealed that the amino acid alterations at this position neatly swapped the DNA binding specificity for IRF5 and IRF7 at the 3-prime end of the DNA binding sites (Figure 3.6D). Mutations at this position had the effect of making IRF5 more tolerant of alternate bases in this region, while making IRF7 less tolerant (e.g., strong selectivity for a cytosine at base position 16). Examining the binding of the mutant IRFs to the IFNA VREs (Figure 3.6A and 3.6B), we found that IRF7(S101K) had drastically reduced binding across the VREs compared to IRF7, while IRF5(K96S) showed increased binding compared to IRF5. These results identify this residue position in helix α3 (Figure 3.6E) as a key determinant of IRF binding specificity that can alter the base preferences in the 3-prime flanks of the canonical ISRE. Furthermore, base contacts mediated by residues at this position are critical for the selective binding of IRF7 over IRF5 across the IFNA VREs. Therefore, poor binding of IRF5 to VREs is partially explained by the bases flanking the core ISRE elements that are unfavorable for interaction with IRFs containing a lysine residue at this position in the DBD (i.e., IRF5 and IRF7 S101K).

3.2.6 Binding affinity contributes to ISRE function and selectivity

We next sought to determine the extent to which IRF3/5/7 binding differences translate into IRF-specific gene regulation. We examined whether IRF3/5/7-specific binding site variants discovered in our dataset would drive IRF-specific gene activation. Dimer-specific binding site variants (I3, I5, I7) that show preferential binding to IRF3, IRF5 or IRF7, respectively, were chosen from the PBM data (Figures 3.9A, 3.9B). To relate our findings to IFNA gene regulation, we examined the impact of these site variants on gene expression in the context of the 250-bp human IFNA14 promoter. We observed no significant IRF binding to the IFNA14 VRE in our PBM experiments (Figures 3.6A, 3.6B), and others have similarly reported no IRF3/7 binding to this element (Yeow, Au, Juang, Fields, Dent, Gewert, Pitha-Rowe, 2000); therefore, it provided a useful promoter context in which to examine the impact of binding site variants on gene expression. Binding site variants were inserted simultaneously into the VRE-C and VRE-D elements of the IFNA14 promoter and we examined the ability of each variant promoter to drive reporter gene expression in the presence of constitutively active IRF dimers: IRF3(6D), IRF5(4D) and IRF7(6D) (Figure 3.9 A). For reporter gene experiments, IRF7(6D) was used instead of IRF7(8D) as it is a stronger activating dimer (Caillaud, Hovanessian, Levy, Marie, 2005).

We found that dimer-specific binding sites can promote dimer-specific gene activation. Promoters with IRF3- and IRF5-specific sites (I3, I5) were selectively activated by IRF3 and IRF5, respectively (Figure 3.9 B,C). However, affinity-independent mechanisms also appear to contribute to IRF binding site function. Promoters with IRF7-specific sites (I7) were strongly activated by IRF7 but, unexpectedly, also by IRF5 (Figure 3.9 B,C). Furthermore, IRF5 drove expression from the I7 promoter at a 4.1-fold higher level than for the I5 promoter, despite the fact that IRF5 binds to I7 with lower affinity than to I5 (z-scores 3.4 and 10.7, respectively). We do not believe that heterodimerization with endogenous IRF proteins contributes to our observation that affinity does not correlate with activity. IRF3 and IRF7 are expressed at low levels in our WT and transfected HEK293T cells (Figure 3.8); however, in our unstimulated cell-culture conditions IRF proteins are not known to be active. Furthermore, we observe very low reporter gene activation with eGFP transfected in place of our active IRF dimers, suggesting that the endogenous IRF3/7 are not active and would not contribute to gene expression. These data show that IRF3/5/7-specific gene activation depends on both affinity and affinity-independent mechanisms.

We next tested whether ‘common’ binding sites bound by all three IRF dimers would be functional for each dimer. We found that IRF3/5/7 all drove reporter gene expression from the C-2 promoter bound with high affinity by all three IRFs (Figure

3.9B, C). However, expression from the C-2 promoter was not the highest for any of the IRFs, despite C-2 being the highest-affinity binding site for all three IRF dimers (Figure 3.9 B). For example, IRF7-stimulated expression was 6.1-fold lower from the C-2 promoter than from the I7 promoter despite having a much higher binding affinity (z-score 17.0 versus 9.5). This diminished activity for the C-2-promoters suggests that for IRF3/5/7 dimers high-affinity binding may depress gene activation (discussed more below).

3.2.7 Affinity-independent mechanisms of ISRE function

To understand affinity-independent mechanisms of ISRE function, we first investigated whether ISRE half-site spacing contributed to IRF-dependent reporter activity. We compared the activity of binding site variants that differed only in their half-site spacer length (i.e., 2-bp versus 3-bp sites). Specifically, we compared the activity of a 2-bp (I7-2) and 3-bp (I7) version of our IRF7-specific site, and a 2-bp (C-2) and 3-bp (C-3) version of a high-affinity site common to IRF3/5/7 (Figure 3.9 A). These matched pairs of binding-site variants differed only by a single base in the spacer sequence (Figure 3.9 D).

Spacer variants for the common high-affinity sites (C-2 and C-3) actually led to similar reporter activity levels for all three IRFs (Figure 3.9E), despite the fact that all IRFs bind with lower affinity to the C-3 variant (Figure 3.9 D). Similarly, spacer variants for the IRF7-specific site (I7-2, I7) also led to similar expression levels for IRF7 despite lower binding affinity to the 3-bp variant. This surprising congruence in activity for IRF-binding-site spacer variants suggests that either (1) 3-bp sites are more functionally active and can make-up for a lower binding affinity; or (2) that DNA-sequence features, which are virtually identical between spacer pairs, are critical to the activity level of IRF sites.

Additional promoter comparisons clarify that DNA-sequence features beyond spacer length influence the activity of IRF binding sites. First, binding sites with similar IRF binding affinity can exhibit different activity: C-3 and I7 have similar IRF7 binding affinity (z-scores 11.2 and 9.5), and the same spacer lengths, but I7 induces 6.5-fold higher gene activation (Figure 3.9 D, E). Second, we noted that the wild-type IFNA14 promoter, intended as a control in our assay, is activated by IRF7 despite low binding affinity to the WT-C and WT-D sites on PBM (z-scores < 4.0) (Figure 3.9 D,E). These data support a model in which IRF-site DNA sequence features, beyond half-site spacing, contribute to the regulatory activity of an IRF site in an affinity-independent manner.

To determine the DNA sequence features of IRF binding sites that may affect their activity, we compared the DNA sequences of the variant sites tested in reporter assays (Figure 3.9 A). The I7-2 and C-2 sites promote different levels of gene activation (from 2.6-fold for IRF3 to 6.1-fold for IRF7) despite having identical core ISRE sequences — C-2 and I7-2 share an identical 5’-CCGAAACCGAAA-3’ core. Furthermore, the most transcriptionally active of these sites, I7-2, is lower affinity than C-2. This differential activity and DNA sequence similarity indicates that DNA bases flanking the 12-bp core element can modulate the transcriptional activity of an IRF site even at the expense of binding affinity.

3.3 Limitations

In this study we systematically characterized the DNA-binding landscapes of homodimeric IRF3/5/7 proteins, using PBMs, and examined the impact of small variations in IRF binding sites on reporter gene expression. While our results provide new insights into the potential role of affinity as a distinguishing mechanism of IRF3/5/7 gene expression and function, this research is not without limitations.

First, a notable limitation of the PBM platform is that it is is an in vitro protein-DNA binding assay that does not approximate the cellular transcription factor (TF) DNA binding environment of the nucleus. TF-DNA binding and transcriptional regulation in the nucleus involves complex 3D changes in genome architecture as well as recently described changes in matter-phase from a liquid to gel surrounding enhancer regions (Shrinivas et al., 2019). These changes in matter state may cause localized changes in protein concentration and pH generally altering interaction kinetics. These conditions are not captured by current in vitro binding assays including PBMs and EMSAs. Additionally, in vitro protein-DNA binding experiments rely on conditions that are not found in the cell nucleus. Notably PBMs utilize printed spots of DNA with many copies of each DNA molecule closely packed and covalently linked to a glass slide. This environment likely alters the movement of DNA and adds constraints to the accessibility of a DNA molecule by a TF in comparison to a DNA molecule in solution. Despite these limitations, PBMs provide protein-DNA binding affinity data in agreement with other in vitro methodologies and allow the high-throughput measurement TF-DNA binding (Badis et al., 2009; Berger et al., 2006, 2008; Linnell et al., 2004; Siggers et al., 2012a). Furthermore, DNA-binding motifs derived from in vivo methodologies exhibit striking agreement with in vitro-defined DNA-binding motifs, suggesting that DNA binding specificity in vivo adheres to basic biophysical parameters that can be determined reliably using these in vitro methods. However, deviations form in vitro and in vivo results occur and provide opportunities for identifying novel mechanisms of specificity; therefore, we believe that careful in vitro biophysical characterization, and comparison of these results to in vivo experiments, is a necessary step in understanding the mechanisms of TF regulatory specificity.

An additional limitation that may impact the interpretation of our data is the use of *E. Coli* expressed IRF3/5/7 proteins. Exogenous expression of IRF3/5/7 and subsequent affinity tag purification produced heterogenous protein samples with multiple degradation products. The heterogeneous nature of these protein samples is visually apparent in the IRF EMSAs present in this chapter (Figures 3.3, 3.6). The multiple bands present in the IRF EMSAs likely represent mixtures of full-length IRF proteins dimerized with those that have lost their GST affinity tags. IRF monomers bind DNA with approximately 200-fold lower affinity than a constitutively active phosphomimetic IRF3 (Kd 1226 (WT) vs. Kd 5.8 (IRF3-5D)) Dragan:2007ga}, so they are unlikely to be present in the bands in our EMSAs. Another aspect of *E. coli*-based protein production that can impact the function of expressed human proteins is the lack of mammalian post-translational modifications (PTMs). Fortunately, IRF3/5/7 are only known to be modified in response to innate immune stimulation (phosphorylation) or for suppression of the IRF-dependent immune response via ubiquitination or acetylation.

In our mutational study of the structural determinants of IRF5 specificity we found that the IRF7-S101K and IRF5K96S mutants largely swapped DNA binding patterns, however the changes in specificity were incomplete. Additional amino acids near the residue we mutated may confer additional selectivity if they were mutated at the same time. IRF5-S97 and IRF7-T102 are additional residues that uniquely vary between IRF3/5/7, future studies examining the structural determinants of IRF DNA-binding could mutate these and check for complete conversion in specificity.

We chose to use phosphomimetic IRF proteins in our PBM and EMSA experiments in part due to concerns regarding the heterogeneity in producing large quantities of purified, phosphorylated IRFs. IRF3/5/7 protein samples used on the PBM were challenging to produce and the addition of a phosphorylation step would likely increase the heterogeneity of the samples. We were also interested in making systematic measurements of IRF3/5/7 DNA-binding affinity and chose not to use stimulated cell lysates in our PBM experiments to guarantee that observed differences in IRF3/5/7 DNA binding were due to inherent differences between the proteins rather than due to potential cofactor interactions. Additionally, given the complex regulation of the IRFs in both a cell-type and stimulus specific manner, finding stimulation conditions that similarly activate IRF3, IRF5 and IRF7 may be impossible. We chose to over-express our phosphomimetic IRF constructs in reporter assays to standardize comparisons between our PBM data and our reporters.

Studies of phosphomimetic IRFs show that they are constitutively active and dimeric; able to interact with transcriptional coactivators (i.e. CBP); and recapitulate IRF dependent gene regulation in both reporter assays and gene expression assays. Despite this, these constitutively active IRF mutants are a convenient experimental tool that may fail to capture the full impact of post-translational modifications (PTMs) (e.g. phosphorylation, acetylation, etc.) on IRF regulation. For example, in the paper characterizing the crystal structure of the IRF5 dimerization domain, Chen et al. (2008) suggest that the IRF5-S436D phosphomimetic (S462D in our construct) may not fully recapitulate dimer stabilizing effects hypothesized to occur with serine phosphorylation at that site (Chen et al., 2008b; Marié et al., 2000; Ren et al., 2014; Ryzhakov et al., 2015; Takahasi et al., 2010). Additionally, acetylation and ubiquitination are known to impact IRF7dependent gene regulation by suppressing the DNA-binding activity of IRF7 (acetylation) and flagging IRF7 for degradation. These are PTMs that are beyond the scope of our research, but are integral to the regulation of IRF7 activity and no mimetic mutants are known to recapitulate the effects of these PTMs. However, we are confident that the DNA-binding specificity of these *E. coli* produced IRF dimers provides a metric by which to compare the individual dimers and to connect inherent DNA binding differences with in vivo regulatory differences.

Heterodimer formation is an aspect of IRF3/5/7 DNA-binding activity that may have important implications for IRF DNA-binding and IRF-dependent gene regulation. IRF3, IRF5 and IRF7 have been shown to interact via immunoprecipitation western blots (Barnes et al., 2003) and it is suggested that IRF3/5/7 heterodimers regulate the expression of IFNβ and the IFNα genes (Honda and Taniguchi, 2006) . We hypothesize that heterodimer formation between IRF5 and either IRF3 or IRF7 could expand (Honda and Taniguchi, 2006b; Suhara et al., 2000; Wathelet et al., 1998)the DNA-binding repertoire of IRF5 allowing it to bind more strongly to the IFNα proximal promoters. We did not perform heterodimer PBM studies in part due to the potential complexity of interpreting heterogeneous mixtures of IRF5:IRF3/7 heterodimers and each respective IRF homodimer that may spontaneously form depending on the interaction kinetics of IRF heterodimers versus IRF homodimers. In future experiments examining IRF5:IRF3/7 heterodimer binding, the IFNα subset of IRF binding sites would be specifically informative. Given the notable lack of strong IRF5 binding to these sites, any significant increase in IRF5 binding could be attributed to heterodimer formation or potentially some form of indirect recruitment of IRF5 by IRF7, although there is no evidence for this kind of IRF homodimer indirect recruitment in the literature. These heterodimer PBM experiments would require testing IRF specific antibodies on the PBM, or producing IRF proteins with different epitope tags for detection.

Another context not captured in our in vitro experiments is the impact of chromatin state on IRF binding. Chromatin state is an essential component of gene regulation that dictates which genes are available as TF gene targets (Medzhitov and Horng, 2009). The PBM platform is unable to assess the impact of chromatin state on TF-DNA binding interactions. In general, biochemical protein-DNA-binding assays like PBMs, EMSAs, HT-SELEX and others, are unable to assay the impact of chromatin state on TF binding. Chromatin immunoprecipitation sequencing (ChIP-seq) is a well-known method for measuring chromatin state and TF-DNA interactions. ChIP-seq data can provide a snapshot of TF genome occupancy as well as chromatin state differences like histone modifications associated with enhancers and promoters. ChIP-seq provides broad TF occupancy windows that subsequently require TF binding site analysis. This limitation necessitates follow up using detailed biochemical methods such as PBMs and EMSAs to verify and characterize TF binding to generate a mechanistic model of TF dependent gene regulation. PBMs can incorporate ChIP-seq data to examine how genomic TF binding sites may be differentially bound. Unfortunately for our analysis of IRF3/5/7, there are few ChIP-seq data sets for these proteins. ChIP-seq data is dependent on finding antibodies with strong affinity and avidity. The lack of high-quality antibodies is a known problem in the IRF research field, exemplified by a paper that only assesses the variable quality of IRF5 targeted antibodies on the market (Li et al., 2016).

3.4 Discussion

IRF3/5/7 are central regulators of the host-defense program to pathogens (Honda and Taniguchi, 2006; Lazear et al., 2013; Stetson and Medzhitov, 2006a). Here, we addressed the ability of inherent IRF3/5/7 DNA-binding differences to define dimer-specific gene regulation. We characterized the DNA-binding preferences of IRF3/5/7 homodimers, and identified both common and dimer-specific DNA-binding preferences. We demonstrate that dimer-specific binding sites can promote dimer-specific reporter gene expression, showing that sufficient DNA-binding differences exist for IRF3/5/7 to induce unique target gene sets. We also found that affinity-independent mechanisms contribute to IRF3/5/7 binding site activity and, therefore, may also contribute to dimer-specific gene regulation (discussed more below). Currently, there are no genome-wide chromatin immunoprecipiation (ChIP) studies for IRF7 and IRF5 that would allow a comparison of in vivo IRF3/5/7 binding to our PBM data. Specifically, there are no ChIP studies for IRF7, and the single IRF5 study (Saliba et al., 2014) found no ISRE motif enriched in the IRF5-bound ChIP peaks, suggesting that under the conditions assayed IRF5 does not function as a canonical DNA-bound dimer.

Genome-wide binding studies for activated IRF3/5/7 dimers would help to clarify whether the inherent DNA-binding differences described here, and in other studies (Jolma et al., 2013), define the distinct global binding patterns and gene regulatory programs for IRF3/5/7.

IRF3/5/7 are central regulators of type I IFN expression, and the host-defense response to viruses and intracellular pathogens (Honda et al., 2006; Stetson and Medzhitov, 2006a; Tamura et al., 2008). To understand the role of DNA binding in IFN regulation, we mapped the binding of IRF3/5/7 dimers to all human and mouse IFN gene VREs. We found that IRF3 and IRF7 bind across many of the VRE sub-elements, but that the VRE-A/B sub-element is bound almost exclusively by IRF7. This suggests that the previously described role for IRF7-VRE-A/B binding to control late-phase IFN expression (Civas et al., 2002) is common to many IFN genes. Examining the landscape of IRF3/5/7 binding to the individual ISRE sites within the VREs, we find that for the 61 sites bound by at least one IRF dimer 93% (57/61) bound to IRF7, 60% (37/61) bound to IRF3, and only 5% (3/61) bound to IRF5. These data support the current understanding that IRF3 and IRF7 are the primary regulators of the type I IFNs (Honda and Taniguchi, 2006; Tamura et al., 2008).

The most striking results from our analysis of the type I IFN VREs was absence of IRF5 homodimer binding to all human and most mouse VREs. Our PBM binding data (Figure 3.5) demonstrate that there are many IRF site variants that can be bound by all 3 IRF dimers; therefore, the conspicuous absence of IRF5 binding to VREs from two evolutionarily distant species suggests that there has been selection against IRF5 homodimer binding to these loci. We demonstrate that a single amino acid difference between IRF7 and IRF5 was critical to their differential binding profiles across the VREs. Furthermore, this amino acid altered the binding preference for DNA bases immediately 3-prime to the canonical ISRE site (Figure 3.6D). Therefore, IRF5 homodimer binding to the VREs is inhibited by the evolutionary retention of specific ISRE variants with 3'-flanking bases that are unfavorable for IRF5.

IRF5 can regulate specific IFNA genes in a cell- and stimulus-specific manner (Barnes et al., 2003, 2001; Tamura et al., 2008; Yanai et al., 2007). Furthermore, IRF5 forms heterodimers with IRF3, and IRF3 can enhance the recruitment of IRF5 to IFNA promoters in virus-infected cells (Barnes et al., 2003, 2001). This suggests that IRF3:IRF5 heterodimers are the dimer species critical for IRF5-dependent regulation of the IFNA genes, as opposed to IRF5 homodimers, which we find do not bind well to IFNA VREs. We propose that a heterodimer of IRF3:IRF5 could avoid the unfavorable IRF5 binding that we observe if the 3'-end of the binding site was occupied by IRF3. In other words, heterodimerizing with IRF3 would allow IRF5 to bind the VREs by avoiding the non-optimal half-site sequence. Future studies with IRF heterodimers should further clarify these selection rules for IRF proteins for the IFNA VREs.

Finally, analyzing the role of DNA-binding affinity in IRF3/5/7-dependent gene regulation revealed clear affinity-independent mechanisms. We found that DNA sequence features of IRF binding sites could enhance their activity even at the expense of binding affinity. Comparison of sequence variants revealed that these differences did not need to occur in the core ISRE motif, but could be in the flanking bases. For example, the C-2 and I7 sites have identical 5'-CCGAAACCGAAA-3' core sequences yet very different IRF7 binding affinity (z-scores 17.0 and 9.5, respectively) and IRF7-dependent gene expression activity (6.1-fold higher for I7) (Figure 3.9). The mechanism of this uncoupling of affinity and activity remains unclear. However, a plausible mechanism is DNA-based allostery, in which IRF dimers adopt alternate protein conformations based on the DNA sequence of the binding sites and these structural differences affect gene activation. DNA-based allostery has been described for other factors, such as glucocorticoid receptors (Meijsing et al., 2009; Watson et al., 2013) and NF-κB (Wang et al., 2012), where in certain situations affinity and activity do not correlate. Future studies should clarify these details and provide a clearer picture of how affinity-dependent and affinity-independent mechanisms regulate IRF activity. The PBM dataset of IRF binding sites sequences and affinities generated here will provide an invaluable framework for dissecting the roles of affinity and activity.

Figure 3.1 - PBM-based Analysis of IRF3/5/7 DNA Binding.

(A) Schematic of IRF dimers and PBM probe design. **(B)** Phosphomimetic IRF variants used in our PBM experiments. Positions of Ser/Thr to Asp mutations are indicated. **(C)** PBM z-score distributions for IRF dimer binding to 10,044 synthetic SNV probes. Highlighted are z-scores for literature described IRF binding sites and a high-affinity consensus site bound by all three IRF dimers. **(D)** PBM-derived DNA-binding logos for IRF3/5/7.

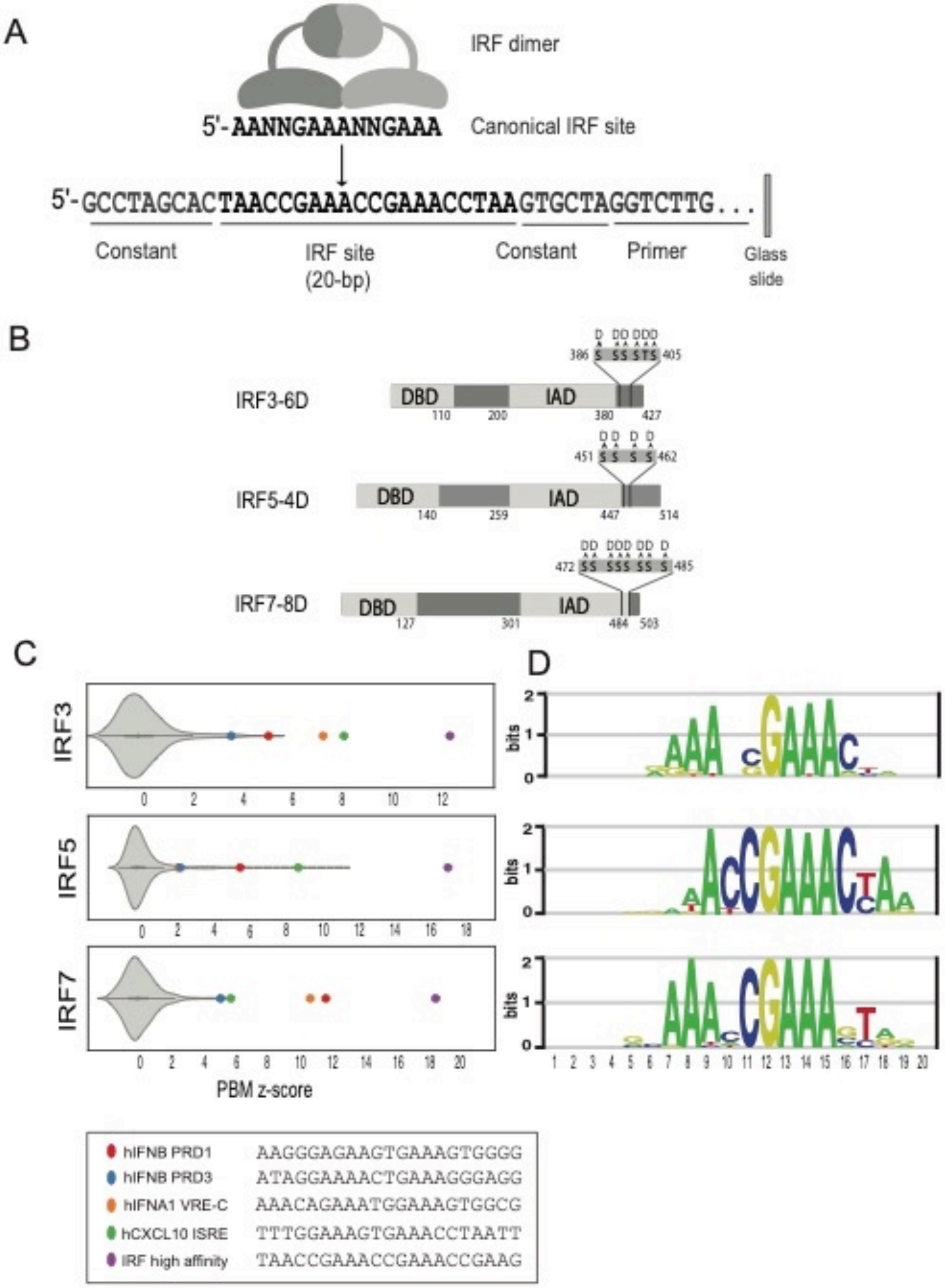

A
IRF dimer
5'-AANNGAAANNGAAA
Canonical IRF site
5'-GCCTAGCACTAACCGAAACCGAAACCTAAGTGCTAGGTCTTG...
Constant
IRF site
(20-bp)
Constant
Primer
Glass
slide
B
IRF3-6D
D DD DDD
386 S SS STS 405
DBD
IAD
110 200 380 427
IRF5-4D
DD D D
451 SS S S 462
DBD
IAD
140 259 447 514
IRF7-8D
DD DDD DD D
472 SS SSS SS S 485
DBD
IAD
127 301 484 503
C
IRF3
0 2 4 6 8 10 12
IRF5
0 2 4 6 8 10 12 14 16 18
IRF7
0 2 4 6 8 10 12 14 16 18 20
PBM z-score
D
bits
2 1 0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
● hIFNB PRD1 AAGGGAGAAGTGAAAGTGGGG
● hIFNB PRD3 ATAGGAAAACTGAAAGGGAGG
● hIFNA1 VRE-C AAACAGAAATGGAAAGTGGCG
● hCXCL10 ISRE TTTGGAAAGTGAAACCTAATT
● IRF high affinity TAACCGAAACCGAAACCGAAG

Figure 3.2 - Seed-sequence and Parameter-dependence of IRF DNA-binding Logos.

(A) IRF5 binding logos from HT-SELEX experiments and PBM are compared. PBM-derived IRF5 binding logos are shown for two separate β parameter values to illustrate the impact of this parameter on the binding logos. (B) IRF5 binding logos generated using our SNV approach (Chapter 2 - Methods) from select seed sequences are shown. DNA sequence of the starting seeds is indicated and the core half-site elements are in bold. Two separate types of logo (or motif) are identified and represented: a dominant motif for which base selectivity appears predominantly in one half-site, and an alternate motif that show base selectivity across the whole ISRE.

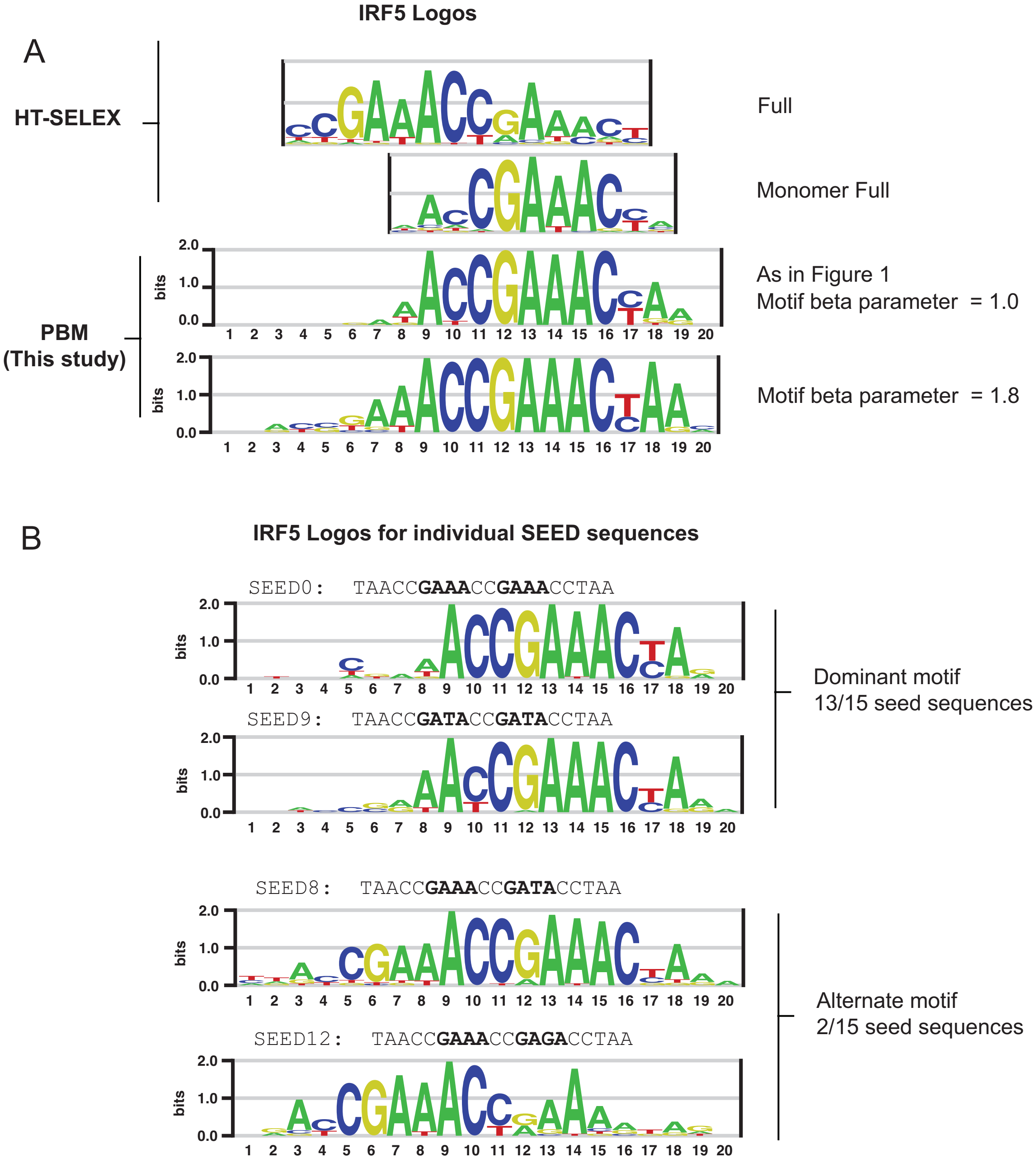

IRF5 Logos
A
HT-SELEX
Full
Monomer Full
PBM
(This study)
As in Figure 1
Motif beta parameter = 1.0
Motif beta parameter = 1.8
bits
2.0
1.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
B
IRF5 Logos for individual SEED sequences
SEED0: TAACCGAAACCGAAACCTAA
SEED9: TAACCGATACCGATACCTAA
Dominant motif
13/15 seed sequences
SEED8: TAACCGAAACCGATACCTAA
SEED12: TAACCGAAACCGAGACCTAA
Alternate motif
2/15 seed sequences

Figure 3.3 - DNA-binding of IRF3/5/7 to Select ISREs.

EMSAs for IRF3, IRF5 and IRF7 on select DNA sequences from our PBM experiments. PBM z-scores for the DNA sequences used in EMSA experiments are included. Estimated fold-differences in binding affinity are indicated for different DNA sequences. DNA probe concentrations of 1 nM were used in all experiments. The C-2 PBM probe is bound with high-affinity by IRF3, IRF5 and IRF7 compared to IFNβ PRD3 element. Refer to Table 2.2 for EMSA probe sequence.

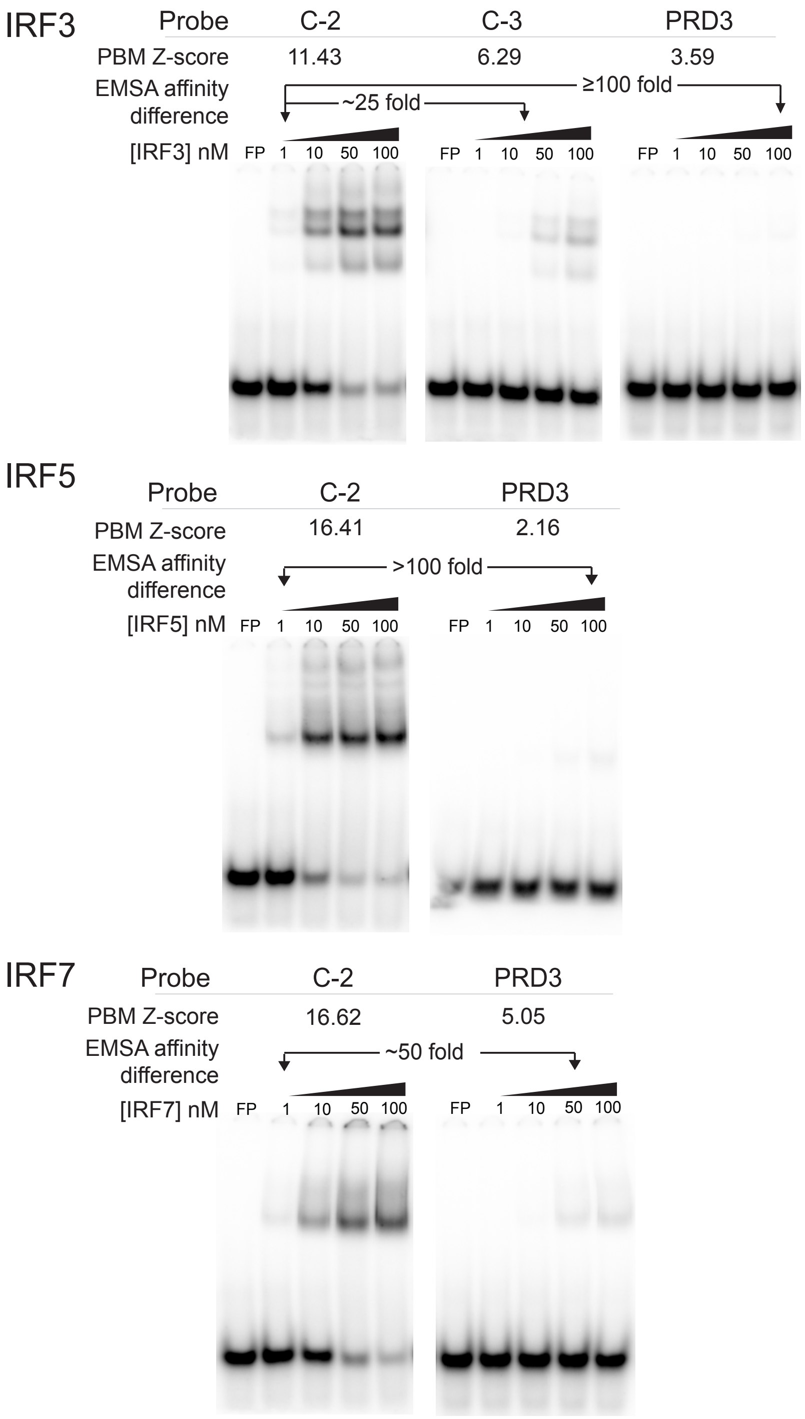

IRF3
Probe
C-2
C-3
PRD3
PBM Z-score
11.43
6.29
3.59
EMSA affinity difference
~25 fold
≥100 fold
[IRF3] nM
FP 1 10 50 100
FP 1 10 50 100
FP 1 10 50 100
IRF5
Probe
C-2
PRD3
PBM Z-score
16.41
2.16
EMSA affinity difference
>100 fold
[IRF5] nM
FP 1 10 50 100
FP 1 10 50 100
IRF7
Probe
C-2
PRD3
PBM Z-score
16.62
5.05
EMSA affinity difference
~50 fold
[IRF7] nM
FP 1 10 50 100
FP 1 10 50 100

Figure 3.4 - IRF3/5/7 DNA-binding Preferences.

(A) Pairwise comparison of z-scores for IRF3, IRF5 and IRF7 binding to 10,044 synthetic IRF sites (black dots) and 644 random background sequences (blue dots). (**B)** Pairwise comparison of PBM z-scores (same sequences as in (A)) for IRF7 performed at two single concentrations (i.e., pseudo-replicate experiments). **(C)** Shown are the same pairwise comparison data as in (A) with all binding sequences matching the pattern 5'-GATANNGATA-3' (N is any base) highlighted in red.

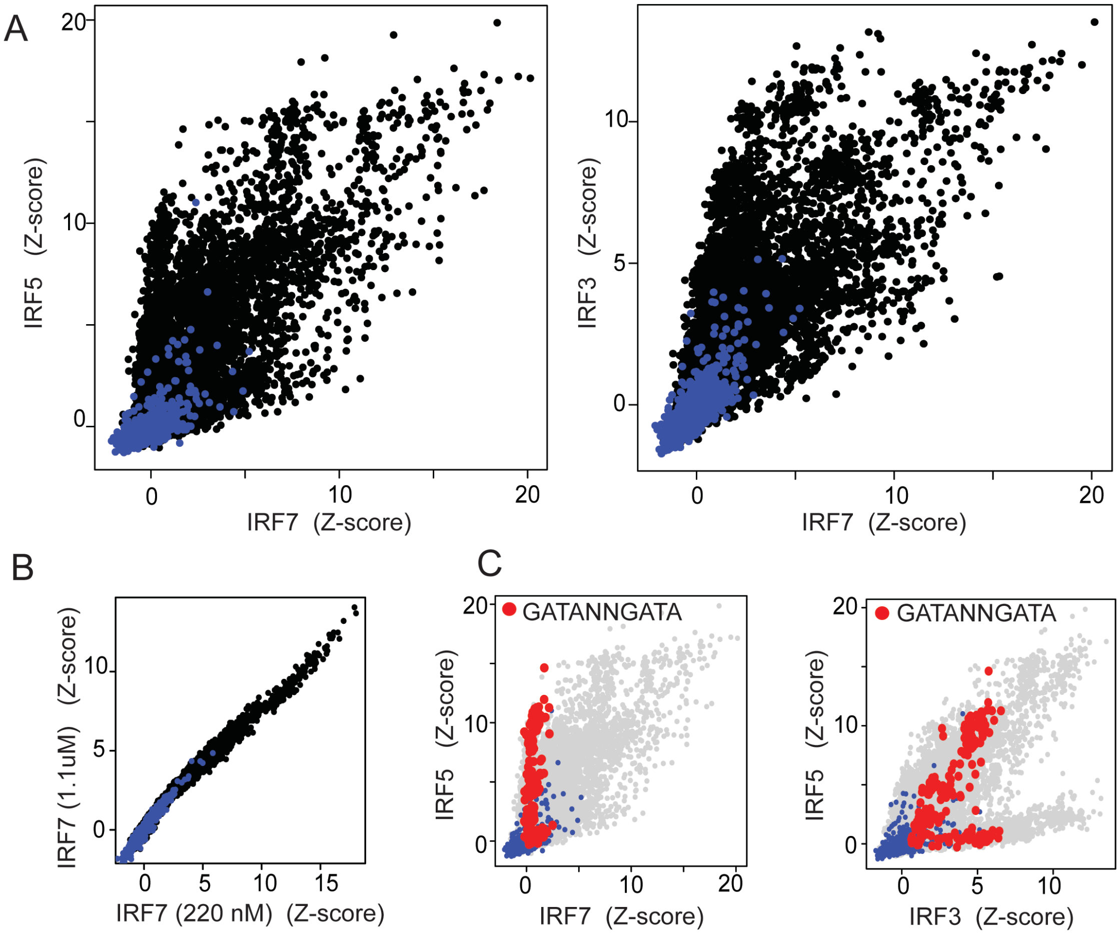


Figure 3.5 - IRF3/5/7 Binding to ISREs with Alternate Half-site Spacing.

(A) Binding of IRF3/5/7 to binding sites found in the PRDI and PRDIII elements of the IFNB promoter. Two sites (P1, P2) have a 2-bp spacer between GAAA-type half-sites (2-bp sites); one site (P3) has a 3-bp spacer. Error bars show the standard deviation of PBM-based z-scores to 5 replicate DNA probe sequences (z-score for each individual sequence was determined by fitting to experiments at different concentrations as described in Materials and Methods). **(B)** Differential binding (i.e., delta z-scores) of IRF3/5/7 to identical sets of 120 sequence-matched 2- and 3-bp sites that differ by a single base (as shown in schematic). A schematic is shown illustrating how the 'delta z-score' for each sequence-matched pairs is calculated. The delta z-score distribution for all 120 sequence pairs is shown for IRF3/5/7. **(C)** EMSAs are shown for IRF3 binding to 3 DNA site variants: (Monomer) — a single ISRE half-site (i.e., 5'-AANNGAAA-3'); (Proximal dimer) — site with a 2-bp spacer between half-sites (2-bp site); (Extended dimer) — site with an 8-bp spacer between half-site (8-bp site). **(D)** Schematic illustrating the proposed binding arrangement of an IRF dimer to the site variants.

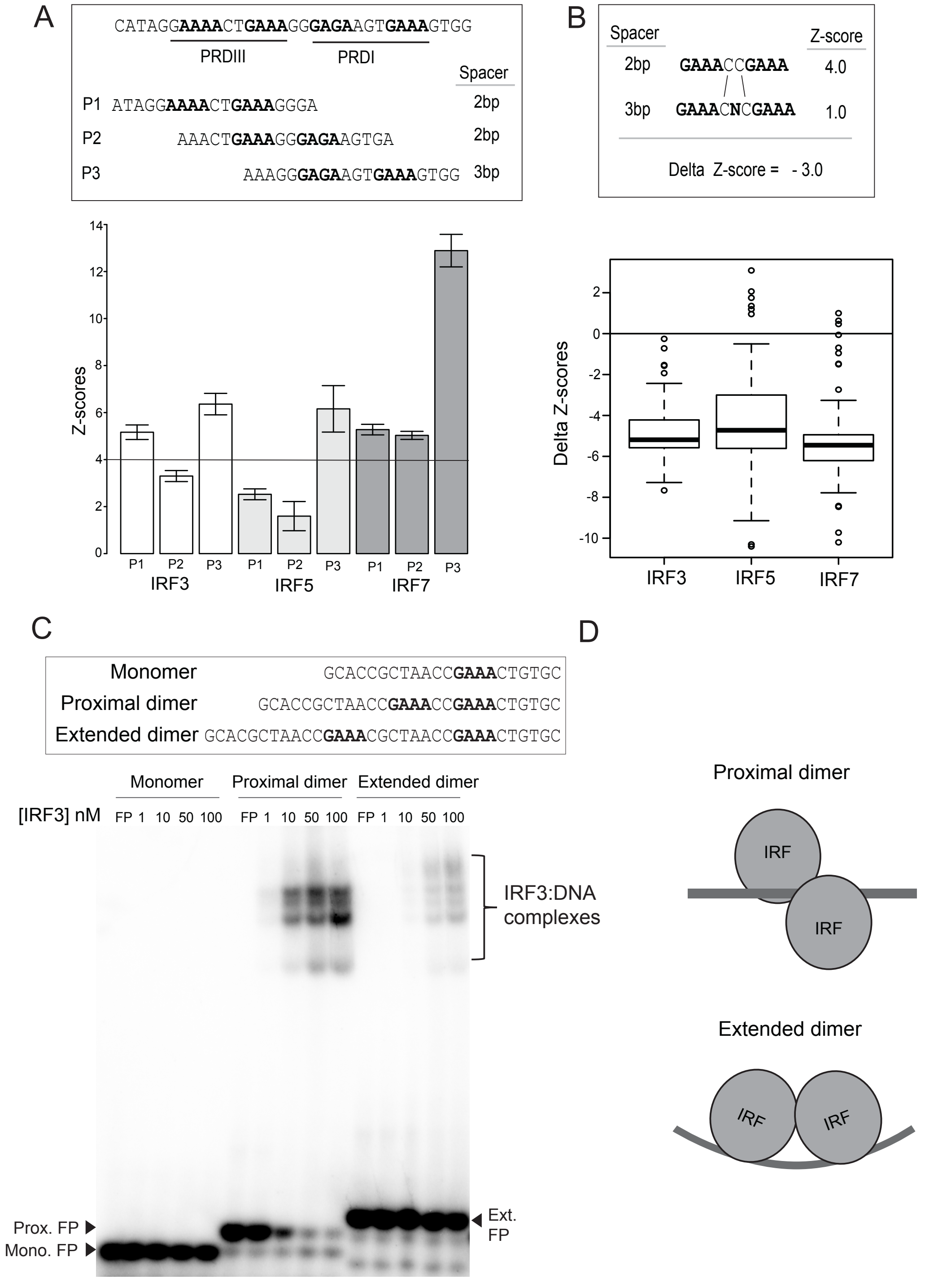

A
CATAGGAAAACTGAAAGGGAGAAGTGAAAGTGG
PRDIII
PRDI
Spacer
P1 ATAGGAAAACTGAAAGGGA 2bp
P2 AAACTGAAAGGGAGAAGTGA 2bp
P3 AAAGGGAGAAGTGAAAGTGG 3bp
Z-scores
0
2
4
6
8
10
12
14
P1 P2 P3 P1 P2 P3 P1 P2 P3
IRF3
IRF5
IRF7
B
Spacer
Z-score
2bp GAAACCGAAA 4.0
3bp GAAACNCGAAA 1.0
Delta Z-score = - 3.0
Delta Z-scores
2
0
-2
-4
-6
-8
-10
IRF3
IRF5
IRF7
C
Monomer GCACCGCTAACCGAAACTGTGC
Proximal dimer GCACCGCTAACCGAAACCGAAACTGTGC
Extended dimer GCACGCTAACCGAAACGCTAACCGAAACTGTGC
Monomer
Proximal dimer
Extended dimer
[IRF3] nM FP 1 10 50 100 FP 1 10 50 100 FP 1 10 50 100
IRF3:DNA complexes
Prox. FP
Mono. FP
Ext. FP
D
Proximal dimer
IRF
IRF
Extended dimer
IRF
IRF

Figure 3.6 - Wild-type and Mutant IRF Dimer Binding to IFNA VREs.

(A) Binding profiles of wild-type and mutant IRFs to human IFNA VREs. Binding sites are illustrated as 10-bp blocks in register with the canonical 5-GAAANNGAAA-3' elements of an IRF binding site. Low-affinity binding sites (z-score < 4.0) are not shown (see chapter 2). Z-scores for each panel range from 0 to 13.0. IFNA gene names are shown. Coordinates for VREs elements are available upon request**.** **(B)** Binding profiles to mouse IFNA VREs are shown, all details as in (A). **(C)** Binding profiles of wild-type IRF3/5/7 to known IRF-binding regulatory elements are shown. Profile details as in (A,B). Coordinates for regulatory element DNA sequences are available upon request**.** **(D)** DNA-binding logos for wild-type and mutant IRF5/7 are shown. Wild-type logos are identical to those in Figure 1 and are shown for contrast. Regions in which the DNA base specificity is altered in the mutant IRFs are highlighted with red shading. DNA logos generated as described in Materials and Methods (i.e., SNV approach). **(E)** Multiple protein-sequence alignment for human and mouse IRF3/5/7 is shown. Alignment is limited to the portion of protein containing alpha helices 2 and 3 that contains the base-contacting residues that match our criteria to be specificity determining (see main text). Identically conserved residues across all IRFs are indicated with a ' * '; residues in which IRF3/5/7 are different (but conserved across species) are indicated with a ' **|** '. Putative specificity-determining residues are highlighted with shaded bars; our selected specificity-determining residue is highlighted in red.

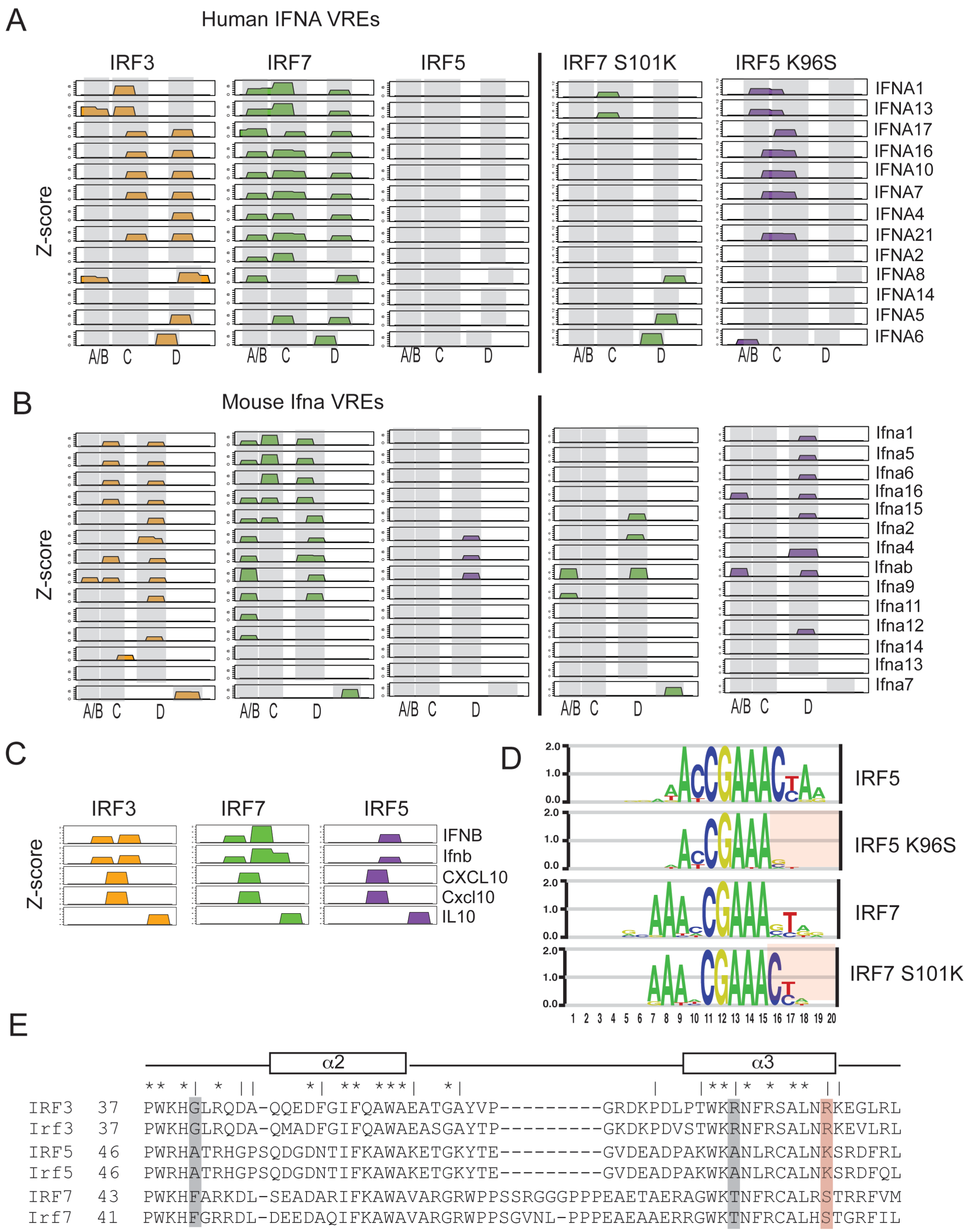
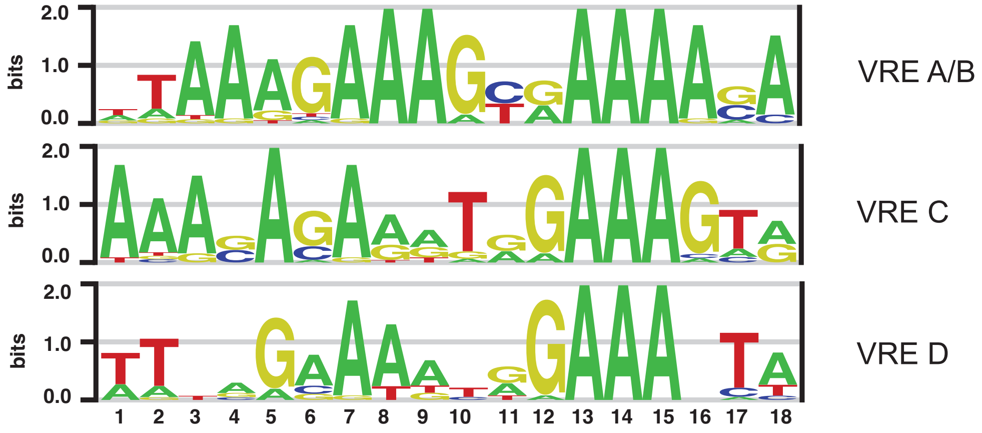

A
Human IFNA VREs
IRF3
IRF7
IRF5
IRF7 S101K
IRF5 K96S
Z-score
IFNA1
IFNA13
IFNA17
IFNA16
IFNA10
IFNA7
IFNA4
IFNA21
IFNA2
IFNA8
IFNA14
IFNA5
IFNA6
A/B C D
B
Mouse Ifna VREs
Z-score
Ifna1
Ifna5
Ifna6
Ifna16
Ifna15
Ifna2
Ifna4
Ifnab
Ifna9
Ifna11
Ifna12
Ifna14
Ifna13
Ifna7
A/B C D
C
IRF3
IRF7
IRF5
Z-score
IFNB
Ifnb
CXCL10
Cxcl10
IL10
D
2.0
1.0
0.0
IRF5
IRF5 K96S
IRF7
IRF7 S101K
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
E
α2
α3
** *| * || *| ** ***| *| | |**|* * ** ||
IRF3 37 PWKHGLRQDA-QQEDFGIFQAWAEATGAYVP---------GRDKPDLPTWKRNFRSALNRKEGLRL
Irf3 37 PWKHGLRQDA-QMADFGIFQAWAEASGAYTP---------GKDKPDVSTWKRNFRSALNRKEVLRL
IRF5 46 PWRHATRHGPSQDGDNTIFKAWAKETGKYTE---------GVDEADPAKWKANLRCALNKSRDFRL
Irf5 46 PWRHATRHGPSQDGDNTIFKAWAKETGKYTE---------GVDEADPAKWKANLRCALNKSRDFQL
IRF7 43 PWKHFARKDL-SEADARIFKAWAVARGRWPPSSRGGGPPPEAETAERAGWKTNFRCALRSTRRFVM
Irf7 41 PWKHFGRRDL-DEEDAQIFKAWAVARGRWPPSGVNL-PPPEAEAAERRGWKTNFRCALHSTGRFIL

Figure 3.7 - DNA-binding Logos Compiled for ISREs from IFNA VREs.

Logos representations are shown for all binding sites bound by IRF3, IRF5 or IRF7 in mouse and human IFNA VREs (as in Figure 3.6). Logos were compiled independently for the sites found in VRE-A/B (20 sites), VRE-C (19 sites) and VRE-D (22 sites), as shown in Figure 3.5. Logos for sites from mouse and human VREs were compiled together.

IFNA VREs Logos



IFNA VREs Logos

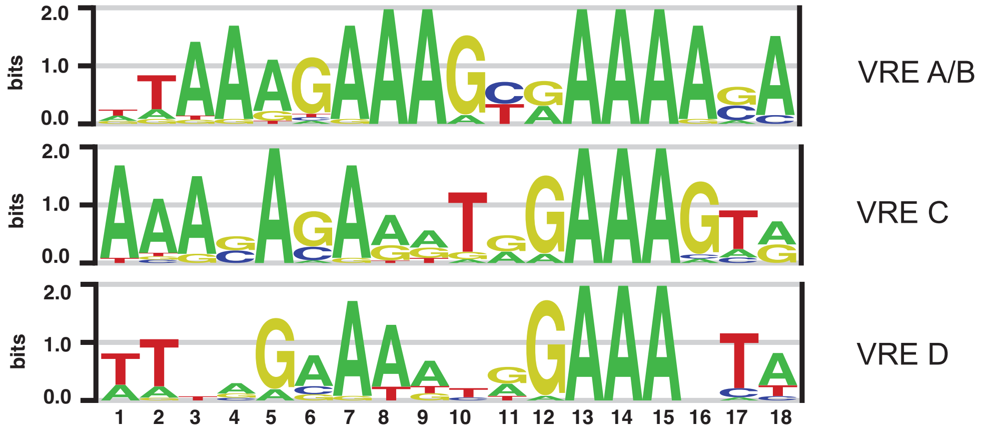


Figure 3.8 - IRF Protein Levels in Transfected HEK293T Cells.

Phosphomimetic IRF protein overexpression levels were assessed using dual-color fluorescent western blot analysis. 30 μg of whole-cell lysate was loaded per condition, with 25ng of purified IRF3/5/7 protein as a positive control. HEK293T plates were PEI transfected as described in the main methods section with 7.2 μg of plasmid DNA consisting of: pGL4.54 TK-Luciferase [2.48 μg]; E1α-eGFP [1.99 μg]; pDEST26 IRF3/5/7 or additional E1α-eGFP [0.25 μg]; pGEM3zf(-) carrier DNA [2.48 μg], these quantities are scaled up from 96-well plates assuming a 20μg typical 10cm plate transfection. (A) IRF3 immunodetection shows the overexpression of His-IRF3-6D protein as well as low ubiquitous levels of endogenous IRF3 present in all conditions. (B) IRF5 immunodetection shows an absence of IRF5 in HEK293T cells and a strong overexpression of His-IRF5-4D protein. (C) IRF7 immunodetection shows basal expression of IRF7 in HEK293T cells; however, the levels of endogenous IRF7 are similar in all transfection conditions with an increase in IRF7 signal in the His-IRF7-6D overexpression condition.

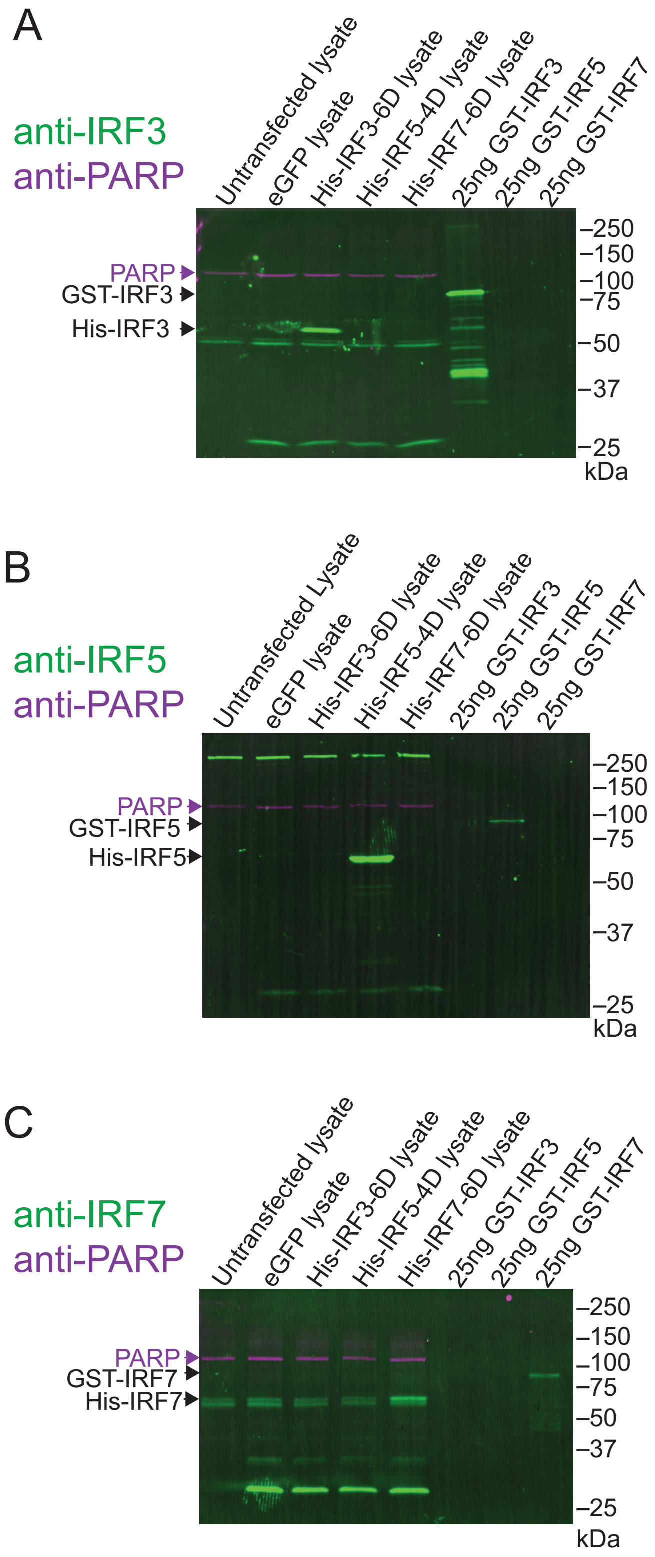

A
anti-IRF3
anti-PARP
Untransfected lysate
eGFP lysate
His-IRF3-6D lysate
His-IRF5-4D lysate
His-IRF7-6D lysate
25ng GST-IRF3
25ng GST-IRF5
25ng GST-IRF7
PARP
GST-IRF3
His-IRF3
–250
–150
–100
–75
–50
–37
–25
kDa
B
anti-IRF5
anti-PARP
Untransfected Lysate
eGFP lysate
His-IRF3-6D lysate
His-IRF5-4D lysate
His-IRF7-6D lysate
25ng GST-IRF3
25ng GST-IRF5
25ng GST-IRF7
PARP
GST-IRF5
His-IRF5
–250
–150
–100
–75
–50
–37
–25
kDa
C
anti-IRF7
anti-PARP
Untransfected lysate
eGFP lysate
His-IRF3-6D lysate
His-IRF5-4D lysate
His-IRF7-6D lysate
25ng GST-IRF3
25ng GST-IRF5
25ng GST-IRF7
PARP
GST-IRF7
His-IRF7
–250
–150
–100
–75
–50
–37
–25
kDa

Figure 3.9 - Comparison of DNA Binding and Gene Activation by ISRE Variants

(A) Schematic for promoter variants based on IFNA14 VRE. Promoter variants contain simultaneous alterations of C-site and D-site with the binding site shown in the box. Wild-type C-site (WT-C) is shown for context. **(B, D)** PBM-determined z-scores are shown for IRF3/5/7 to binding site sequence variants listed in (A). Error bars are calculated as in Figure 3. **(C, E)** Reporter gene activation levels for promoter variants in HEK293T cells over-expressing the phosphomimetic, constitutively active IRF dimers: IRF3(6D), IRF5(4D) and IRF7(6D). Normalized fold-induction (relative to I7-2 promoter levels) is shown for each promoter. Mean and standard error of the mean (sem) values were calculated over at least 15 replicate measurements**.**

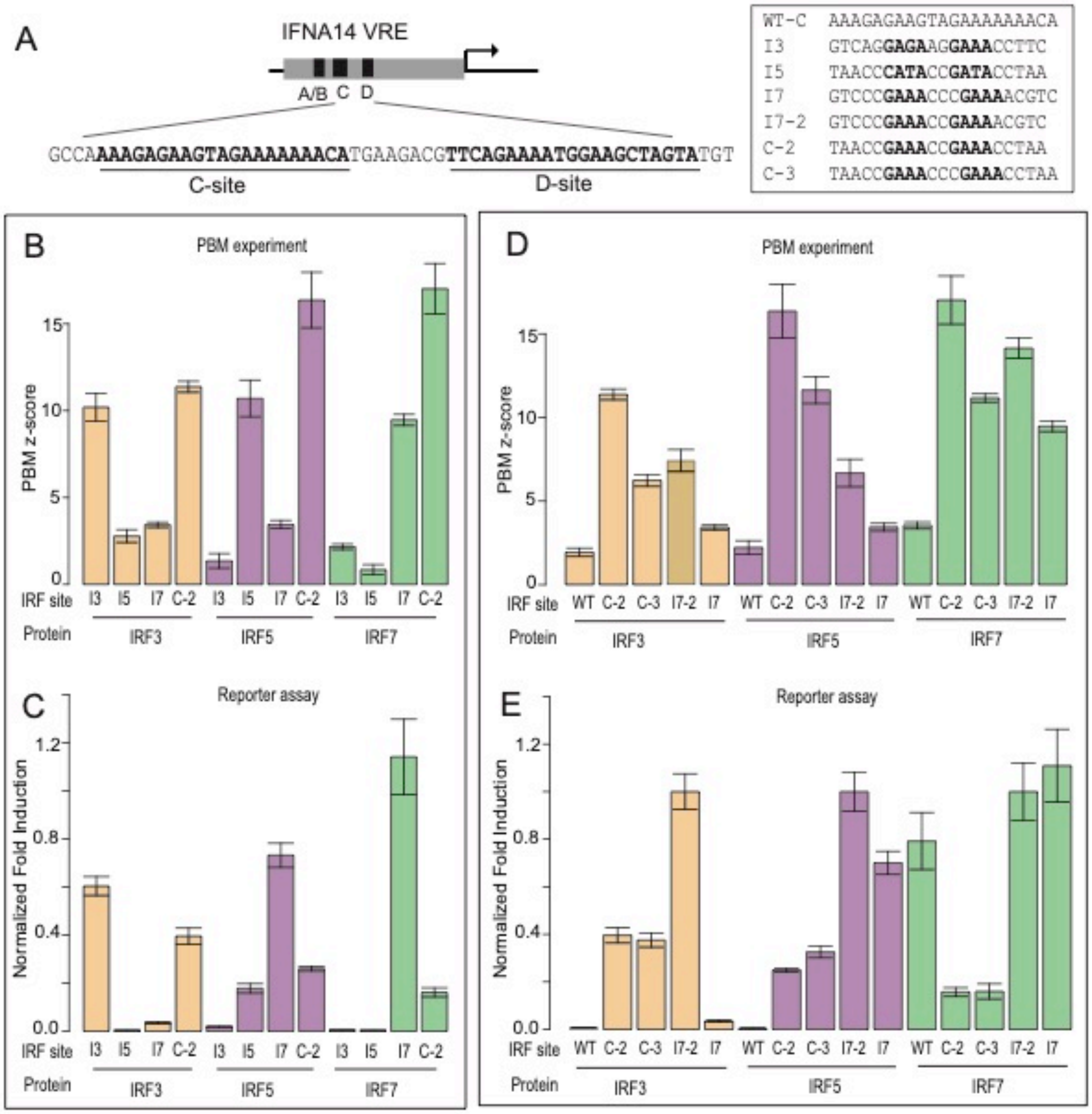


Figure 3.10 – Reporter assay luciferase (pGL4.54) plots

Bar plots showing the mean of the luciferase (constitutive transfection control plasmid) data. Error bars show standard deviation.

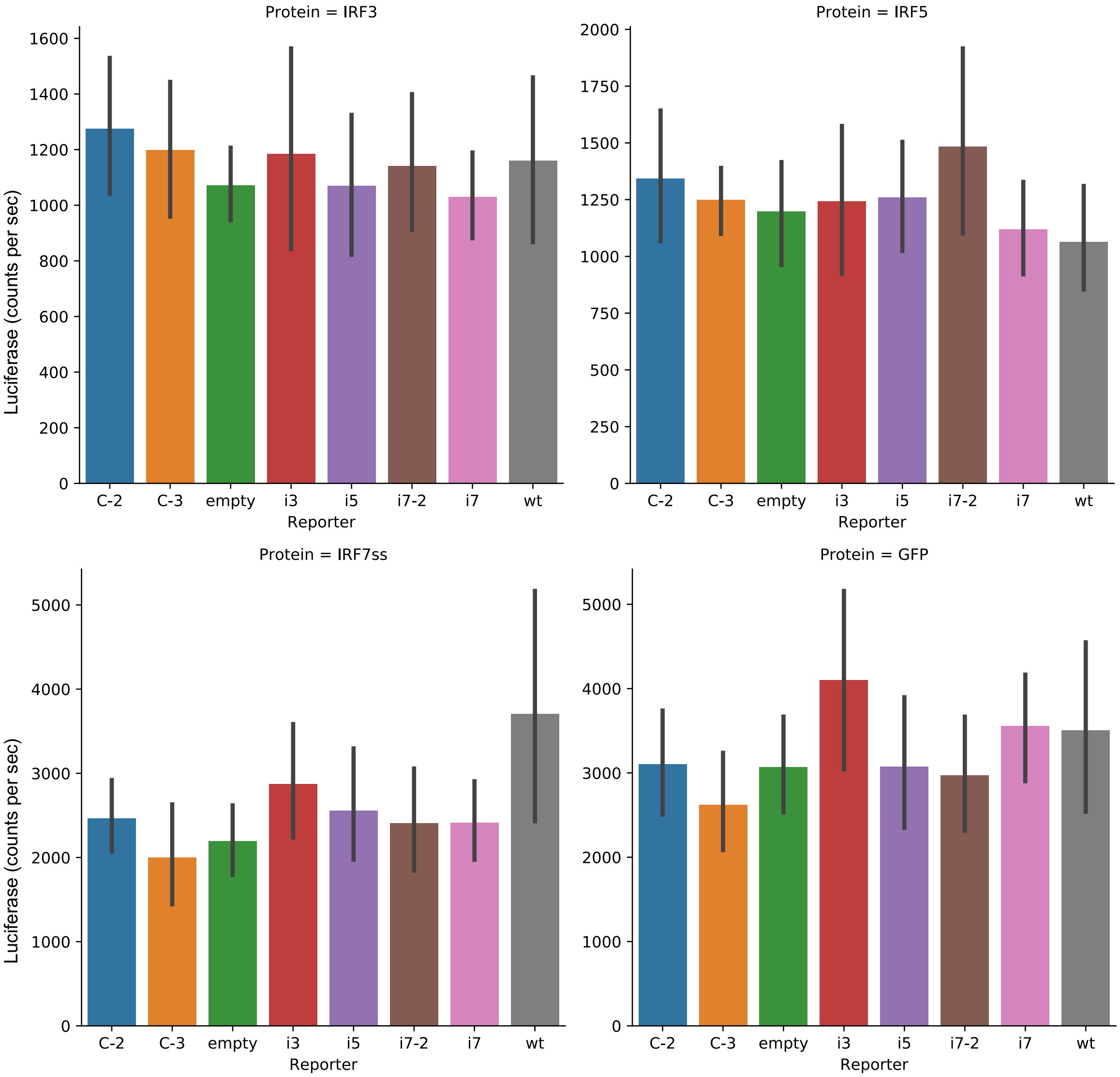


Figure 3.11 – Reporter assay nano luciferase plots

Bar plots showing mean nano luciferase data (experimental reporter plasmid), error bars show standard deviation.

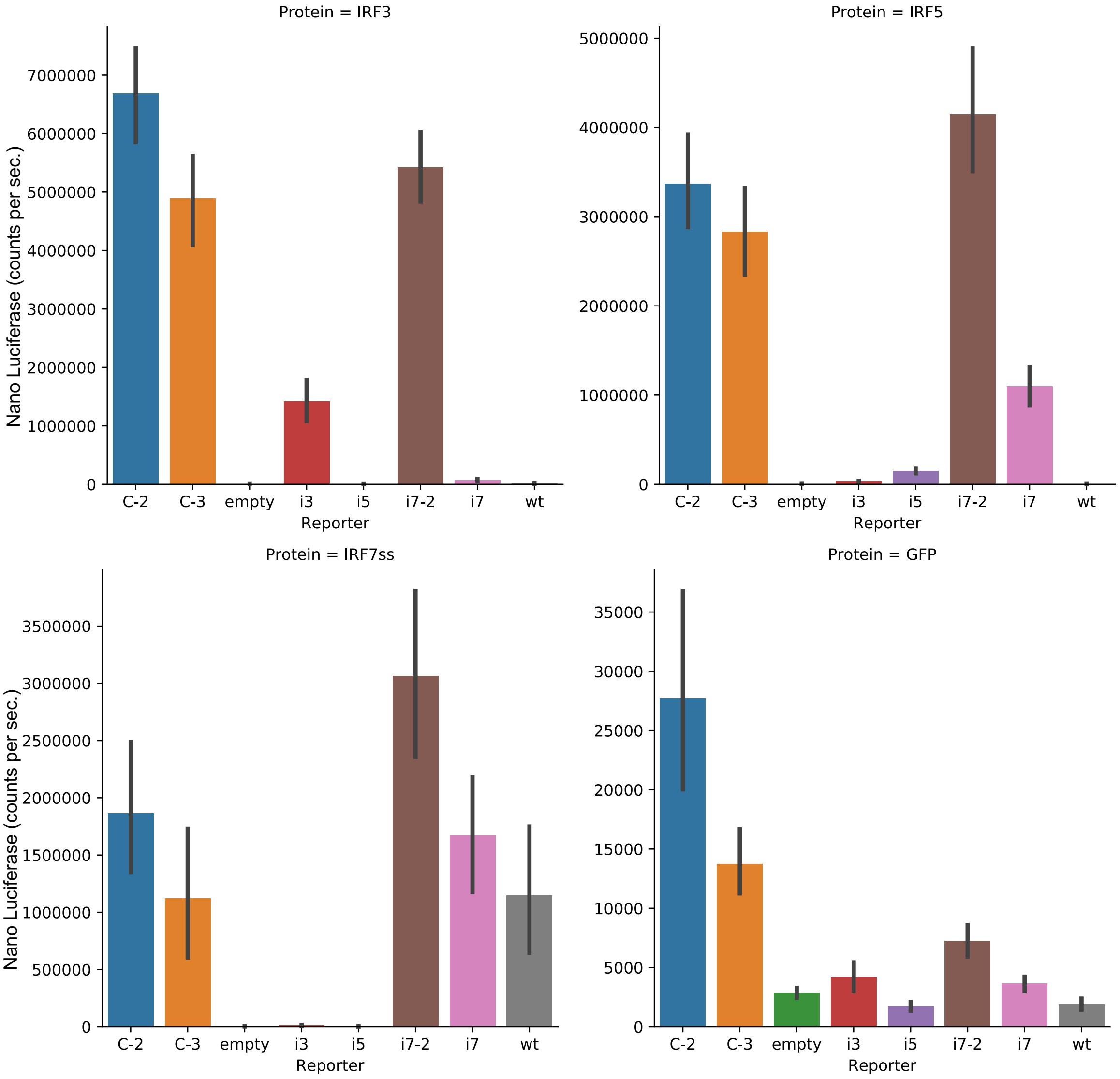


Figure 3.12 – Reporter assay transfection normalized data plots (Norm)

Bar plots showing reporter assay data normalized to the luciferase constitutive transfection control plasmid, error bars represent standard deviation. C-2 and C-3 constructs in the GFP control condition show elevated background expression levels. This may be due to other transcription factors utilizing those sites. Final reporter assay data was normalized to the GFP signal to account for this background expression. (see Figure 3.13)

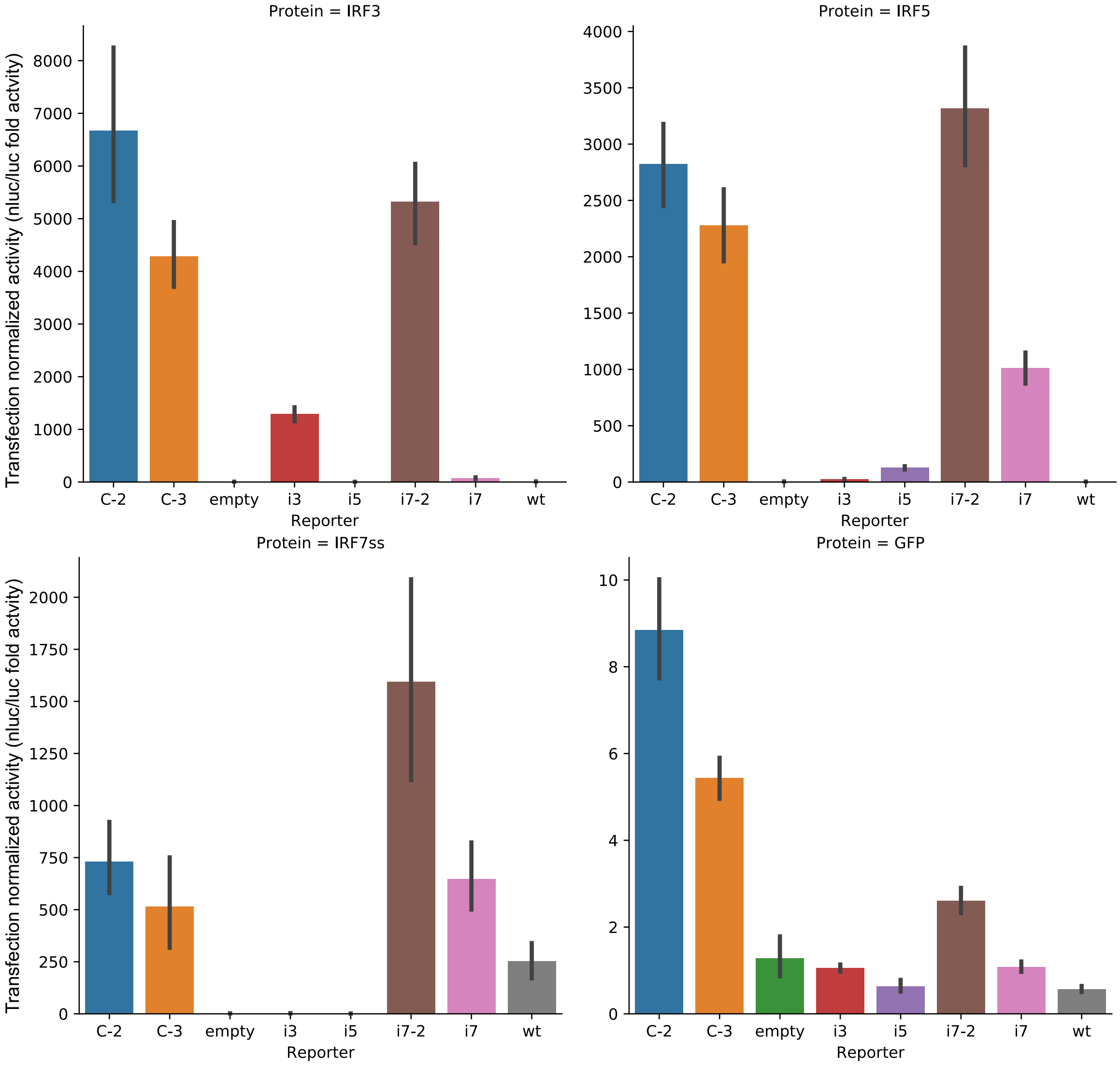


Figure 3.13 – Reporter assay GFP normalized fold-induction plot (gfp_norm)

Bar plots show reporter assay activity relative to GFP (no IRF) expression, error bars represent standard deviations. GFP expression condition was used to place the cells under transcriptional/translational load to control for the impact of exogenous protein expression.

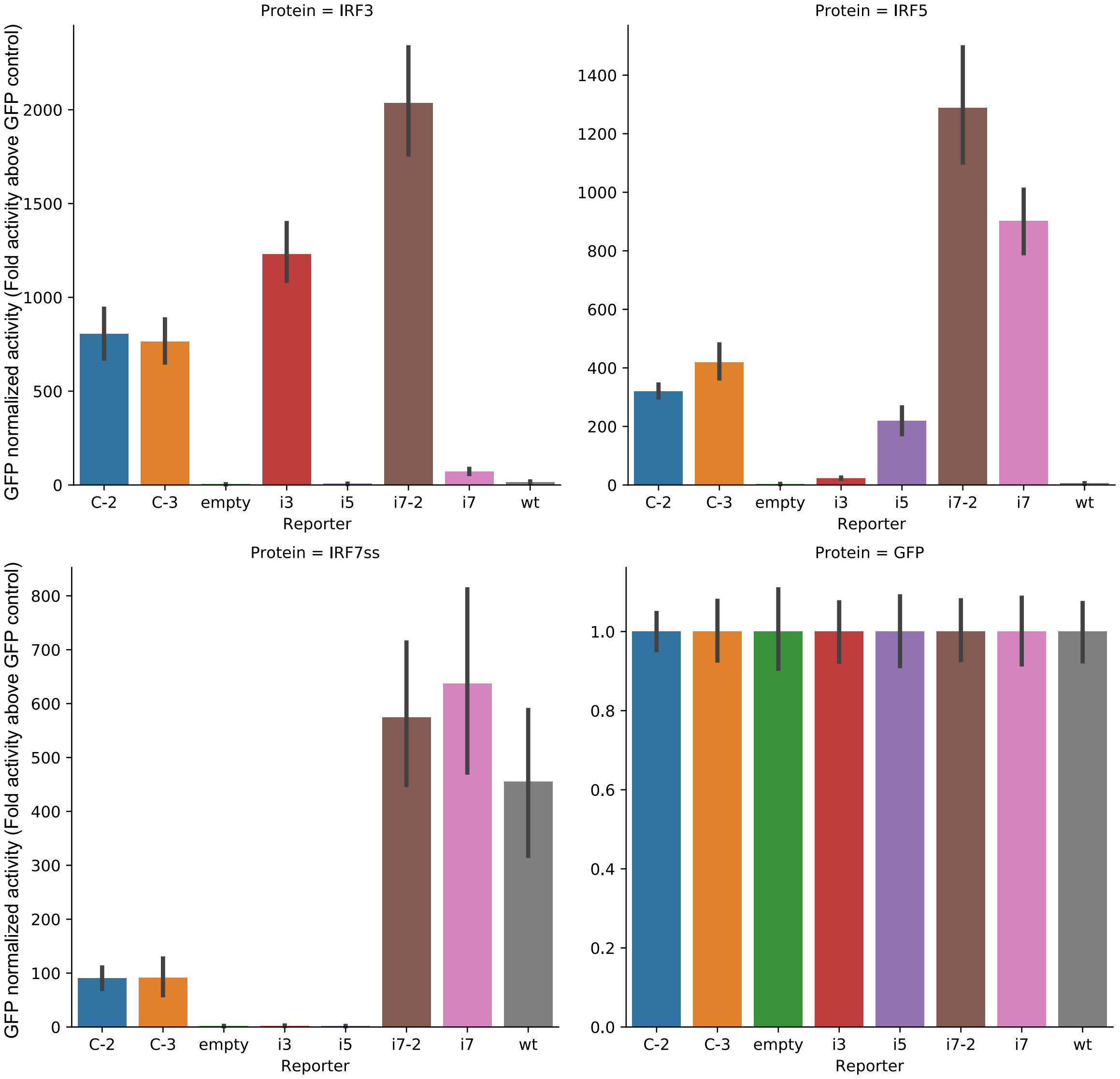


Figure 3.14 – Reporter assay data scaled to i7-2 reporter condition

Bar plots show reporter assay data scaled to each protein condition's respective i7-2 activity, error bars represent standard deviation. This aesthetic scaling was used to simplify the final plots in Figure 3.9 without altering the relationships between data within a protein condition.

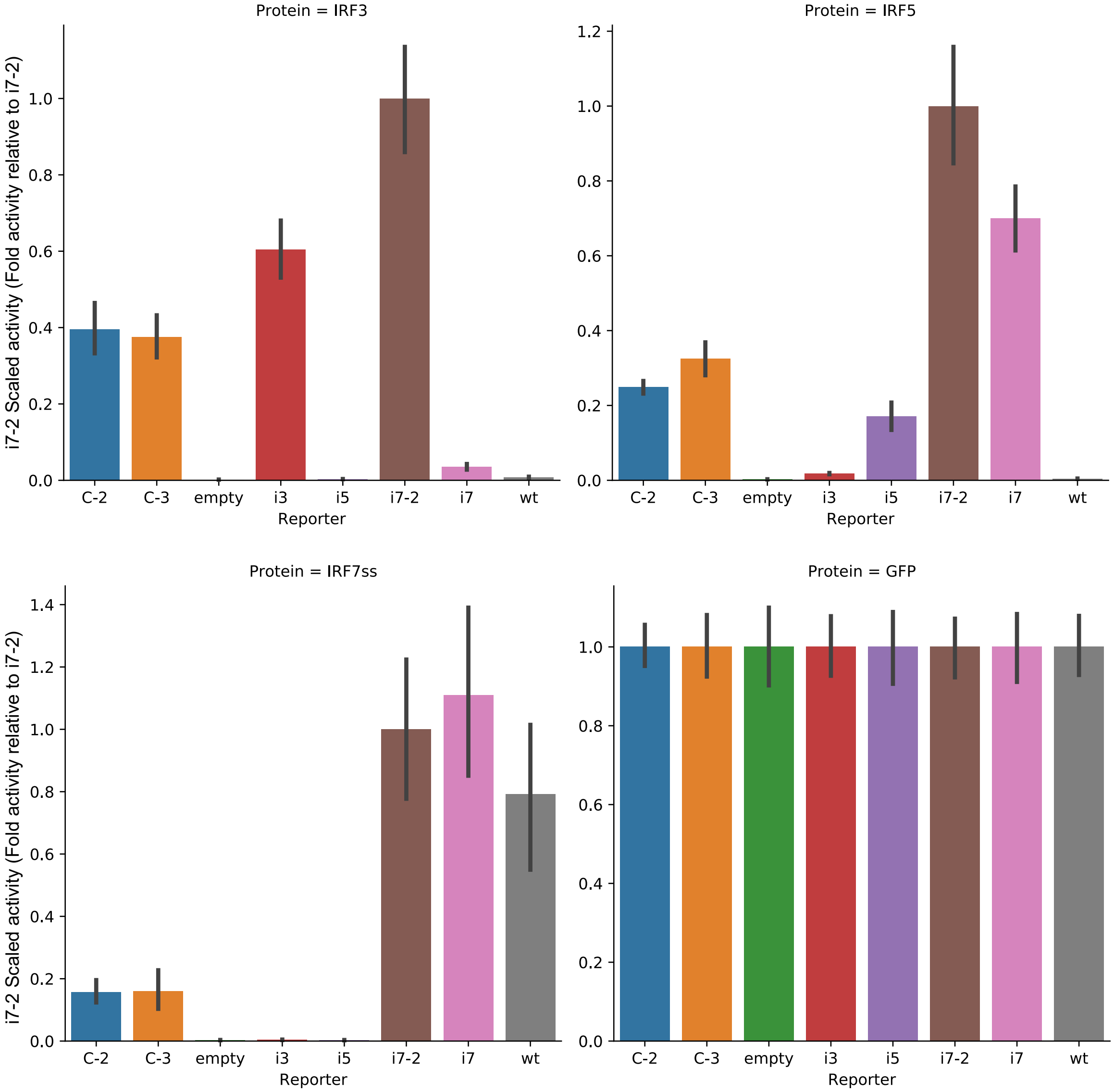


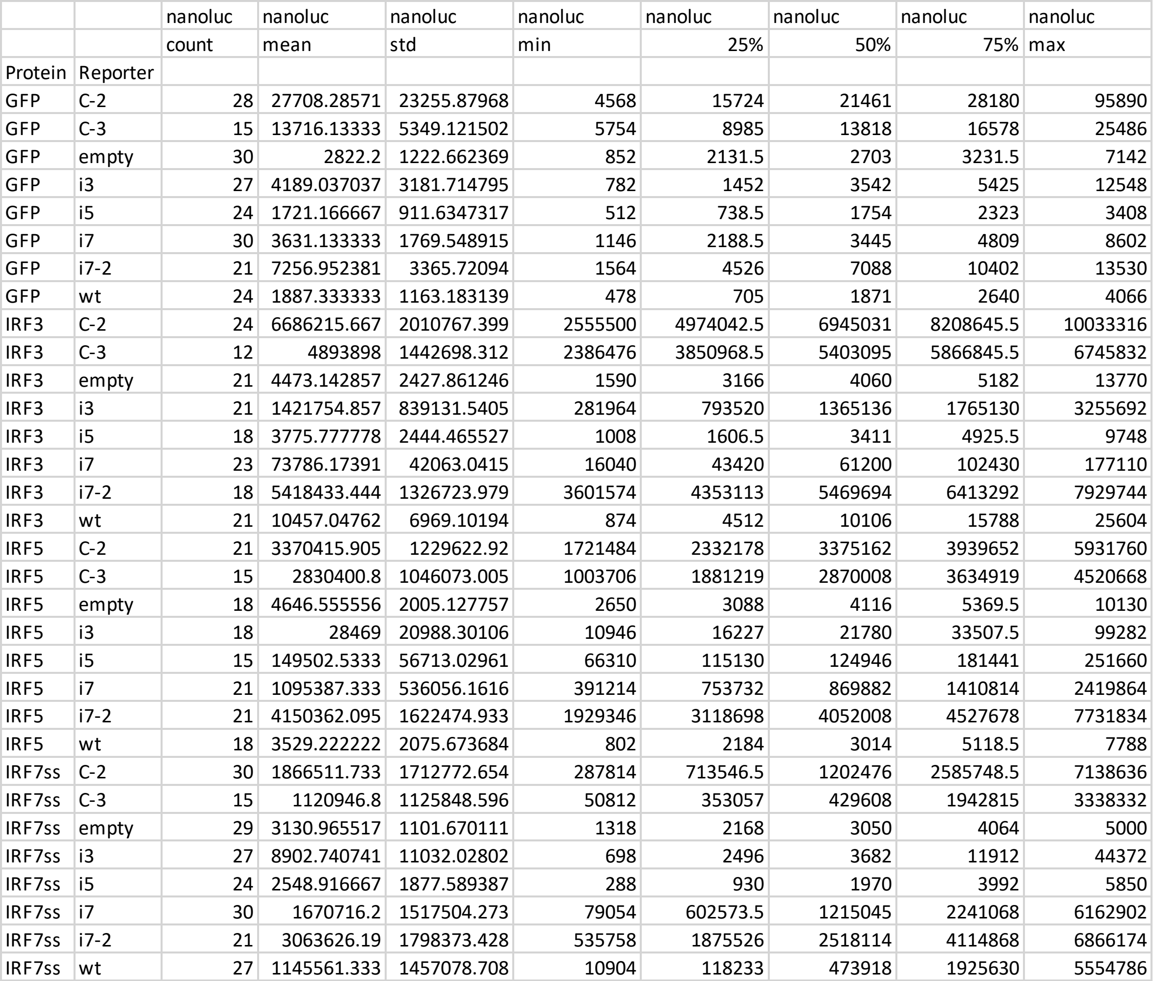
Table 3.1 – Reporter assay luciferase data descriptive statistics

Table includes descriptive statistics of raw luciferase values (transfection control plasmid) for each Protein by Reporter condition. Values are in counts per minute as measured on the Victor-3 luminometer (see Chapter 2 – Materials and Methods)

		luc	luc	luc	luc	luc	luc	luc	luc
		count	mean	std	min	25%	50%	75%	max
Protein	Reporter								
GFP	C-2	28	3103.714286	1692.09077	558	1725.75	2840.5	3838.25	6536
GFP	C-3	15	2620.466667	1146.293896	1118	1668.5	2464	3340.5	4650
GFP	empty	30	3071.4	1532.397324	617	2077.75	2867	3973	6796
GFP	i3	27	4100.518519	2865.232054	624	1376.5	4214	5895	12500
GFP	i5	24	3076.5	2010.252699	889	1500.75	2769.5	3995	8496
GFP	i7	30	3556.133333	1842.746577	1202	2128	2955.5	4863.75	7310
GFP	i7-2	21	2968.52381	1586.142447	490	1736	2830	3949	6322
GFP	wt	24	3502.833333	2546.467659	1060	1452	2523.5	4644	9204
IRF3	C-2	24	1275.541667	629.6927463	356	829	1124	1442	2836
IRF3	C-3	12	1199.416667	447.7705294	682	826.25	1106	1479.75	1994
IRF3	empty	21	1072.190476	307.2556947	538	863	1069	1250	1689
IRF3	i3	21	1184.952381	842.2655446	238	544	940	1447	2855
IRF3	i5	18	1069.944444	595.0555557	383	617.5	854.5	1666	1943
IRF3	i7	23	1030.347826	384.0456423	337	771.5	1012	1218.5	1857
IRF3	i7-2	18	1141.833333	560.3042503	582	743.75	975.5	1328.75	2728
IRF3	wt	21	1161.285714	712.6125976	183	515	1074	1648	2365
IRF5	C-2	21	1342.52381	705.7898143	421	945	1146	1778	2883
IRF5	C-3	15	1249.533333	320.5100548	619	1070	1317	1454.5	1716
IRF5	empty	18	1198.388889	523.3633468	521	756.75	1125	1671.25	2098
IRF5	i3	18	1242.722222	740.5358001	423	615.5	1288.5	1604.5	2794
IRF5	i5	15	1261.2	497.5959635	459	869	1236	1702.5	1942
IRF5	i7	21	1119.904762	465.3776858	399	806	1059	1293	2203
IRF5	i7-2	21	1484.809524	1023.100465	592	885	1129	1651	4321
IRF5	wt	18	1064.944444	514.3484139	365	694.75	873	1443.75	2121
IRF7ss	C-2	30	2463.166667	1261.822521	914	1598	2132	3192.5	5672
IRF7ss	C-3	15	1999.266667	1196.863906	705	1175.5	1772	2244.5	4950
IRF7ss	empty	29	2195.172414	1134.396538	625	1306	2033	3120	4496
IRF7ss	i3	27	2872.37037	1803.820713	441	1450	2325	3968.5	7901
IRF7ss	i5	24	2558.25	1673.31967	498	1418.25	1980	3028.5	7067
IRF7ss	i7	30	2415.233333	1341.68334	399	1619.75	2023.5	3069.75	6425
IRF7ss	i7-2	21	2405.190476	1494.838574	1018	1350	1601	2808	5708
IRF7ss	wt	27	3702.925926	3695.934522	583	1452	2222	3763.5	15117

Table 3.2 – Reporter assay Nano Luciferase descriptive statistics

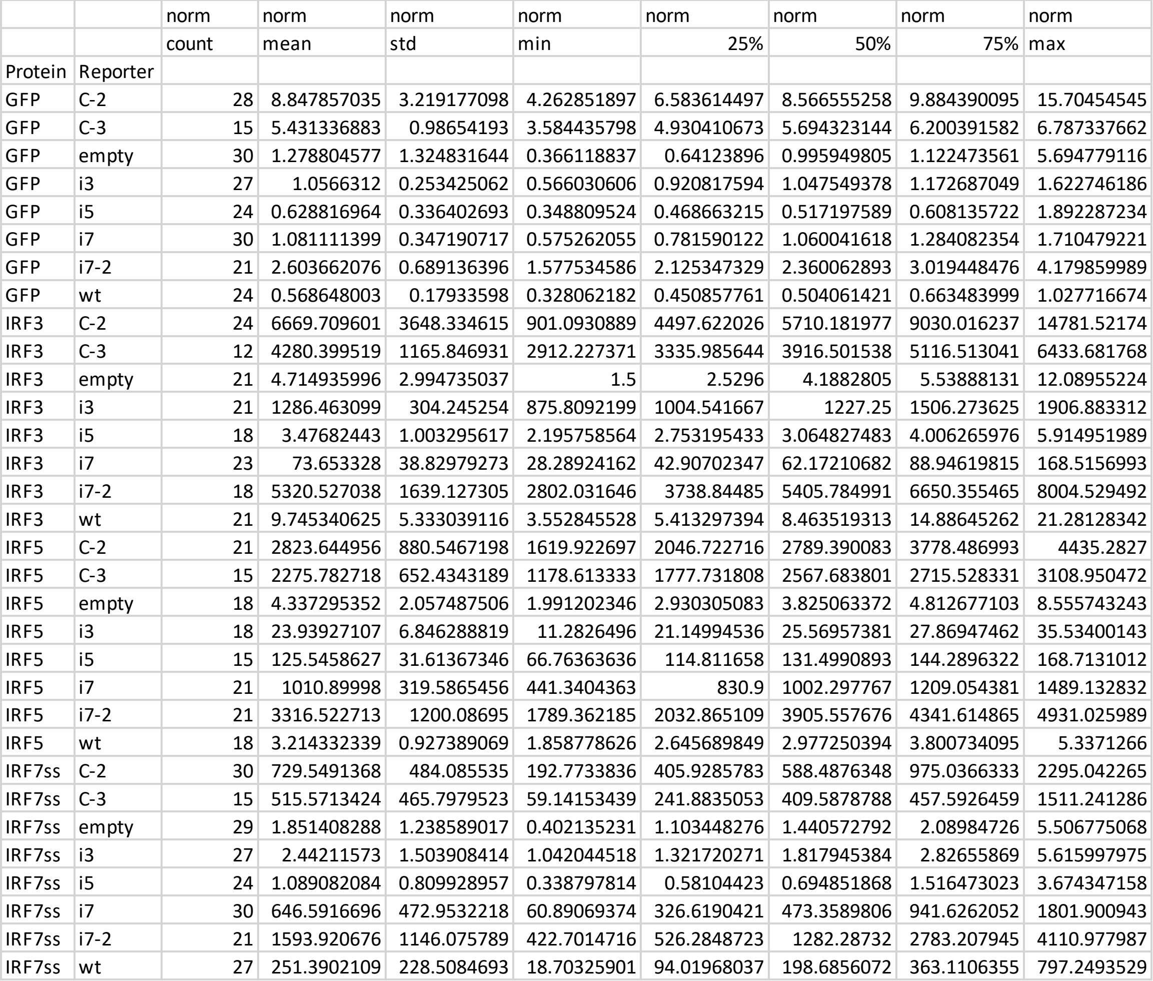
Table includes descriptive statistics of raw nano-luciferase (experimental plasmid) values for each Protein by Reporter condition. Values are in counts per minute as measured on the Victor-3 luminometer (see Chapter 2 – Materials and Methods)



		nanoluc	nanoluc	nanoluc	nanoluc	nanoluc	nanoluc	nanoluc	nanoluc
		count	mean	std	min	25%	50%	75%	max
Protein	Reporter								
GFP	C-2	28	27708.28571	23255.87968	4568	15724	21461	28180	95890
GFP	C-3	15	13716.13333	5349.121502	5754	8985	13818	16578	25486
GFP	empty	30	2822.2	1222.662369	852	2131.5	2703	3231.5	7142
GFP	i3	27	4189.037037	3181.714795	782	1452	3542	5425	12548
GFP	i5	24	1721.166667	911.6347317	512	738.5	1754	2323	3408
GFP	i7	30	3631.133333	1769.548915	1146	2188.5	3445	4809	8602
GFP	i7-2	21	7256.952381	3365.72094	1564	4526	7088	10402	13530
GFP	wt	24	1887.333333	1163.183139	478	705	1871	2640	4066
IRF3	C-2	24	6686215.667	2010767.399	2555500	4974042.5	6945031	8208645.5	10033316
IRF3	C-3	12	4893898	1442698.312	2386476	3850968.5	5403095	5866845.5	6745832
IRF3	empty	21	4473.142857	2427.861246	1590	3166	4060	5182	13770
IRF3	i3	21	1421754.857	839131.5405	281964	793520	1365136	1765130	3255692
IRF3	i5	18	3775.777778	2444.465527	1008	1606.5	3411	4925.5	9748
IRF3	i7	23	73786.17391	42063.0415	16040	43420	61200	102430	177110
IRF3	i7-2	18	5418433.444	1326723.979	3601574	4353113	5469694	6413292	7929744
IRF3	wt	21	10457.04762	6969.10194	874	4512	10106	15788	25604
IRF5	C-2	21	3370415.905	1229622.92	1721484	2332178	3375162	3939652	5931760
IRF5	C-3	15	2830400.8	1046073.005	1003706	1881219	2870008	3634919	4520668
IRF5	empty	18	4646.555556	2005.127757	2650	3088	4116	5369.5	10130
IRF5	i3	18	28469	20988.30106	10946	16227	21780	33507.5	99282
IRF5	i5	15	149502.5333	56713.02961	66310	115130	124946	181441	251660
IRF5	i7	21	1095387.333	536056.1616	391214	753732	869882	1410814	2419864
IRF5	i7-2	21	4150362.095	1622474.933	1929346	3118698	4052008	4527678	7731834
IRF5	wt	18	3529.222222	2075.673684	802	2184	3014	5118.5	7788
IRF7ss	C-2	30	1866511.733	1712772.654	287814	713546.5	1202476	2585748.5	7138636
IRF7ss	C-3	15	1120946.8	1125848.596	50812	353057	429608	1942815	3338332
IRF7ss	empty	29	3130.965517	1101.670111	1318	2168	3050	4064	5000
IRF7ss	i3	27	8902.740741	11032.02802	698	2496	3682	11912	44372
IRF7ss	i5	24	2548.916667	1877.589387	288	930	1970	3992	5850
IRF7ss	i7	30	1670716.2	1517504.273	79054	602573.5	1215045	2241068	6162902
IRF7ss	i7-2	21	3063626.19	1798373.428	535758	1875526	2518114	4114868	6866174
IRF7ss	wt	27	1145561.333	1457078.708	10904	118233	473918	1925630	5554786

Table 3.3 – Reporter assay transfection control normalized data

Table includes descriptive statistics of transfection control normalized (NanoLuc/Luc) values for each Protein by Reporter condition. Values are ratios of NanoLuc/Luc (see Chapter 2 – Materials and Methods)



		norm	norm	norm	norm	norm	norm	norm	norm
		count	mean	std	min	25%	50%	75%	max
Protein	Reporter								
GFP	C-2	28	8.847857035	3.219177098	4.262851897	6.583614497	8.566555258	9.884390095	15.70454545
GFP	C-3	15	5.431336883	0.98654193	3.584435798	4.930410673	5.694323144	6.200391582	6.787337662
GFP	empty	30	1.278804577	1.324831644	0.366118837	0.64123896	0.995949805	1.122473561	5.694779116
GFP	i3	27	1.0566312	0.253425062	0.566030606	0.920817594	1.047549378	1.172687049	1.622746186
GFP	i5	24	0.628816964	0.336402693	0.348809524	0.468663215	0.517197589	0.608135722	1.892287234
GFP	i7	30	1.081111399	0.347190717	0.575262055	0.781590122	1.060041618	1.284082354	1.710479221
GFP	i7-2	21	2.603662076	0.689136396	1.577534586	2.125347329	2.360062893	3.019448476	4.179859989
GFP	wt	24	0.568648003	0.17933598	0.328062182	0.450857761	0.504061421	0.663483999	1.027716674
IRF3	C-2	24	6669.709601	3648.334615	901.0930889	4497.622026	5710.181977	9030.016237	14781.52174
IRF3	C-3	12	4280.399519	1165.846931	2912.227371	3335.985644	3916.501538	5116.513041	6433.681768
IRF3	empty	21	4.714935996	2.994735037	1.5	2.5296	4.1882805	5.53888131	12.08955224
IRF3	i3	21	1286.463099	304.245254	875.8092199	1004.541667	1227.25	1506.273625	1906.883312
IRF3	i5	18	3.47682443	1.003295617	2.195758564	2.753195433	3.064827483	4.006265976	5.914951989
IRF3	i7	23	73.653328	38.82979273	28.28924162	42.90702347	62.17210682	88.94619815	168.5156993
IRF3	i7-2	18	5320.527038	1639.127305	2802.031646	3738.84485	5405.784991	6650.355465	8004.529492
IRF3	wt	21	9.745340625	5.333039116	3.552845528	5.413297394	8.463519313	14.88645262	21.28128342
IRF5	C-2	21	2823.644956	880.5467198	1619.922697	2046.722716	2789.390083	3778.486993	4435.2827
IRF5	C-3	15	2275.782718	652.4343189	1178.613333	1777.731808	2567.683801	2715.528331	3108.950472
IRF5	empty	18	4.337295352	2.057487506	1.991202346	2.930305083	3.825063372	4.812677103	8.555743243
IRF5	i3	18	23.93927107	6.846288819	11.2826496	21.14994536	25.56957381	27.86947462	35.53400143
IRF5	i5	15	125.5458627	31.61367346	66.76363636	114.811658	131.4990893	144.2896322	168.7131012
IRF5	i7	21	1010.89998	319.5865456	441.3404363	830.9	1002.297767	1209.054381	1489.132832
IRF5	i7-2	21	3316.522713	1200.08695	1789.362185	2032.865109	3905.557676	4341.614865	4931.025989
IRF5	wt	18	3.214332339	0.927389069	1.858778626	2.645689849	2.977250394	3.800734095	5.3371266
IRF7ss	C-2	30	729.5491368	484.085535	192.7733836	405.9285783	588.4876348	975.0366333	2295.042265
IRF7ss	C-3	15	515.5713424	465.7979523	59.14153439	241.8835053	409.5878788	457.5926459	1511.241286
IRF7ss	empty	29	1.851408288	1.238589017	0.402135231	1.103448276	1.440572792	2.08984726	5.506775068
IRF7ss	i3	27	2.44211573	1.503908414	1.042044518	1.321720271	1.817945384	2.82655869	5.615997975
IRF7ss	i5	24	1.089082084	0.809928957	0.338797814	0.58104423	0.694851868	1.516473023	3.674347158
IRF7ss	i7	30	646.5916696	472.9532218	60.89069374	326.6190421	473.3589806	941.6262052	1801.900943
IRF7ss	i7-2	21	1593.920676	1146.075789	422.7014716	526.2848723	1282.28732	2783.207945	4110.977987
IRF7ss	wt	27	251.3902109	228.5084693	18.70325901	94.01968037	198.6856072	363.1106355	797.2493529

Table 3.4 – Reporter assay GFP transfection normalized data

Table includes descriptive statistics of GFP control normalized (Reporter with IRF/Reporter with GFP) values for each Protein by Reporter condition. Values are ratios showing the fold activity of a reporter with IRF over its corresponding GFP control. (see Chapter 2 – Materials and Methods)

		gfp_norm	gfp_norm	gfp_norm	gfp_norm	gfp_norm	gfp_norm	gfp_norm	gfp_norm
		count	mean	std	min	25%	50%	75%	max
Protein	Reporter								
GFP	C-2	28	1	0.133844143	0.80669904	0.887153543	0.995228142	1.040677008	1.30479454
GFP	C-3	15	1	0.158421816	0.772507343	0.890100226	1.012772376	1.090856654	1.40528319
GFP	empty	30	1	0.28518211	0.53872407	0.855121801	0.978546493	1.078667971	1.864761327
GFP	i3	27	1	0.202062847	0.709090822	0.867097693	0.950547788	1.102240891	1.524724173
GFP	i5	24	1	0.231901606	0.493057559	0.892679154	0.987517385	1.06466807	1.581868319
GFP	i7	30	1	0.246478843	0.506419891	0.837444064	0.975643373	1.129358809	1.499302599
GFP	i7-2	21	1	0.185770047	0.686371339	0.902326579	0.958316165	1.155947892	1.385024226
GFP	wt	24	1	0.195636156	0.658131872	0.839095153	0.962117862	1.195119307	1.351314755
IRF3	C-2	24	805.9005182	338.0716455	186.9078327	584.3899646	867.5251569	1013.990102	1408.061472
IRF3	C-3	12	763.7682662	210.5630705	512.0964031	605.2677423	683.9754084	966.1672329	1091.002607
IRF3	empty	21	3.599570888	1.259156271	1.950445034	2.51201248	3.276583541	4.307475591	5.871610726
IRF3	i3	21	1229.405309	367.994617	845.9520425	970.2958766	1135.449965	1380.17153	2192.373069
IRF3	i5	18	6.295637145	1.60915675	4.140946116	5.24143953	5.656898034	7.325301104	10.00860014
IRF3	i7	23	71.40315421	36.21365608	24.90384086	49.36610598	55.08555289	95.33786051	156.718623
IRF3	i7-2	18	2035.5864	643.3341416	1137.779941	1514.682935	2013.656658	2465.941537	3173.63503
IRF3	wt	21	16.05925388	7.438961588	7.398837255	10.22471606	13.1341662	20.16813102	31.94349067
IRF5	C-2	21	320.4181412	51.75808341	229.4121249	277.9845097	325.9981326	359.931951	422.497142
IRF5	C-3	15	419.070192	119.4214688	254.0113719	319.0787718	419.2734	507.9706327	601.3873938
IRF5	empty	18	3.229574785	1.094405792	1.585455205	2.176822697	3.492685931	4.081646161	4.854413892
IRF5	i3	18	23.06507759	5.299854547	14.25464533	20.87159275	23.39697318	25.13352105	34.62661599
IRF5	i5	15	219.7997455	93.92154232	51.53346525	204.5813485	239.9703525	274.4797035	341.9352521
IRF5	i7	21	901.9453599	263.2940065	506.9869802	679.726701	915.756164	1084.405634	1310.926873
IRF5	i7-2	21	1287.859755	483.0767785	726.5800887	827.1839545	1190.887093	1699.992632	2255.918095
IRF5	wt	18	5.364321049	1.76121893	3.032375138	4.389290901	4.776284854	5.660936029	9.031518168
IRF7ss	C-2	30	90.24182339	57.41065909	24.0640019	44.02925934	74.75708472	121.0996669	228.9614173
IRF7ss	C-3	15	91.70673241	71.88875662	10.03399156	46.78932398	81.7131446	96.05723117	240.5170065
IRF7ss	empty	29	1.715038916	0.958684852	0.591720247	1.295781721	1.459815524	1.86528705	4.707274938
IRF7ss	i3	27	2.297324931	1.251133853	0.85470871	1.484843127	1.761098558	2.798645802	4.518048738
IRF7ss	i5	24	1.656442977	0.86318309	0.638933404	1.104665621	1.390584908	1.850670388	3.785873826
IRF7ss	i7	30	637.455357	480.4697319	48.77851978	300.7262728	484.4712773	879.3574741	1675.757426
IRF7ss	i7-2	21	574.6637128	317.2459928	171.6401941	213.7007882	590.8466619	835.6324714	1234.283159
IRF7ss	wt	27	455.0835679	357.1957461	31.18505095	195.7969487	281.3834612	764.1612064	1115.20361

Table 3.5 – Reporter assay data scaled to i7-2 reporter condition

Table includes descriptive statistics of i7-2 reporter scaled data used in reporter assay figures (gfp_norm/gfp_norm_i7-2) values for each Protein by Reporter condition. Values are ratios scaling the data to the i7-2 experimental condition to simplify figures given the range of reporter activities seen across IRF proteins. (see Chapter 2 – Materials and Methods)

		i72_scale	i72_scale	i72_scale	i72_scale	i72_scale	i72_scale	i72_scale	i72_scale
		count	mean	std	min	25%	50%	75%	max
Protein	Reporter								
GFP	C-2	28	1	0.133844143	0.80669904	0.887153543	0.995228142	1.040677008	1.30479454
GFP	C-3	15	1	0.158421816	0.772507343	0.890100226	1.012772376	1.090856654	1.40528319
GFP	empty	30	1	0.28518211	0.53872407	0.855121801	0.978546493	1.078667971	1.864761327
GFP	i3	27	1	0.202062847	0.709090822	0.867097693	0.950547788	1.102240891	1.524724173
GFP	i5	24	1	0.231901606	0.493057559	0.892679154	0.987517385	1.06466807	1.581868319
GFP	i7	30	1	0.246478843	0.506419891	0.837444064	0.975643373	1.129358809	1.499302599
GFP	i7-2	21	1	0.185770047	0.686371339	0.902326579	0.958316165	1.155947892	1.385024226
GFP	wt	24	1	0.195636156	0.658131872	0.839095153	0.962117862	1.195119307	1.351314755
IRF3	C-2	24	0.395905828	0.166080715	0.091820142	0.28708679	0.426179482	0.498131694	0.691722775
IRF3	C-3	12	0.375207982	0.103440989	0.251571932	0.297343184	0.336009028	0.474638283	0.535964775
IRF3	empty	21	0.001768321	0.000618572	0.000958174	0.001234049	0.001609651	0.002116086	0.002884481
IRF3	i3	21	0.603956338	0.180780642	0.415581497	0.476666516	0.557799937	0.678021591	1.077022852
IRF3	i5	18	0.003092788	0.000790513	0.002034277	0.002574904	0.002779002	0.00359862	0.004916814
IRF3	i7	23	0.035077437	0.017790282	0.012234234	0.02425154	0.02706127	0.046835576	0.076989423
IRF3	i7-2	18	1	0.316043643	0.558944558	0.74410152	0.989226819	1.211415805	1.559076554
IRF3	wt	21	0.007889252	0.003654456	0.003634745	0.005022983	0.006452276	0.009907774	0.015692525
IRF5	C-2	21	0.24879894	0.040189223	0.1781344	0.215849986	0.253131703	0.279480704	0.328061453
IRF5	C-3	15	0.325400488	0.092728629	0.197235274	0.247758943	0.325558275	0.394430085	0.466966525
IRF5	empty	18	0.002507707	0.000849786	0.001231078	0.001690264	0.002712008	0.003169325	0.003769365
IRF5	i3	18	0.017909619	0.004115242	0.011068476	0.016206417	0.01816733	0.019515728	0.026886946
IRF5	i5	15	0.17067056	0.072928393	0.040014811	0.158853748	0.186332675	0.213128567	0.26550659
IRF5	i7	21	0.700344394	0.204443074	0.393666297	0.527795592	0.711068236	0.842021525	1.017911204
IRF5	i7-2	21	1	0.375100454	0.564176407	0.642293504	0.924702467	1.320013787	1.751679938
IRF5	wt	18	0.004165299	0.001367555	0.002354585	0.003408206	0.0037087	0.004395615	0.007012812
IRF7ss	C-2	30	0.157034143	0.099903053	0.041874929	0.076617434	0.130088403	0.210731362	0.398426788
IRF7ss	C-3	15	0.159583301	0.125097087	0.017460632	0.081420356	0.142192978	0.167153814	0.418535225
IRF7ss	empty	29	0.002984422	0.001668254	0.001029681	0.002254852	0.002540295	0.003245876	0.008191356
IRF7ss	i3	27	0.003997686	0.002177158	0.00148732	0.002583847	0.003064572	0.004870058	0.007862074
IRF7ss	i5	24	0.002882456	0.001502066	0.001111839	0.001922282	0.002419824	0.003220441	0.006587981
IRF7ss	i7	30	1.109266764	0.836088518	0.084881851	0.523308269	0.8430518	1.530212287	2.916066194
IRF7ss	i7-2	21	1	0.552055029	0.298679367	0.371871032	1.028160729	1.454124304	2.147835563
IRF7ss	wt	27	0.791912831	0.621573519	0.054266609	0.340715699	0.489648911	1.329753714	1.940619505

4 CHAPTER 4 - MASSIVELY PARALLEL REPORTER ASSAYS

4.1 MPRA Overview

Massively parallel reporter assays (MPRAs) use high-throughput sequencing to simultaneously measure the transcriptional activity of many thousands of cis-regulatory elements (CREs). MPRAs achieve this scale by quantitating mRNA abundance transcribed from a complex pool of reporter plasmids after transfection into cells. Every CRE construct in the MPRA plasmid pool is identified by a sequence barcode embedded in the plasmid's 3'-untranslated region (3'-UTR). This transcribed barcode is a key feature of MPRAs and results in reporter mRNAs that identify which CRE induced their expression. After mRNA extraction, CRE-associated barcodes are quantified using RNA-seq and normalized to the CRE plasmid abundances. By using high-throughput sequencing, systematic comparisons of many CREs can be performed at a scale that would be prohibitive using low-throughput luciferase assays.

4.2 Proposed IRF-MPRA experimental design

The relationship between transcription factor DNA-binding affinity and transcriptional activity is complex and requires protein-specific empirical testing (Mulero et al., 2017). The Interferon Regulatory Factor (IRF) MPRA library described herein was designed to test the relationship between *in vitro* DNA binding affinity and IRF-dependent transcriptional regulation in cells. The IRF-MPRA design leverages IRF DNA-binding affinity data from PBM experiments (Andrilenas et al., 2018) to dissect IRF-dependent transcriptional regulation in a systematic manner. This IRF-MPRA experiment is intended to be tested in immune-activated cells (e.g., THP-1s with immune ligand stimulation) or cell lines expressing activatable or constitutively active IRFs of interest.

The highly parallel nature of MPRAs allows multiple experimental hypotheses to be tested at once. Included in this MPRA library are multiple binding site experiments designed to test different aspects of IRF-dependent gene regulation such as: the relationship between DNA binding affinity and transcriptional activity and the impact of binding site spacing on reporter expression. The experiments included in the IRF-MPRA design are described below.

4.2.1 MPRA design to assay the impact of Single Nucleotide Variants (SNVs) on IRF-dependent gene expression

To interrogate the impact of single nucleotide variation on IRF-dependent gene regulation we designed sets of regulatory elements that include every possible single nucleotide variant (SNV) of four different IRF binding sites (Figure 4.1). Nearly all IRF binding sites in this design are included in PBM experiments discussed in Chapter 3. This allows for direct comparisons between MPRA transcriptional output and the PBM measured DNA-binding affinities of IRF3, IRF5 and IRF7 to the SNV-modified IRF binding sites. To date, data directly comparing PBM derived protein:DNA-binding affinities and transcriptional regulation at the scale of MPRAs has not been generated, with most studies using binding models derived from HT-SELEX experiments or other methods (Hughes et al., 2018). We anticipate that the data generated from the SNV subset of the IRF-MPRA may elucidate details of the complex relationship between DNA-binding affinity and transcriptional output, and help to identify how DNA sequence can affect transcriptional activity and binding affinity differently.

The IRF-MPRA SNV design has multiple potential benefits. (1) SNVs are associated with changes in gene expression and health outcomes between people (Ghodke-Puranik and Niewold, 2015; Matta and Barnes, 2019); however, the mechanisms underlying the impact of SNVs seen in GWAS studies is not well understood. Systematically assaying the impact of SNVs on IRF-dependent gene expression may provide important insight into what SNVs may impact IRF gene regulation. (2) Using the SNV modeling approach used with IRF PBM data (chapter 3) we can generate a PWM-motif illustrating IRF-dependent transcriptional output rather than DNA binding affinity (Andrilenas et al., 2018; Mohaghegh et al., 2019). By modeling the impact each SNV has on transcriptional output, we can provide an IRF DNA sequence logo that represents transcriptional regulation, which may differ from IRF DNA-binding affinity. (3) By measuring the impact of individual SNVs across IRF binding sites we can examine the regulatory consequences of subtle differences in DNA binding affinity found in PBM experiments (chapter 3). For example, IRF3/5/7 were shown to have decreased DNA binding specificity in the 5' portion of their binding motifs (chapter 3). The MPRA data will allow the regulatory impact of this binding site asymmetry to be assayed, expanding our knowledge of IRF-dependent gene regulation.

The SNV CREs included in this MPRA design are all in a conserved Viral Response Element (VRE)-like arrangement found in the promoters of the type-I-interferon promoters that features two IRF binding sites with a small spacer (Figure 4.2A,B) (Civas et al., 2006, 2002). Single binding site CREs are also included for a limited set of IRF binding sites in the form of binding site ablations discussed further below (Figure 4.3).

As part of the SNV experimental subset, multiple variants of the high-affinity common IRF binding site were included to explore the effects of different flanking sequences. Traditional reporter assays as, discussed in Chapter 3, suggest that flanking regions adjacent to the main binding site may influence gene expression (Andrilenas et al., 2018). We include the two flanking sequences interrogated in those assays (TAA and GTC) in our MPRA design. A non-consensus GATA half-site core sequence is included to examine the effects of lower affinity half-sites on IRF-dependent gene expression. Figure 4.1B lists all starting sequences (seeds) used to generate SNV probes included in the IRF-MPRA design.

4.2.1.1 Design to assay previously identified IRF homolog specific binding sites

Protein binding microarray data (PBM) and luciferase reporter assay data (chapter 3) identified a small set of IRF binding sites with increased IRF-homolog specificity(Andrilenas et al., 2018). These sequences have been included in the IRF-MPRA design to corroborate previous low-throughput studies and to provide known points of comparison between the existing data and this high-throughput design. The DNA-binding affinities for most of these sites have been measured using PBM (Figure 4.2C).

4.2.1.2 Design to assay VRE ablation sequences

As discussed in Chapter 1, many promoters of Interferon Regulatory Factor (IRF) driven genes feature a conserved “Viral Response Element” sequence structure featuring multiple IRF dimer binding sites (Barnes et al., 2002b; Civas et al., 2002). Numerous studies, including our own (Chapter 3) have explored the impact of altering IRF binding sites on transcriptional regulation; many have done so in a native or multi binding site context (Civas et al., 2006; Génin et al., 2009). The VRE ablation design included in this MPRA design feature position matched ablations of IRF binding sites (Figure 4.3). These ablations were included to interrogate whether a single IRF binding site is sufficient to drive gene expression. This sequence design allows for the direct comparison of intact, sequence matched VRE regulatory elements to ablation VREs with either the Transcription Start Site (TSS)-distal or TSS-proximal binding site substituted with a "null" sequence (Figure 4.2C).

4.2.1.3 Design to assay IRF VRE binding site spacing on IRF-dependent gene expression

The VRE cis-regulatory element is an important feature of IRF-dependent gene regulation that has been the focus of many studies (Civas et al., 2006; Génin et al., 2009). What remains unclear is the relationship between multiple IRF binding sites and whether IRF-driven gene expression requires a specific spacing between binding sites for efficient transcriptional activation. To explore this question, CRE plasmids that vary the length of the spacer between the IRF binding sites are included in this design.

4.2.1.4 Design to assay the impact of binding site orientation on IRF-dependent transcriptional activity

Intimate DNA-binding protein complexes like the Interferon Beta enhanceosome, suggest that transcription factor orientation is an important component of transcriptional complex assembly (Panne et al., 2007). At the same time, other studies find that TF binding site orientation has negligible impact on gene expression (Lis and Walther, 2016). The impact of IRF DNA-binding site orientation on transcriptional output has yet to be systematically assayed. To address this gap in the literature, we designed a subset of the IRF-MPRA design to include CREs from other experimental subsets with the IRF binding sites reverse-complemented to change the orientation of IRF dimers that may bind. This design will allow the direct comparison between CREs with IRF binding sites in both orientations.

4.3 MPRA technical design

In brief, MPRAs require the construction of a complex plasmid library with thousands of unique members followed by transfection into cells and subsequent high-throughput DNA and RNA sequencing (Figure 4.4). This section provides a technical overview of the CRE-seq from plasmid library construction to sequencing and quantification.

In MPRAs, the complex plasmid library containing the CREs being assayed, starts as an oligo pool often synthesized using microarray synthesis which allows for the customization of each member oligo (King et al., 2018; LeProust et al., 2010; White, 2015; White et al., 2016). We used Agilent OLS oligo pools (Agilent Inc.) which are delivered as 10 picomoles of single stranded DNA, which much be amplified via PCR and inserted into a modified mammalian expression vector backbone. An open reading frame containing a minimal promoter and the coding sequence for a characterized exogenous gene are inserted between the CRE and barcode (Figure 4.4). After integration, the barcode is contained within the vector sequence corresponding to the 3’ untranslated region of the mRNA that will be transcribed upon expression of the reporter gene. The MPRA plasmid library is transfected into cells and the resulting mRNA is extracted and sequenced. A critical aspect of the MPRA technique is the inclusion of the CRE-associated barcode within the mRNA that is produced from the expression plasmid library (King et al., 2018). The gene regulatory activity of a given CRE is quantitated by counting the number of reads that contain a specific barcode. These values are normalized by their representation in the MPRA plasmid library as determined by the barcode counts from the transfected MPRA plasmid pool.

Our MPRA protocol is based on a detailed Cis-Regulatory Element Sequencing (CRE-seq) protocol shared by the Cohen Lab at Washington University, St. Louis, Missouri (King et al., 2018; Kwasnieski et al., 2012; White et al., 2016). Modifications have been made to improve key steps in the MPRA library cloning process as well as the inclusion of Unique Molecular Identifiers that have been used to improve RNA-seq quantitation in single-cell sequencing experiments (Zhang et al., 2019; Ziegenhain et al., 2017). The following sections delve into the technical decisions made while designing the IRF-MPRA experiment.

4.3.1 Oligo library, cloning and plasmid design

4.3.1.1 Complex oligo pools and library design

As mentioned above, MPRA experiments begin as custom single stranded DNA (ssDNA) oligo pools that require PCR amplification before cloning into the target vector backbone. We sourced our starting MPRA oligo pool from Agilent, which is able to manufacture oligo pools with each oligo at a custom and independent length up to 210bp. The IRF-MPRA design contains 3,131 unique oligos with a maximum length of 153bp including cloning sequences and barcodes (Figure 4.5). We ordered a 15,000 oligo pool containing multiple independent MPRA subsets from different projects and used a subset specific primer strategy to facilitate multiplexing. To mitigate cloning contamination by multiplexed MPRA designs, subset-specific amplification primers were used. These primers were designed with specific attention toward potential PCR mispriming. Although unamplified ssDNA oligos may be purified along with dsDNA during PCR clean-up, contaminating ssDNA will be unable to participate in subsequent restriction digest and ligation steps (Horspool et al., 2010; Nishigaki et al., 1985). Thus, It is highly unlikely that contamination from other MPRA experimental subsets will occur.

4.3.1.2 Barcodes

CRE-associated barcodes are an essential component of the MPRA methodology. Our MPRA library design includes barcodes as part of each oligo, resulting in a defined mapping between every cis-regulatory element (CRE) and its associated barcodes. Each CRE has multiple barcodes that are used as replicates and averaged to mitigate potential effects of barcode sequences on transcription and mRNA stability. Variations in the 3' untranslated region (UTR) of mRNA transcripts can impact mRNA stability which could alter the barcode counts of MPRA experiments (Pesole et al., 2001; Rabani et al., 2017). Having multiple barcodes for each CRE has been shown to increase the correlation between MPRA replicates (Tewhey et al., 2016); however strong guidelines for the number of barcodes to include do not exist.

Analysis done in Tewhey et al. (2016) suggests that barcode counts up to 50 barcodes per CRE are beneficial for increasing MPRA replicate correlations; however, their MPRA preparation process adds barcodes in a separate emulsion PCR step, which requires paired-end HT-sequencing to map CREs to barcodes. While this PCR barcoding process increases the number of barcodes that can be incorporated into an MPRA library, it may also increase variability at multiple protocol steps. Stochastic PCR bias, derived from random templates outperforming others in a sequence-independent manner, can result in the loss of barcodes or the increase in variability in separately handled samples (Best et al., 2015; Smith et al., 2014). This potential increase in variability may account for the high suggested barcode count found in Tewhey et al. (2016) in comparison to CRE-seq experiments (Fiore and Cohen, 2016; King et al., 2018).

CRE-seq experiments, where barcodes are included in the initial oligo-pool design, have used barcode quantities between 3 - 10 barcodes per CRE (Fiore and Cohen, 2016; King et al., 2018). Ten barcodes per CRE was suggested to be sufficient for MPRA libraries used in cells with a transfection efficiency of 10% - 40%, ensuring that at least 4 barcodes were sufficiently represented from the starting 10 (personal communication with Cohen Lab). By Including more than the minimum number of barcodes we may be able to use the IRF-MPRA design in primary immune cells or challenging cell-lines, like THP-1s, assuming a moderate transfection efficiency can be reached (Auwerx, 1991; Schnoor et al., 2009).

Barcodes were designed to avoid sequences that match IRF binding sites and restriction enzyme recognition sites used in the MPRA cloning procedures (Acc65I, XbaI, BbsI, BsaI, NheI, and more). See chapter 2 for more information.

4.3.1.3 Vector design and MPRA CRE-library cloning strategy

Briefly, successful MPRA library generation involves two rounds of restriction enzyme digestion and ligation; the first inserts the CRE-library, the second inserts the open reading frame (ORF) containing a minimal promoter and eGFP gene. Each round of MPRA library cloning requires large scale colony picking (>15,000 colonies; see Chapter 2 - Methods) and plasmid DNA purification to ensure full representation of all CRE plasmids.

Vector design is an important component of experiments that require large scale cloning. One limitation of existing CRE-seq implementations are the cloning efficiencies and the chance of “background” contaminating vector which does not contain an insert. To address these concerns, we generated a modified MPRA reporter vector (pNL-MPRA) that is designed to reduce background vector contamination. First, we opted to include a ccdB death gene cassette (Bernard et al., 1994; Dao-Thi et al., 2005; Hu et al., 2010), which is removed during CRE insertion, to remove background during the first cloning step (Figure 4.4). Second, we adapted the starting reporter vector (pNL3.1; see Chapter 2 for more details) to be compatible with asymmetrical type-IIS restriction enzymes for CRE insertion and ORF insertion. Type-IIS restriction sites were selected for cloning steps because they allow for efficient, simultaneous restriction digest and ligation reactions that proceed to an indigestible end product that lacks the initial restriction enzyme site (Pingoud et al., 2014, 2005; Roberts, 2003). This processive restriction/ligation reaction design has been used widely in synthetic biology and complex DNA assembly (Andreou and Nakayama, 2018; Engler and Marillonnet, 2013; Iverson et al., 2016; Werner et al., 2014).

4.3.2 Unique Molecular Identifiers and MPRA sequencing library preparation

4.3.2.1 <u>Rationale for Unique Molecular Identifiers (UMIs)</u>

Accurate quantitation of high-throughput sequencing (HTS) is essential for MPRAs. MPRA experiments use barcode counts from raw high-throughput sequencing (HTS) reads to measure the transcriptional activity of different CRE constructs. Ideally, sequenced barcode counts represent the barcode abundances from a starting experimental sample; however, sources of bias and variation during

HTS library preparation may lead to inaccurate quantitation. These technical errors, when coupled with PCR based amplification, distort the relationship between the starting composition of an experimental mRNA/DNA sample and the resulting HTS data. Methods to measure and mitigate bias in quantitative HTS experiments exist, but have not been implemented for MPRAs (Alon et al., 2011; Marx, 2017; Zheng et al., 2011). Here, we briefly discuss sources of HTS bias and propose a technical solution adapted to MPRA protocols.

During HTS library preparation two sources of bias influence the composition of sequencing reads that may negatively impact MPRA barcode quantitation. HTS library preparation is known to bias sequencing data through sequence dependent and stochastic processes (Best et al., 2015; van Dijk et al., 2014). Sequence dependent bias may result from differences in PCR amplification stemming from properties such as GC content or even context specific bias in polymerase efficiency (Thielecke et al., 2017). In clonal cell population experiments, with similar conditions to MPRAs, differences between barcode sequences have been found to bias barcode counts (Thielecke et al., 2017). This illustrates the influence that small sequence differences can have on HTS read composition, and emphasizes the importance of having multiple barcodes per MPRA CRE as within-sample replicates. Although having multiple barcodes per MPRA CRE may improve data reliability through averaging, it does not fundamentally address problems with HTS read quantitation nor stochastic sources of bias.

Multiple studies suggest that stochastic processes during PCR, where PCR composition becomes biased due to random, sequence-independent differences in template amplification, may be a primary driver of skewed HTS library composition (Best et al., 2015; Kebschull and Zador, 2015). Best (2015) found that heterogeneous amplification efficiency during HTS protocols stemmed from stochastic variation in early cycles of PCR. In their studies of T-cell receptor repertoire, small differences in template abundance and amplification efficiency in the initial cycles of PCR led to appreciable heterogeneity in HTS read composition across replicates and samples (Best et al., 2015). As mentioned previously, the DNA duplication inherent in HTS library preparation coupled with sequence dependent and stochastic sources of bias may distort the relationship between the initial mRNA/DNA template composition and resulting HTS barcode counts. This is especially problematic for MPRAs, because they rely on two HTS measurements (mRNA and plasmid DNA) to normalize CRE barcode counts (Fiore and Cohen, 2016; King et al., 2018; White, 2015). Tracking the initial sample composition through PCR amplification may improve MPRA quantitation and has been implemented for single-cell RNA sequencing using unique molecular identifiers (Hashimshony et al., 2016; Islam et al., 2014).

Multiple single-cell RNA-sequencing (scRNA-seq) protocols use Unique Molecular Identifiers (UMIs) to address sources of PCR bias by tagging every mRNA molecule at the beginning of HTS library preparation (Bageritz and Raddi, 2019; Hashimshony et al., 2016; Islam et al., 2014; Jaitin et al., 2014; Ziegenhain et al., 2017). UMIs improve HTS quantitation and measure amplification bias by tracking PCR duplication events: individual copies of the same mRNA/DNA will receive different UMIs such that copies of identical starting material can be distinguished from PCR duplicates which will share the same UMI. In RNA-seq, UMIs are often random sequences, 5 to 20 nucleotides in length, that are added to each mRNA molecule during cDNA first-strand synthesis while some methods add UMIs to sequencing adapters (Hong and Gresham, 2017; Islam et al., 2014; MacConaill et al., 2018; Pflug and von Haeseler, 2018). In the technical solution proposed below, we add UMIs during both cDNA synthesis and adapter ligation to track bias independently at each PCR step.

UMIs are acknowledged as a way to improve the accuracy of MPRA barcode counts and detect PCR amplification bias (Best et al., 2015; Kinney and McCandlish, 2019; Ogawa et al., 2017; Smith et al., 2017); however, to date, no published MPRA experiment has included UMIs. Implementing UMIs for MPRA experiments requires modifications to existing MPRA and HTS library preparation protocols. These modifications must work for both mRNA and plasmid DNA derived HTS libraries, which specifically excludes many UMI tagging methods used in scRNA-seq targeted for mRNA labeling (Islam et al., 2014; Ziegenhain et al., 2017). The impact of PCR bias on MPRA HTS library preparation has not been measured and the inclusion of UMIs in MPRA experiments may improve barcode quantitation. Below we propose protocol modifications that will allow UMI tagging of samples from both MPRA mRNA and plasmid pools.

4.3.2.2 Standard HTS library preparation and CRE-seq protocols are incompatible with UMIs

Standard protocols for HTS library preparation are inefficient for MPRA experiments and incompatible with UMI incorporation. Standard Illumina sequencing protocols are designed to provide uniform coverage of input mRNA and DNA samples (NEB and Illumina promotional materials). Standard HTS library preparation often relies on random primer cDNA synthesis and library fragmentation before adapter ligation. While these steps provide uniform sequencing coverage, they are incompatible with UMI incorporation due to the lack of UMI incorporation into the initial cDNA synthesis process. Additionally, sequencing MPRA experiments requires a targeted sequencing strategy, where regions of interest are selectivity amplified for sequencing, to obtain the greatest data yields. For example, MPRA plasmid-derived mRNAs all contain the same reporter gene sequence (eGFP) and the barcode region of the mRNA is the only experimentally informative portion (Figure 4.4, 4.5). It is therefore inefficient to use HTS protocols that will uniformly sequence the eGFP gene as part of the sequencing library.

While the standard CRE-seq protocol from the Cohen lab addresses issues with sequencing efficiency it is not compatible with UMIs. In the CRE-seq protocol, cDNA reverse transcription is initiated using poly-dT primers that select for poly-A tail containing RNA. The resulting single stranded cDNA library is then amplified using PCR targeted to the MPRA barcode region. The PCR primers contain restriction sites used to add custom illumina sequencing adapters. After adapter ligation the HTS library is amplified and enriched for amplicons containing the Illumina P5 and P7 adapter sequences. A parallel process is used for plasmid DNA (pDNA) HTS library preparation which starts at the barcode region PCR amplification step. This process does not allow incorporation of UMIs. Our proposed solution is described below.

4.3.2.3 Proposal for UMI incorporating sequence-specific first-strand DNA synthesis from mRNA and plasmid DNA (pDNA)

Incorporating UMIs into existing CRE-seq methods requires protocol modifications. CRE-seq studies to date have generated sequencing libraries using poly-dT primer-based cDNA synthesis followed by sequence-specific amplification of the MPRA barcode region. To effectively identify the starting composition of an MPRA sample, UMIs must be incorporated into every cDNA during first-strand synthesis. To accomplish this, we propose using sequence-specific primers (ss-primer) targeting the MPRA barcode region to initiate cDNA synthesis (Figure 4.6). These ss-primers will contain a trailing 5-bp UMI and restriction site-containing segment used for PCR enrichment and adapter ligation. A second primer targeted to a portion of the eGFP coding sequence will be used to incorporate the other required restriction site used for adapter ligation. Enrichment of the restriction site-containing amplicon can proceed using primers targeting the restriction enzyme sites (see figure 4.6). We can use a similar method to add UMIs to MPRA plasmid sequencing.

To incorporate UMIs into pDNA sequencing, we propose a ss-primer strategy that begins with a single PCR cycle used to incorporate the UMI sequence and restriction site for adapter ligation (see figure 4.6). This process parallels the cDNA library preparation process described above with one important modification. To prevent circular amplification of the small pNL-MPRA plasmid we propose an initial restriction digest of the pDNA sample targeting a site within the CRE/ORF region of the MPRA library. In the case of the IRF-MPRA design, a HindIII site is present just before the minimal promoter. Just as in the cDNA protocol, the second adapter restriction site can be added using a single PCR cycle, then the adapter-ready amplicon can be enriched using PCR.

4.3.2.4 Additional Sequencing library preparation considerations - Preventing UMI oligo carryover during library preparation.

In experiments that include UMIs, careful control of unwanted library amplification is important. In standard cDNA preparation and downstream sequencing library construction, small quantities of primer may contaminate subsequent protocol steps. For example, the Invitrogen SuperScript IV protocol recommends using an unpurified 1st strand synthesis reaction as template material for subsequent amplification (See Invitrogen SuperScript IV manual). The unpurified reaction contains the synthesized cDNA first-strand as well as any excess primer that was not incorporated into the cDNA. In the case of poly-dT cDNA protocols, any excess primer will not participate in the sequence-specific amplification of template, however it is possible that small quantities of additional full-length (poly-A containing) cDNA molecules could be generated during PCR amplification. The quantity of this off-target amplification is likely to be small; however, the CRE-seq protocol uses many cycles of PCR amplification prior to adapter ligation (21+13 cycles). As discussed earlier, careful quantitation of MPRA sequencing data is essential for accurate interpretation of MPRA experiments. Accidental UMI incorporation after the initial 1st strand synthesis complete defeats the intention of UMIs as a method of tracking PCR duplication to better estimate barcode counts.

There are few methods to selectively remove primers without purification. Single-cell sequencing protocols have used template switching RNA primers to generate cDNA without subsequent primer contamination during amplification by elegantly relying on the hydrolysis of RNA at high temperature in the presence of magnesium ions contained in most polymerase buffers (Islam et al., 2014). Unfortunately, this template switching technique only tags cDNA molecules at the 5' end when using a poly-dT oligo (Islam et al., (2014) (Figure 4.7). A requirement for using UMIs in MPRA experiments is that the UMIs must be incorporated adjacent to the CRE-barcodes so they are included in the sequencing read. This requires that sequence-specific primers be used to synthesize the first strand of cDNA which means that a template switching RNA oligo strategy cannot be used. Additionally, the UMI incorporation techniques used in scRNA-seq protocols are incompatible with sequencing MPRA pDNA pools for barcode normalization given the lack of reverse transcriptase with template switching properties in standard PCR.

4.3.2.5 Nonsense-mediated primer exclusion (NOPE) - a solution for UMI primer contamination

Nonsense-mediated primer exclusion (NOPE) is a method recently developed which addresses the technical complications of UMI primer contamination during sequence-specific cDNA generation (Shagin et al., (2017). The NOPE method sequesters first strand synthesis primers preventing additional incorporation of UMI oligos during subsequent second-strand synthesis and PCR amplification. Three key features allow the NOPE oligo to sequester UMI containing primers: (1) a region complementary to the target sequence in the UMI oligo; (2) a modified base that prevents 3’ polymerase extension from the oligo; (3) a non-complementary “non-sense” region that is incorporated to the 3’ end of the UMI-oligo during PCR (See figure 4.8). In the work by Shagin et al. (2017) the NOPE technique was tested with genomic DNA. In the IRF-MPRA experiment, the NOPE oligo will be necessary in both mRNA and plasmid DNA sequencing preparation due to the carryover of UMI containing oligos in both processes.

In Shagin et al. (2017), polymerase extension from the 3' end of the NOPE oligo was inhibited by the addition of a bulky Black-Hole quencher (Millipore Sigma) which likely prevented extension through steric hindrance. For the IRF-MPRA experiment we propose using an inverted-dT (IDT-DNA) that prevents elongation and also has the added benefit of being resistant to 3' exonuclease activity exhibited by high-fidelity proofreading DNA polymerases (Ortigão et al., 1992). If oligos are not exonuclease resistant then the NOPE oligos may lose their extension inhibiting modifications across PCR cycles which may allow them to participate in amplification (Ott and Eckstein, 1987). The other modification that is important is the non-complementary “non-sense” tail that is added to each UMI-oligo. This sequence can vary, but a set of nucleotides with a high AT content may reduce the possibility of the non-sense tail mispriming, decreasing the likelihood of unwanted polymerization. The NOPE oligo would be added after the initial first strand cDNA synthesis step for mRNA samples, or after a single cycle of PCR in the case of pDNA, along with the primer for second strand synthesis (see Figure 4.6). Additional amplification, and selection for full-length target amplicons (those with both the R1 and R2 restriction sites), can be performed with additional primers corresponding to R1 and R2 sites. Throughout this process, the PCR products that are generated from the NOPE-oligo:UMI-oligo duplex will lack the primer site used to generate the second strand of cDNA and will fail to be amplified. A similar protocol was used in the original NOPE oligo paper (see figure 4.8).

4.3.2.6 Sequencing library preparation - adapter design and library complexity

In standard Illumina RNA-seq library preparation, P5 and P7 illumina adapters are added to fragmented DNA/cDNA using T/A cloning ligation (Figure 4.9). As part of the CRE-seq protocol, restriction enzyme (RE) sites are included in the 1st and 2nd strand cDNA synthesis primers (similar to Figure 4.6). The Cohen lab CRE-seq protocol uses a linear adapter that uses specific restriction digest sites to provide end-specificity. In our modified design we use type-IIS BbsI sites to generate sticky overhangs for ligation. By using end-specific overhangs we can ensure that the 'Read1' sequencing primer is ligated to the barcode side of the target amplicon which is not the case in standard T/A adapter ligation (Figure 4.9). This process guarantees that all sequencing reads will be in the same orientation, allowing for efficient use of single-end sequencing reads and maximizing data yield.

As part of the adapter ligation process, we propose including an adapter-UMI adjacent to the Illumina 'Read1' primer sequence (Figure 4.10). This adapter-UMI has two purposes: **(1)** the UMI allows additional tracking of PCR duplication events after adapter ligation. Similar strategies have been used in other studies (Hong and Gresham, 2017; MacConaill et al., 2018; Pflug and von Haeseler, 2018). **(2)** This UMI will improve sequencing by increasing library complexity. MPRAs involve sequencing highly similar pools of RNA and DNA. Illumina sequencing machines use the first ~5 bases of sequencing to visually distinguish sequence clusters (Figure 4.11) (2011). This process requires sufficient representation of all nucleotides at each sequenced base during cluster calling. Krueger (2011) found that considerable sequencing data were automatically discarded by the Illumina sequencing platform due to low sequence complexity. One recommended way to increase sequence diversity is to spike-in Phi-X lambda phage genome standards (Illumina). Phi-X concentrations of 20-30% of total DNA are recommended by the Cohen Lab CRE-seq protocol, which is a sizable portion of an expensive sequencing lane. By including a random UMI in our adapter design, the resulting increase in sequence diversity may allow significantly less Phi-X spike-in to be used; however, empirical tests will be required to optimize conditions.

One manufacturing complication arises from the inclusion of a random UMI in our adapter design, but is easily solved. Correctly annealing random oligos is recognized as a physically improbable process (Oliphant et al., 1986). In the case of sequencing UMIs it is incredibly important that both strands of the adapter have identical sequences so that UMI diversity is not over-estimated. Preliminary designs for the UMI sequencing adapter used a single stranded DNA oligo to avoid the random annealing problem. This strategy assumed that dsDNA-ssDNA ligation by T4 DNA ligase would be equivalent to a dsDNA-nick repair and highly efficient. Research by Horspool et al. (Horspool et al., 2010) found that T4 DNA ligase required a double-stranded duplex at the site of ligation of at least 5 bp for effective ligation with the greatest efficiency at a length of 6bp or greater (Horspool et al., 2010). This context directly parallels our MPRA sequencing adapter ligation and in light of these data, we propose a modified semi-dsDNA adapter that includes a 6 bp constant sequence adjacent to the ligation site followed by the ssDNA UMI and remaining adapter (Figure 4.10). This strategy will likely solve both the UMI annealing and the T4 ligase efficiency problems; however, empirical tests of this modification will be required. This design reduces the cost of sequencing adapters, which are recommended to include an exonuclease resistant 3' modification and be purified to a higher standard than desalted oligos (NEB and IDT-DNA recommendations). The constant region in this adapter design could also be used for additional sample multiplexing if multiple sets of adapters were ordered. Also, greater sequencing read base diversity could be achieved by varying the length of the constant duplex-region to change the sequencing phase of different samples, thus increasing the complexity of an individual base position.

Lastly, our custom UMI adapter design uses a two-step adapter extension process paralleling that used by the NEBnext library workflow (Figure 4.12). In the NEBnext protocol, truncated adapters are ligated on to the prepared DNA library, then a final PCR step with primers (containing the complete Illumina adapters) is used to amplify the library. This two step adapter preparation step allows for the inclusion of multiplexing indices in the P7-index region of the Illumina adapter (Figure 4.12 and 4.10). In the Cohen lab CRE-seq protocol, they use inline sample multiplexing indices that occupy sequencing read space. This two-step adapter preparation allows for standard NEBnext index sets to be used for sample multiplexing easily expanding the number of replicates and samples that can simultaneously be sequenced.

4.3.2.7 Preliminary IRF-MPRA library sequencing

In an effort to verify the integrity of the IRF-MPRA library and assess barcode representation after CRE insertion into the pNL-MPRA backbone, the IRF-MPRA library was sequenced using the MGH sequencing core amplicon sequencing service. 150,226 paired-end reads, with 75,113 pairs, were obtained from the IRF-MPRA CRE-insert library (See summary Table 4.1). 71,476 barcodes were successfully extracted from the raw sequencing reads with 94.9% (67,866) of these barcodes identically matched designed barcodes. 3,610 extracted barcodes did not match the designed barcode list, and likely represent sequencing errors.

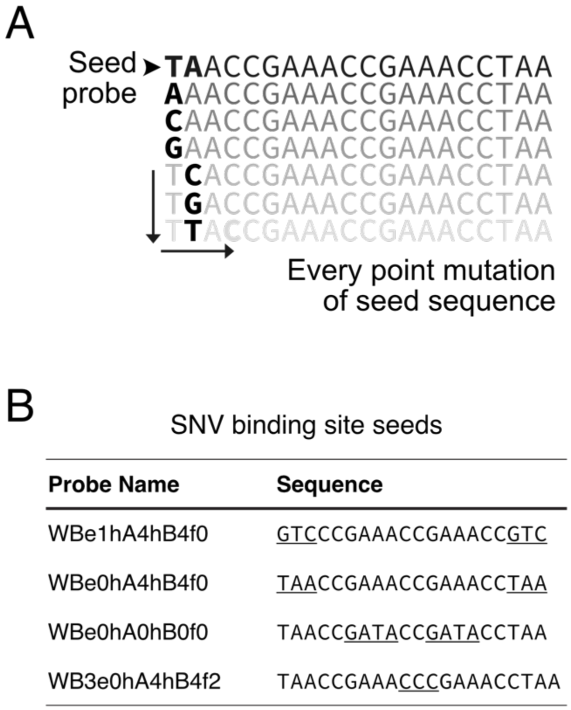
3,159 of 3,160 barcodes were represented in the sequencing data and replicate sequencing may verify the presence of all barcodes in the MPRA library. The mean number of reads per barcode was 21.5 reads/barcode with a minimum of 0 and a maximum of 60 reads/barcode. The distribution of read/barcode frequencies was unimodal and comparison of the mean (21.5) and median (21) does not suggest a skewed distribution (see Table 4.2 and Figure 4.13). These data suggest that there is sufficient barcode representation in the IRF-MPRA CRE insertion library to proceed with ORF cloning.

4.4 Conclusion

To assay the sequence determinants of IRF-dependent transcriptional regulation, we proposed using a modified massively parallel reporter assay (MPRA). MPRAs use high-throughput sequencing to measure the transcriptional activity of many thousands of cis-regulatory elements (CREs) at once. In this chapter we detail modifications to the MPRA technique that may improve key steps in the MPRA library cloning process and improve MPRA high-throughput sequencing. The proposed MPRA modifications leverage unique molecular identifiers to improve accuracy of reporter gene quantitation. The experiments and techniques proposed lay groundwork for future studies of IRF-dependent transcriptional regulation.

Figure 4.1 - The IRF-MPRA design includes Single Nucleotide Variants (SNV) for a subset of IRF binding sites.

A) Diagram of SNV probe generation process used for the IRF-MPRA design. SNV probes begin with a seed probe which is used to generate all possible single base substitutions along that seed sequence. B) List of all PBM derived binding sites used for SNV experiments in the IRF-MPRA design. Underlined regions highlight key differences between the binding sites. The underlined regions in the first two probes highlight the differences in the sequences flanking the IRF binding site. The core half sites of the 3rd sequence feature non-canonical 5'-GATA-3' sites. The 4th sequence features a 3 base pair spacer between the core half sites.

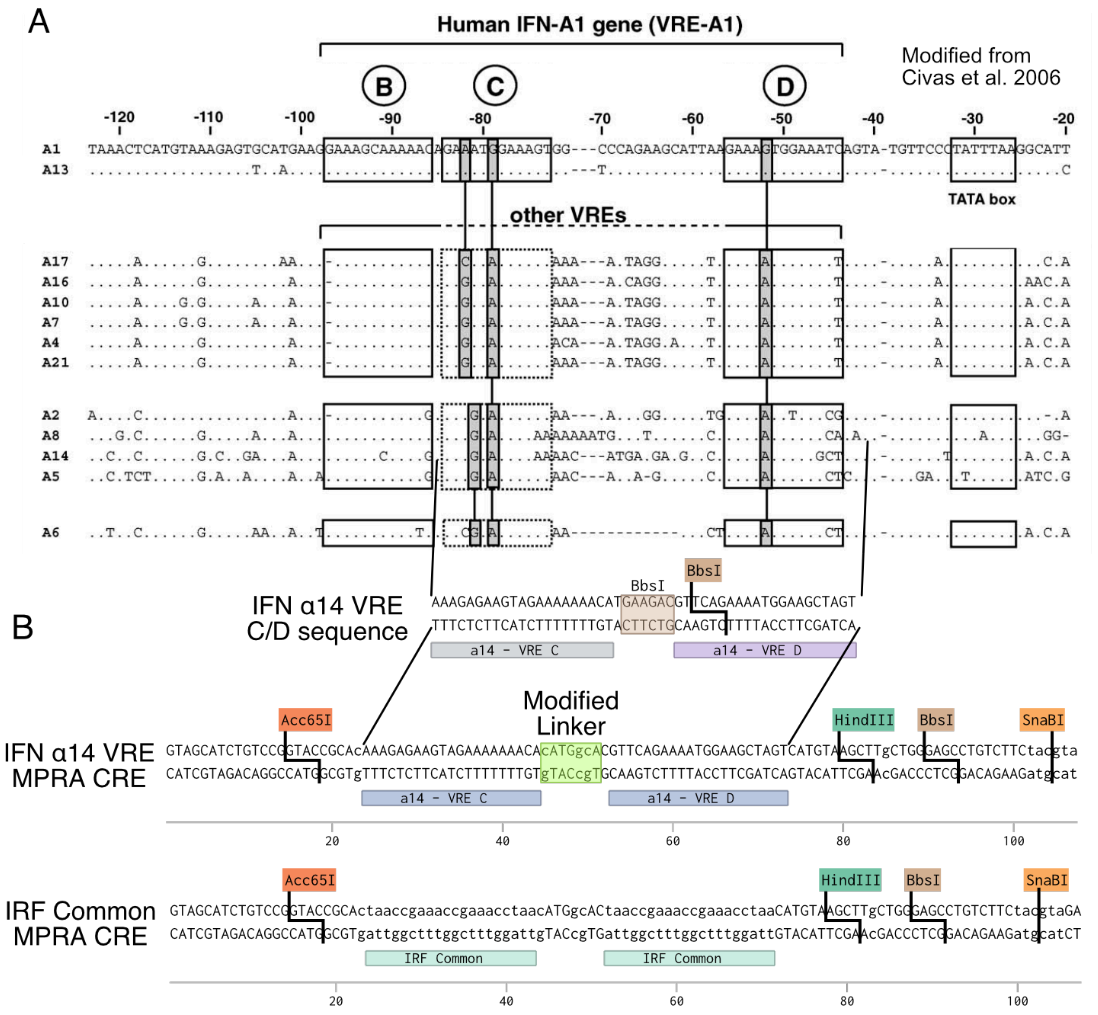


SNV binding site seeds

Probe Name	Sequence
WBe1hA4hB4f0	GTCCCGAAACCGAAACCGTC
WBe0hA4hB4f0	TAACCGAAACCGAAACCTAA
WBe0hA0hB0f0	TAACCGATACCGATACCTAA
WB3e0hA4hB4f2	TAACCGAAACCCGAAACCTAA

Figure 4.2 - The IRF-MPRA design features Cis-Regulatory Elements based on the interferon Viral Response Elements.

(A)The type-I-interferons feature a conserved CRE called the Viral Response Element consisting of multiple IRF binding sites. (B)We substitute IRF binding sites of interest into the existing IFNα-14 context at the VRE-C and VRE-D sites. (C) As part of the MPRA design we include non-SNV sites that were assayed in our IRF3/5/7 PBMs.


A
Human IFN-A1 gene (VRE-A1)
Modified from Civas et al. 2006
B
C
D
-120
-110
-100
-90
-80
-70
-60
-50
-40
-30
-20
A1
A13
TATA box
other VREs
A17
A16
A10
A7
A4
A21
A2
A8
A14
A5
A6
B
IFN α14 VRE C/D sequence
BbsI
AAAGAGAAGTAGAAAAAAACATGAAGACGTTCAGAAAATGGAAGCTAGT
TTTCTCTTCATCTTTTTTTGTACTTCTGCAAGTCTTTTACCTTCGATCA
a14 - VRE C
a14 - VRE D
Modified Linker
IFN α14 VRE MPRA CRE
Acc65I
HindIII
BbsI
SnaBI
GTAGCATCTGTCCGGTACCGCAcAAAGAGAAGTAGAAAAAAACAcATGgcACGTTCAGAAAATGGAAGCTAGTCATGTAAGCTTgCTGGGAGCCTGTCTTCtacgta
CATCGTAGACAGGCCATGGCGTgTTTCTCTTCATCTTTTTTTGTgTACcgTGCAAGTCTTTTACCTTCGATCAGTACATTCGAAcGACCCTCGGACAGAAGatgcat
20
40
60
80
100
IRF Common MPRA CRE
GTAGCATCTGTCCGGTACCGCActaaccgaaaccgaaacctaacATGgcACtaaccgaaaccgaaacctaaCATGTAAGCTTgCTGGGAGCCTGTCTTCtacgtaGA
CATCGTAGACAGGCCATGGCGTgattggctttggctttggattgTACcgTGattggctttggctttggattGTACATTCGAAcGACCCTCGGACAGAAGatgcatCT
IRF Common
IRF Common

C

Non-SNV binding sites		PBM Z-score		
Site Name	**Sequence**	**IRF3**	**IRF 5**	**IRF7**
IRF3 Specific	GTCAGGAGAAGGAAACCTTC	10.4	1.28	2.13
IRF5 Specific	TAACCCATACCGATACCTAA	2.67	10.6	0.70
IRF7 Specific	GTCCCGAAACCCGAAAACGTC	3.41	3.41	9.34
IRF7 Specific 2bp spacer	GTCCCGAAACCGAAAACGTC	7.35	6.62	13.9
IRF Common	TAACCGAAACCGAAACCTAA	11.4	16.4	16.6
IRF Common GTC Flank	GTCCCGAAACCGAAACCTAA	Not measured by PBM		
Null site	GTCAGCAGTAGCAGTCCGTC	-0.69	-0.62	0.01

Figure 4.3 - Diagram of IRF-MPRA experimental subsets

Our MPRA features multiple binding site layouts. The viral response element subset features two IRF binding sites separated by a spacer. The ablation layout substitutes the “null” sequence for each binding site position as illustrated with red-X’s. We test the effect of binding site orientation by reverse-complimenting the IRF binding sites. Lastly, we include a variable spacer design.

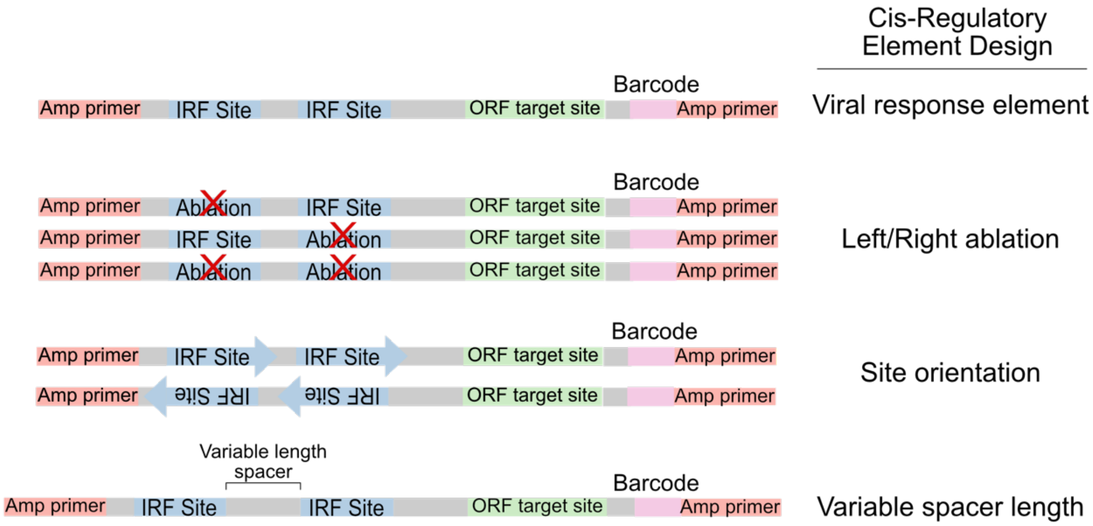


Figure 4.4 - MPRA experimental workflow

MPRA experiments begin with an MPRA pool and an empty destination vector. The CRE insertion step uses restriction digest and ligation to insert the CRE cassette from the MRA pool into the destination vector. Next, an open reading frame (ORF) that contains a minimal promoter and eGFP gene is inserted into the CRE containing vector pool. This library is then grown a large scale and purified before transfection into cells. MPRA plasmid and mRNA is isolated from transfected cells and processed for high-throughput sequencing.

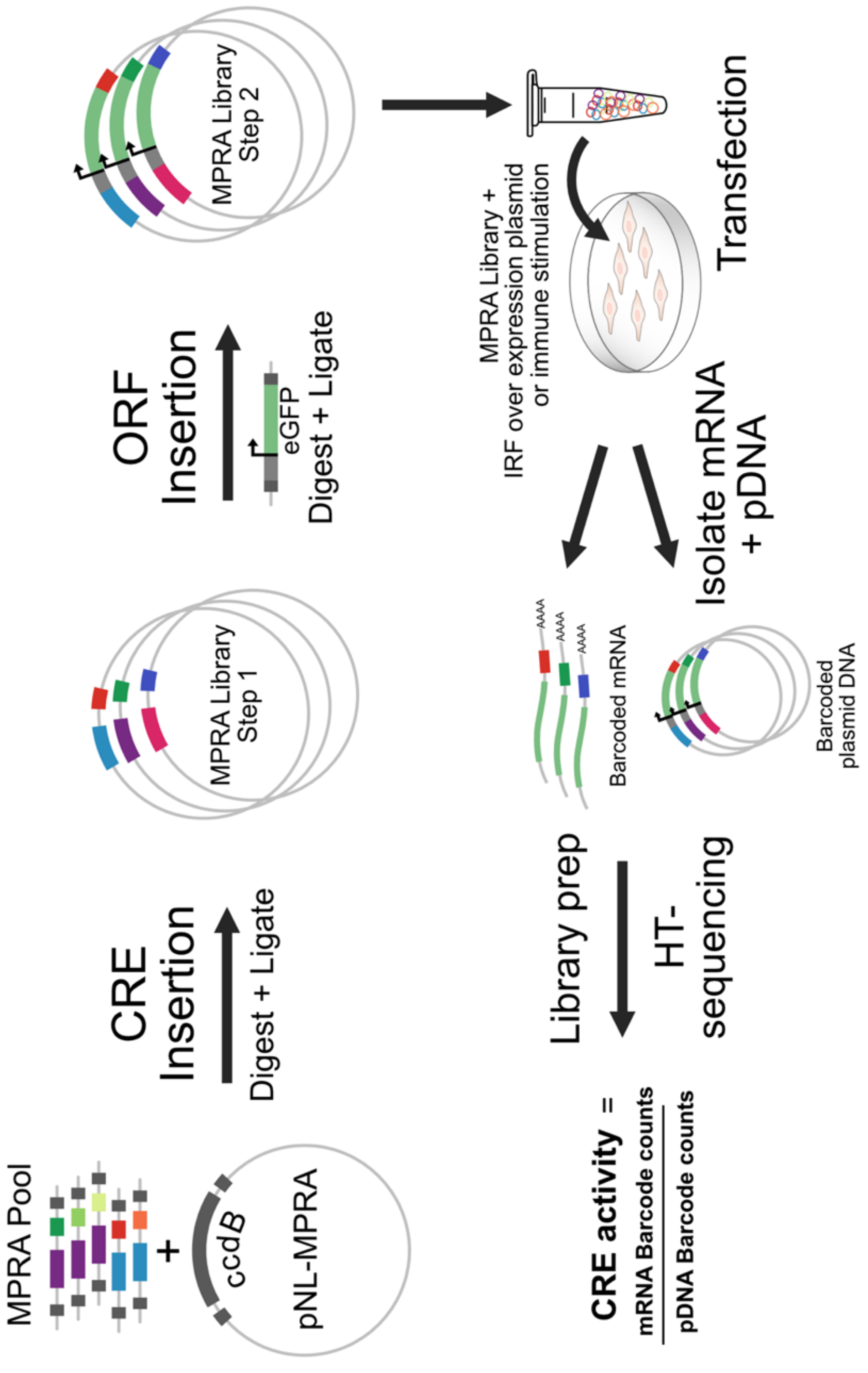

MPRA Pool
+
ccdB
pNL-MPRA
CRE Insertion
Digest + Ligate
MPRA Library Step 1
ORF Insertion
eGFP
Digest + Ligate
MPRA Library Step 2
MPRA Library + IRF over expression plasmid or immune stimulation
Transfection
Isolate mRNA + pDNA
AAAA
AAAA
AAAA
Barcoded mRNA
Barcoded plasmid DNA
Library prep
HT-sequencing
CRE activity = mRNA Barcode counts / pDNA Barcode counts

Figure 4.5 - MPRA CRE insert architecture

Sequence diagram of the IRF-MPRA CRE insert design. The CRE is inserted into the pNL-MPRA backbone using Acc65I and XbaI sites. This is followed by ORF insertion using BbsI sites. A SnaBI site is present to linearize any remaining ORF insert library that did not receive an ORF during cloning.

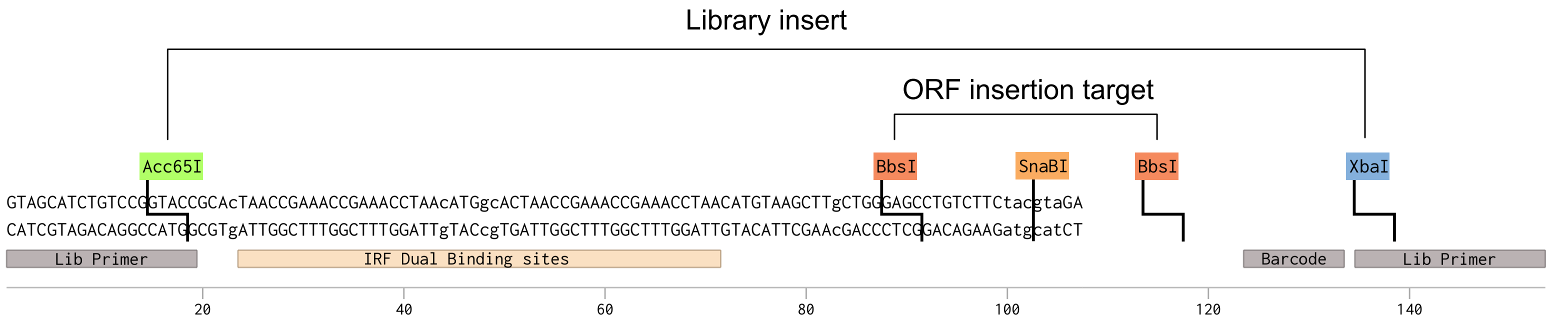

Library insert
ORF insertion target
Acc65I
BbsI
SnaBI
BbsI
XbaI
GTAGCATCTGTCCGGTACCGCAcTAACCGAAACCGAAACCTAAcATGgcACTAACCGAAACCGAAACCTAACATGTAAGCTTgCTGGGAGCCTGTCTTCtacgtaGA
CATCGTAGACAGGCCATGGCGTgATTGGCTTTGGCTTTGGATTgTACcgTGATTGGCTTTGGCTTTGGATTGTACATTCGAAcGACCCTCGGACAGAAGatgcatCT
Lib Primer
IRF Dual Binding sites
Barcode
Lib Primer
20
40
60
80
100
120
140

Figure 4.6 - umiCRE-seq protocol workflow

Isolated MPRA plasmid DNA (pDNA) or mRNA have similar workflows. First UMIs (UMI1) are incorporated via a 1st strand synthesis step using a reverse transcriptase (mRNA) or single cycle PCR reaction (pDNA). An adapter ligation restriction site is included in this step (R1). Next, second-strand amplification primers are added that contain a second adapter ligation site (R2), as well as R1 containing amplification primers. NOPE oligos are added to sequester any remaining UMI containing primer. Next, the amplified sample is restriction digested and adapters are ligated. The adapter closest to the MPRA barcode contains a second UMI (UMI2). Both adapters contain NEB sequencing primers that allow NEBnext multiplexing kits to be used for sample labeling. Finally using the NEB mutiplexing primers, the library is amplified and then finished using the standard illumina P5 and P7 primers.

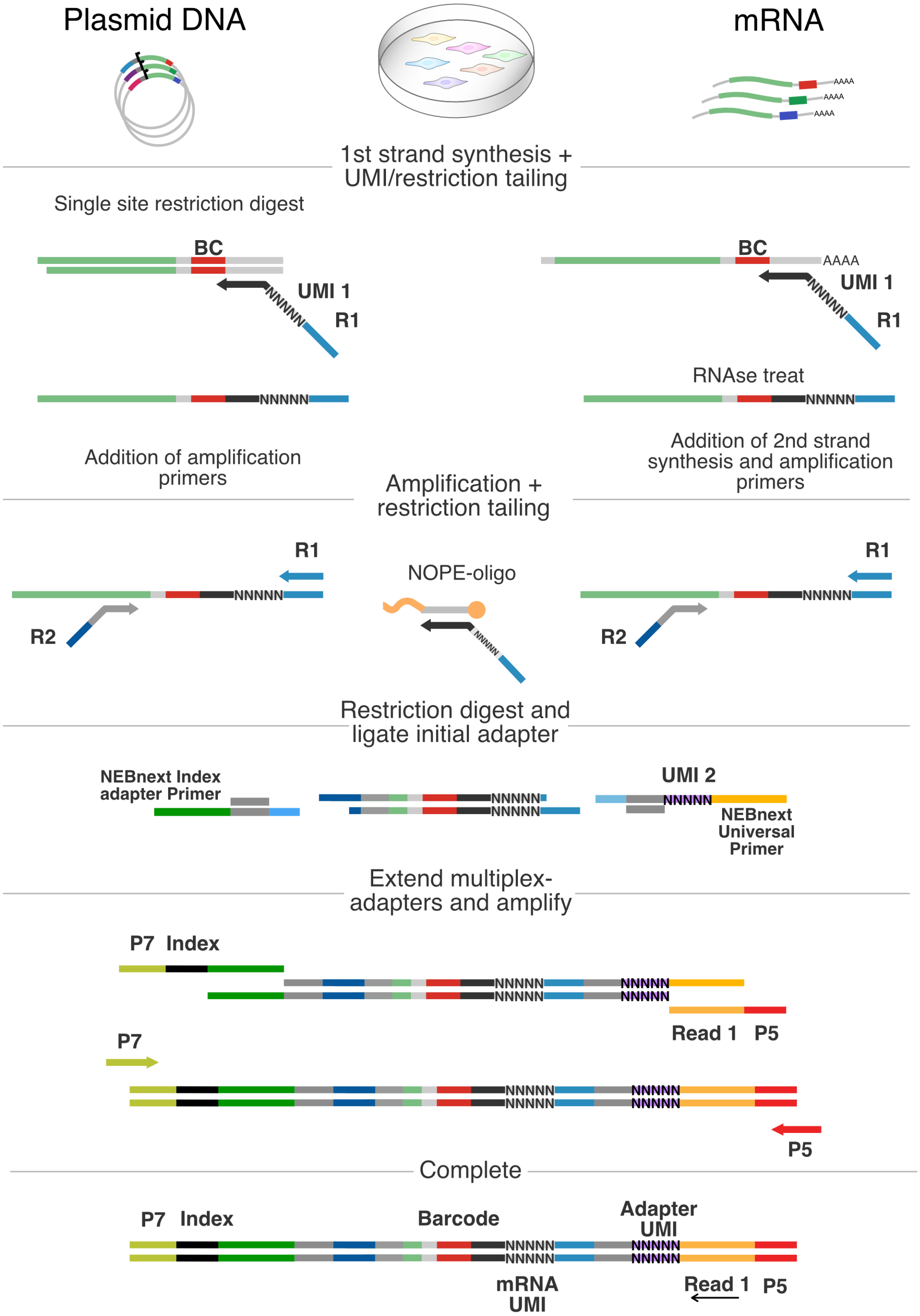

Plasmid DNA
mRNA
AAAA
AAAA
AAAA
1st strand synthesis +
UMI/restriction tailing
Single site restriction digest
BC
UMI 1
NNNNN
R1
BC
AAAA
UMI 1
NNNNN
R1
NNNNN
RNAse treat
NNNNN
Addition of amplification
primers
Amplification +
restriction tailing
Addition of 2nd strand
synthesis and amplification
primers
R1
NNNNN
R2
NOPE-oligo
NNNNN
R1
NNNNN
R2
Restriction digest and
ligate initial adapter
NEBnext Index
adapter Primer
NNNNN
NNNNN
UMI 2
NNNNN
NEBnext
Universal
Primer
Extend multiplex-
adapters and amplify
P7 Index
NNNNN
NNNNN
NNNNN
NNNNN
Read 1 P5
P7
NNNNN
NNNNN
NNNNN
NNNNN
P5
Complete
P7 Index
Barcode
Adapter
UMI
NNNNN
NNNNN
NNNNN
NNNNN
mRNA
UMI
Read 1 P5

Figure 4.7 - scRNA-seq UMI addition schematic

Schematic of UMI incorporation procedure used for scRNA-seq from Islam et al. (2014). This process is incompatible with pDNA as it relies on RNA hydrolysis to remove unincorporated primers and it also requires a template-switching enzyme (i.e. reverse transcriptase). **Taken from Islam et al.** (2014)

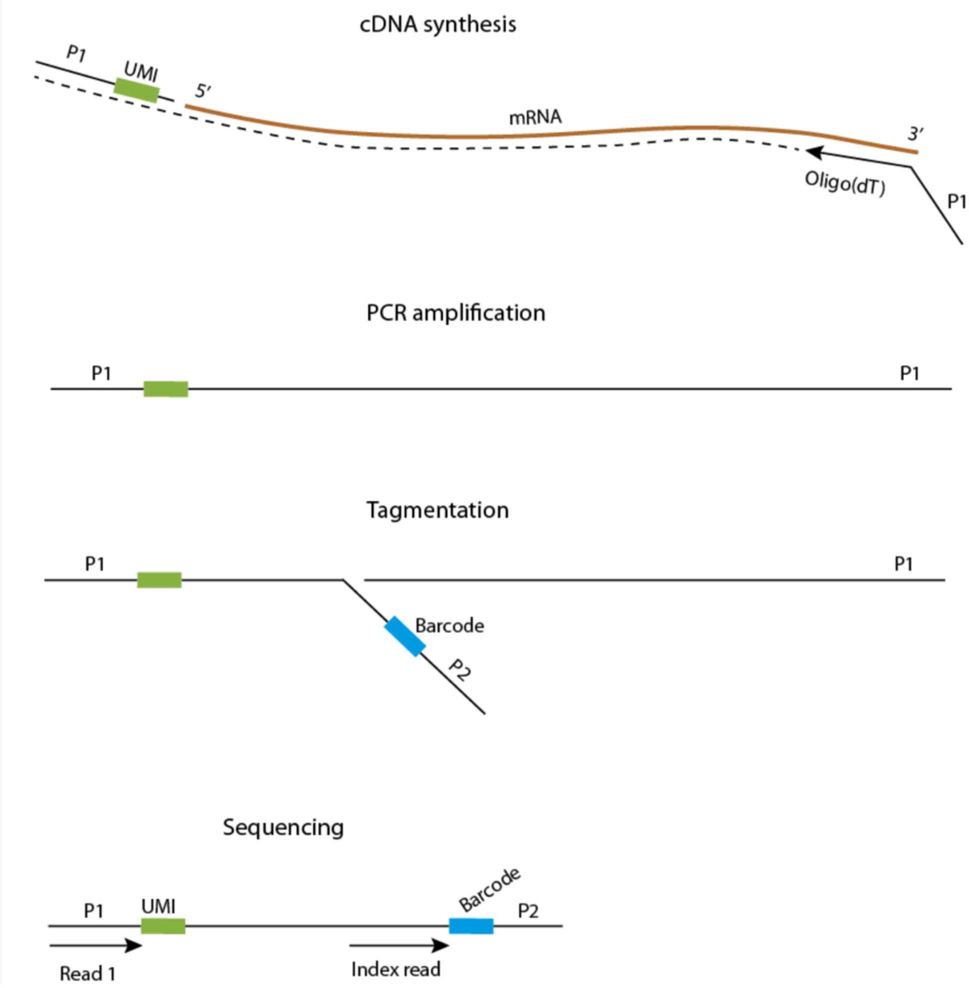
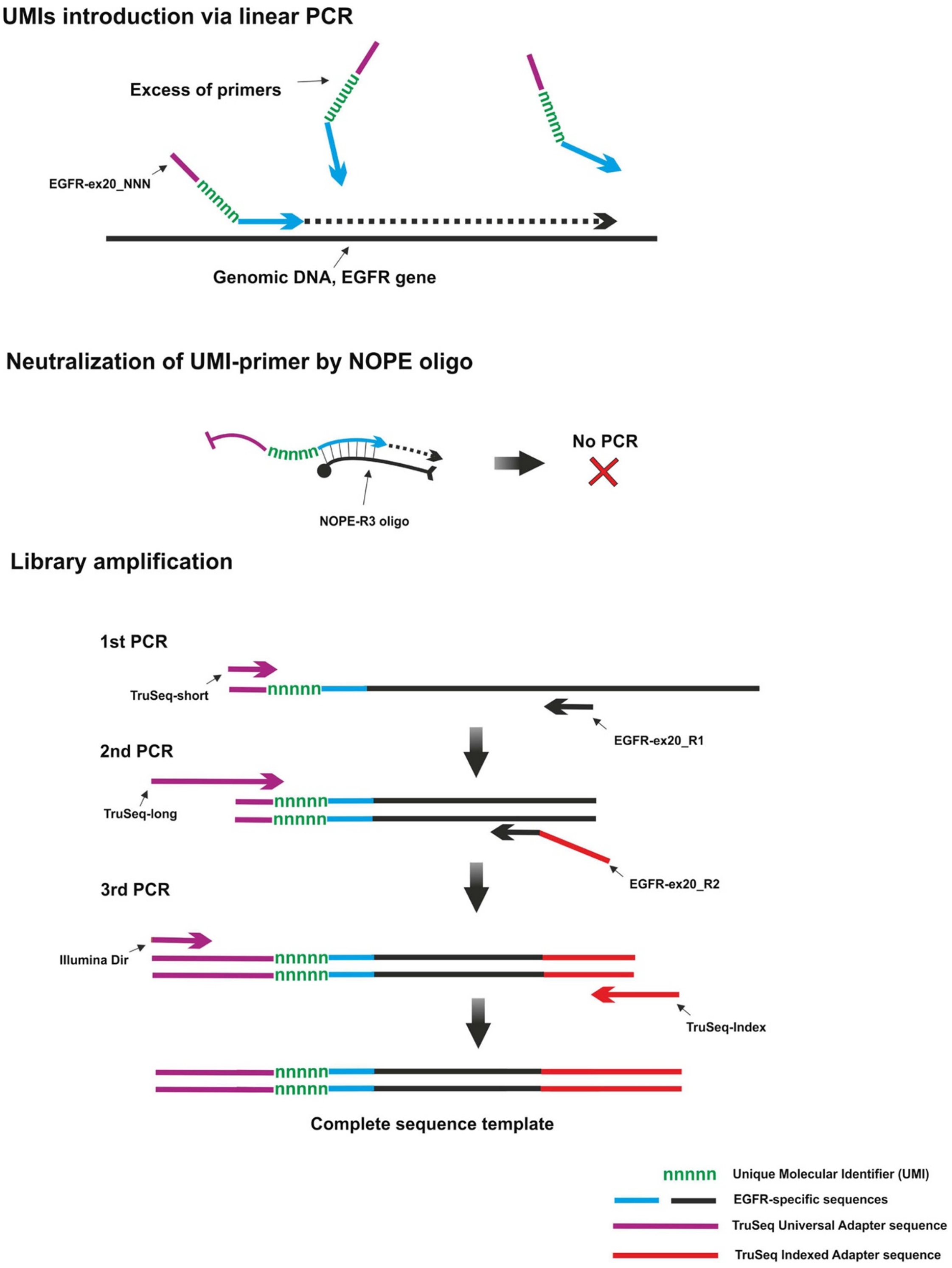


Figure 4.8 - NOPE workflow schematic

Taken from Shagin et al. (2017)**.** NOPE oligos neutralize UMI-primers by sequestering them and preventing PCR extension. NOPE are complementary to the gene specific annealing region present in the UMI containing amplification primer. An essential component of the NOPE process is a “nonsense” region present in the NOPE oligo that prevents sequence specific annealing once the NOPE oligo has been incorporated into excess UMI primer during PCR. An oligo modification on the NOPE oligo also prevents extension of the UMI primers.



UMIs introduction via linear PCR

Neutralization of UMI-primer by NOPE oligo

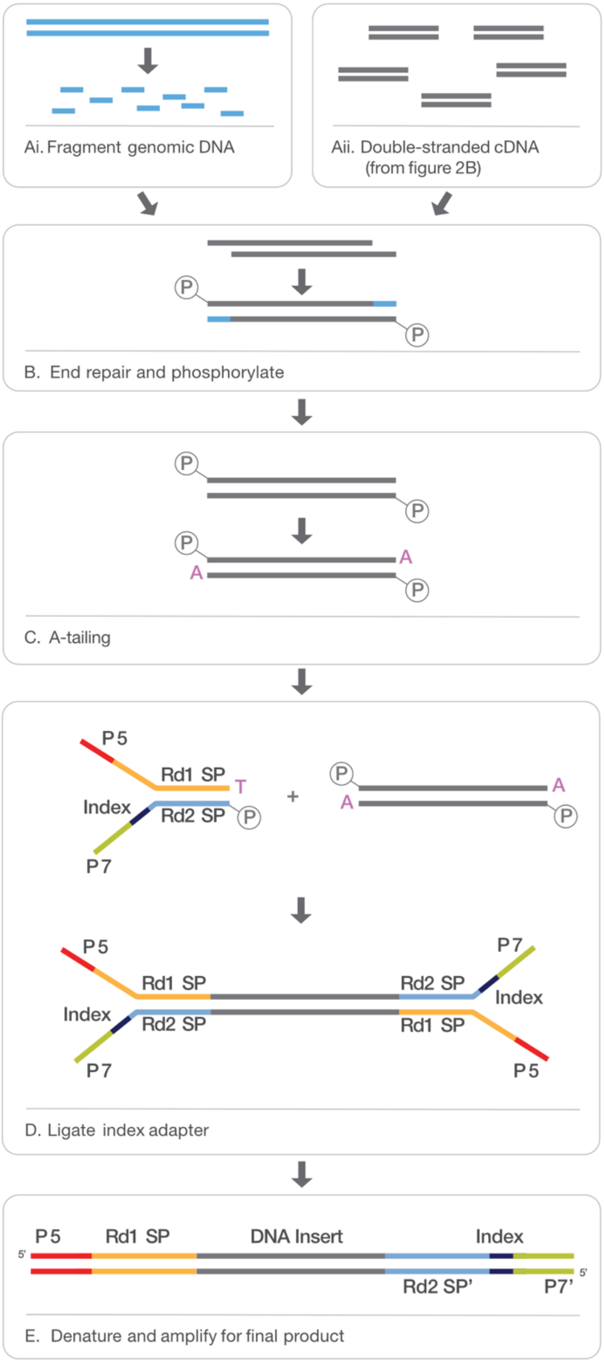
Library amplification

Complete sequence template

nnnnn Unique Molecular Identifier (UMI)
EGFR-specific sequences
TruSeq Universal Adapter sequence
TruSeq Indexed Adapter sequence

Figure 4.9 - Illumina sequencing library workflow

Illumina sequencing uses fragmented starting DNA and T/A ligation which are incompatible with Unique Molecular Identifiers and inefficient for MPRA experiments. T/A ligation utilizes single nucleotide base pair overhangs to add adapters to the sequencing library. **Modified from Illumina.com**



Modified from Illumina.com

Figure 4.10 - Diagram of the proposed UMI containing sequencing adapter.

Graphical depiction of the partially double-stranded UMI containing adapter. The double-stranded constant annealing site allows for efficient T4 ligase activity when ligating this adapter to the restriction digested MPRA library. The direction of the sequencing read is indicated beneath the Read 1 region.

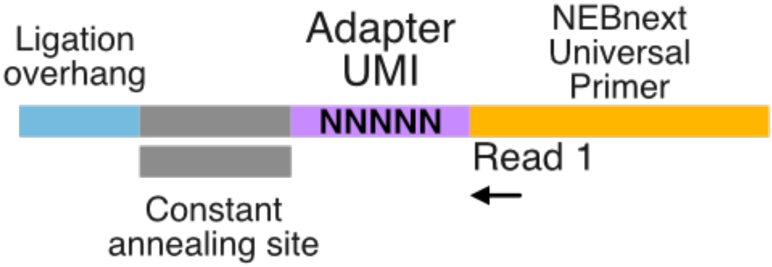


Figure 4.11 - Successful cluster differentiation requires nucleotide diversity

Low complexity samples can lead to problems with cluster differentiation leading to loss of sequencing data. This figure illustrates the challenge of low diversity samples during early cluster calling sequencing cycles. Circles represent sequencing clusters on an Illumina flow cell. Enlarged cluster circles present in cycles 1-4 illustrate how the Illumina sequencing machine may fail to differentiate adjacent clusters that share the same base call during a cycle. Adjacent clusters that look the same cannot be differentiated by the sequencing machine and are discarded as bad data. This illustrates the need for diverse samples. **Modified from Krueger et al.** (2011)

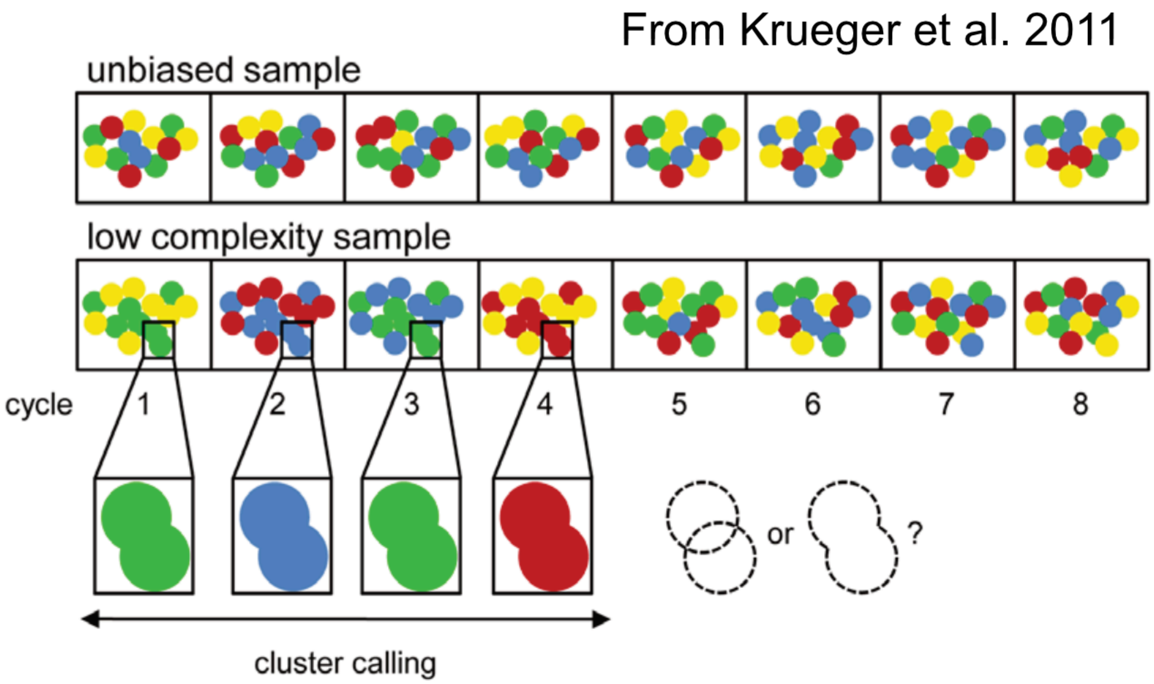


Figure 4.12 - Overview of NEBnext library preparation

NEBnext uses a two-step adapter addition process that allows for straightforward sample multiplexing. Our custom MPRA adapters are compatible with standard NEBnext multiplexing indices. Adapter ligated MPRA sequencing libraries could utilize NEB components used in the PCR enrichment step onward. During PCR enrichment, multiplexing indices can be added to different samples. **Adapted from NEB.com**

From NEB.com

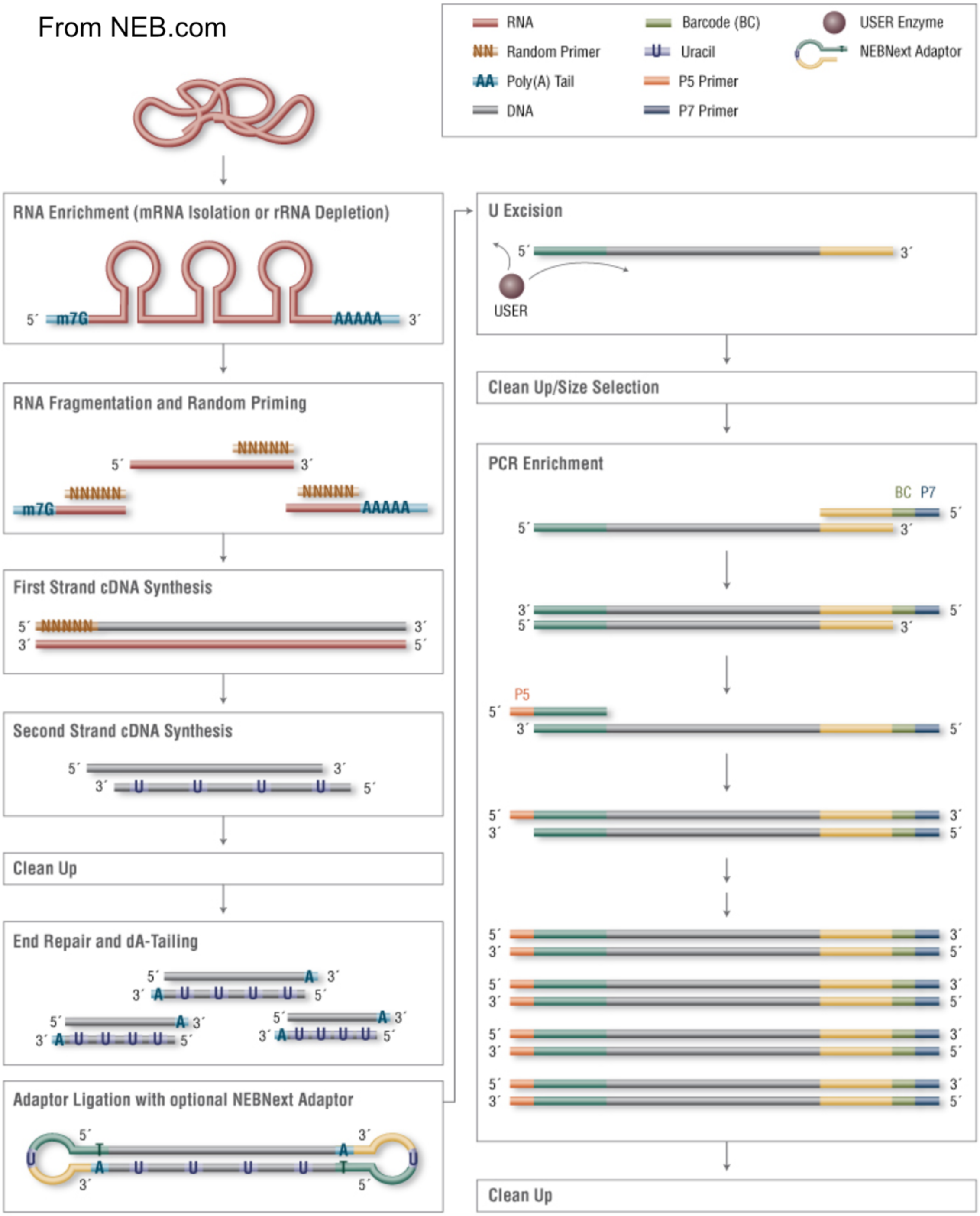

RNA
NN Random Primer
AA Poly(A) Tail
DNA
Barcode (BC)
U Uracil
P5 Primer
P7 Primer
USER Enzyme
NEBNext Adaptor
RNA Enrichment (mRNA Isolation or rRNA Depletion)
5´ m7G AAAAA 3´
RNA Fragmentation and Random Priming
NNNNN
5´ 3´
m7G NNNNN
NNNNN AAAAA
First Strand cDNA Synthesis
5´ NNNNN 3´
3´ 5´
Second Strand cDNA Synthesis
5´ 3´
3´ U U U U 5´
Clean Up
End Repair and dA-Tailing
5´ A 3´
3´ A U U U U U 5´
Adaptor Ligation with optional NEBNext Adaptor
5´ T A 3´
U A U U U U T U
3´ 5´
U Excision
5´ 3´
USER
Clean Up/Size Selection
PCR Enrichment
BC P7
5´
5´ 3´
3´ 5´
5´ 3´
P5
5´
3´ 5´
5´ 3´
3´ 5´
Clean Up

Figure 4.13 - Histogram of preliminary MPRA library sequencing data.

Histogram showing the frequency of barcode counts (i.e. count of barcode counts). Each bar displays the number of retrieved barcodes that contain a given number of sequencing reads. For example, 154 barcodes had a sequencing read count of 19 reads per barcode.

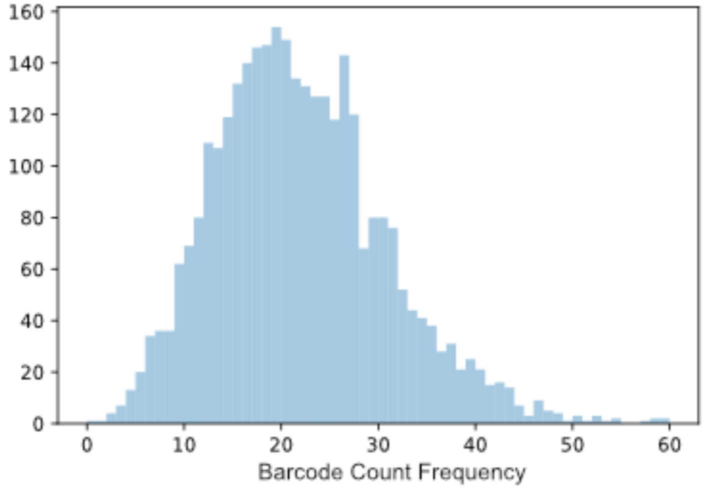
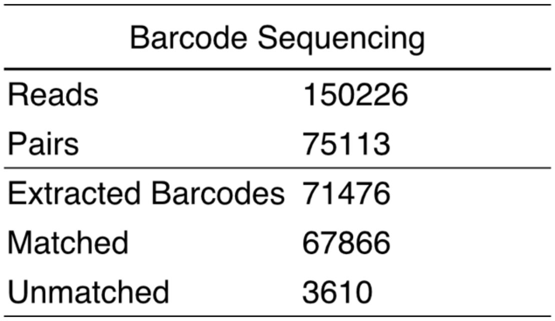


Table 4.4.1 – Summary table of MPRA library preliminary barcode sequencing.

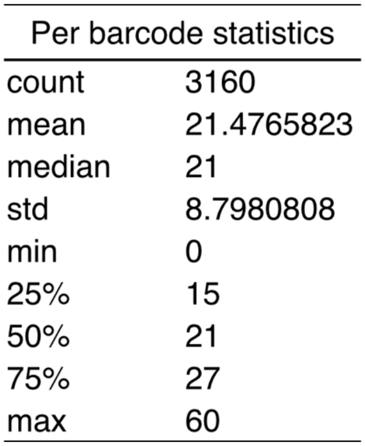
For quality control, we sequenced the initial CRE insertion MPRA library. MGH sequencing core Illumina paired end sequencing resulted in 150,226 reads sorted into 75,113 pairs. From these data we extracted 71,476 barcodes with 67,866 of them matching our designed sequences, 3610 extracted barcodes deviated from the expected sequences.



Barcode Sequencing	
Reads	150226
Pairs	75113
Extracted Barcodes	71476
Matched	67866
Unmatched	3610

Table 4.4.2 – Sequence read per barcode descriptive statistics

Preliminary MPRA library quality control sequencing resulted in retrieval of 3,160 barcodes. The mean number of sequencing reads per barcode was 21 per barcode (SD ± 8.79). One barcode of the designed 3161 barcodes was not retrieved from sequencing.



Per barcode statistics	
count	3160
mean	21.4765823
median	21
std	8.7980808
min	0
25%	15
50%	21
75%	27
max	60

5 CHAPTER FIVE

5.1 Summary of findings

IRF3/5/7 are central regulators of the host-defense program to pathogens (Honda and Taniguchi, 2006; Lazear et al., 2013; Stetson and Medzhitov, 2006a). IRF3/5/7 are homologous transcription factors with highly-conserved DNA-binding domains and can all bind the consensus IRF binding site (5'-GAAANNGAAA-3'). Despite their ability to bind the same consensus sequence, IRF3/5/7 drive overlapping and distinct gene programs.

In Chapter 3, using PBMs we assessed the ability of inherent IRF3/5/7 DNA-binding differences to define dimer-specific gene regulation. The differences we found provide a mechanism for explaining the differential IRF-dependent regulation of immune genes. We characterized the DNA-binding preferences of IRF3/5/7 homodimers, and demonstrated that dimer-specific binding can promote dimer-specific gene expression. These results support the claim that DNA-binding differences between IRF3/5/7 are sufficient to induce unique target gene sets. For example, we found that IRF5 is largely excluded from binding IRF regulatory elements from the human and mouse IFNα promoters. This conserved exclusion is mediated by selection against specific sequences by IRF5. Mutational analysis of IRF5 and IRF7 found a single amino acid residue that contributes to this binding selectivity. Lastly, we found evidence of clear affinity-independent mechanisms in IRF-dependent gene regulation. More specifically, we found that DNA sequence features in IRF binding sites can enhance reporter gene activity at the expense of binding affinity. These expression modulating sequence features can be regions flanking the core IRF motif and may have implications for our understanding of IRF-dependent gene regulation as a whole.

In Chapter 4, we propose a modified Massively Parallel Reporter Assay (MPRA) protocol that utilizes Unique Molecular Identifiers (UMIs) to increase the potential quantifiability of MPRAs. We describe key technical hurdles and address these issues with proposed solutions. The groundwork provided in chapter 4 provides a foundation for future graduate research.

5.2 Discussion

5.2.1 IRF-dependent gene regulation

Our PBM data in Chapter 3 shows clear DNA-binding differences between IRF3/5/7. Each IRF maintains the ability to bind the same common IRF binding site (5'-CCGAAACCGAAACC-3') while differentially tolerating deviations from this high-affinity IRF-site. Our reporter data found that preferential IRF-binding to IRF-sites can drive reporter gene expression in an IRF-homolog specific manner. For example, IRF3 specific binding sites only drove reporter gene expression in cells over expressing phosphomimetic IRF3. This suggests that sequence differences in IRF binding sites are sufficient to drive gene expression in an IRF-homolog specific manner. Despite the fact that binding affinity does not appear to predict the expression level in our reporter assays, the absence of binding (i.e., the inability of IRF7 or IRF5 to bind to our IRF3-specific site) appears to predict the inability to drive reporter gene expression. Therefore, we conclude based on our PBM data that IRF3 and IRF7 can regulate the type-I-interferons (IFNα and

IFNβ) as well as CXCL10, and IL10, which is consistent with the existing literature (Honda and Taniguchi, 2006)

As summarized above, IRF5 was shown to not bind the human IFNα VRE regions with strong affinity, suggesting that IRF5 homodimers may not directly regulate the IFNα genes. It is possible that aggregate, low affinity DNA binding by IRF5 to the IFNα promoters could drive gene expression a mechanism which has been shown for other TFs (Crocker et al., 2015). In Chapter 3 we see that IRF5 shows appreciable binding to some of the mouse IFNα promoters, suggesting that the role of IRF5 in IFNα regulation may differ between humans and mice, potentially complicating the combined interpretation of human and mouse derived data across the field. We also find that IRF5 binds to the IFNβ promoter sequences and IRF5 has been shown to induce IFNβ production (del Fresno et al., 2013). Given this, IRF5 may indirectly induce the expression of IFNα genes via production of IFNβ which activates the ISGF3 TF complex (STAT1-STAT2-IRF9) through the Interferon receptor (Fink and Grandvaux, 2013). ISGF3 is known to bind ISREs (Cheon et al., 2013) and activation of this complex, downstream of IRF5 induced IFNβ signaling, could contribute to IFNα gene expression. Many viruses impair IRF-dependent gene activation as a mechanism of infection, making cross-talk between anti-pathogen signaling pathways an important component of a robust immune response (Chan and Gack, 2016; Fensterl et al., 2015; Kawai and Akira, 2011). It has been suggested that IRF5 binding may be facilitated by RelA and that IRF5 may also be indirectly recruited to RelA binding sites (Saliba et al., 2014). This mechanism could provide another route for IRF5 activation to contribute to the interferon response.

Interactions between IRF homologs both through heterodimerization and potential competition for binding sites is an additional layer of IRF-dependent gene regulation. As discussed in Chapter 3 (Section 3.3 - Limitations), IRF3/5/7 heterodimerization is an important aspect of IRF-dependent gene regulation. Heterodimerization is a mechanism that could expand the DNA-binding repertoire of the IRFs, notably IRF5. IRF3+7 heterodimers are thought to contribute to the temporal dynamics of IFNα expression during the anti-viral response with the ratio of IRF3:IRF3+7:IRF7 regulating which set of IFNα genes are expressed at a given time (Genin et al., 2009a). Future experiments to understand IRF heterodimerization may be able to use both purified or endogenously activated IRFs (nextPBM) to study heterodimer DNA-binding specificity. However, these experiments will need to establish the stability of IRF heterodimers, potentially using EMSAs or native westerns to assess the fraction of heterodimer to homodimer. IRF3/5/7 have been described to function as both transcriptional activators and repressors in different promoter contexts (Barnes et al., 2002b; Honda et al., 2006). Competition between IRFs at IRF binding sites, in conjunction with differential recruitment of transcriptional activators and repressors (discussed below), may contribute to these dichotomous regulatory effects. Unfortunately, research in the Siggers lab suggests PBMs are not able to assess competition in protein binding, but traditional competition EMSAs may provide insight for sequences of interest.

As mentioned above, Saliba et al. (2014) and Krausgruber et al. (2010) suggest that IRF5 and RelA interact at multiple pro-inflammatory gene promoters (TNFα, IL-6, IL-1a) after LPS stimulation in macrophages (Saliba et al., 2014). This is an interaction that the authors suggest could explain the dual role of IRF5 as an activator and inhibitor of gene expression. IRF5 ChIP-seq data and motif analysis from Saliba et al. (2014) fails to retrieve canonical ISRE motif, but suggests that IRF5 may bind a composite ETS-IRF site with PU.1; although, the ETS-IRF site may be utilized by multiple ETS factors in addition to PU.1. IRF4 and IRF8 are known to bind both ETS-IRF and NFAT-IRF composite elements cooperatively with PU.1 or NFATc2 respectively. If IRF5 interacts cooperatively with ETS factors, in a similar manner to IRF4 and IRF8, then this would be an additional mechanism of IRF5dependent gene regulation. In our PBM data, IRF5 shows a strong binding preference for the 3' half site of the ISRE with little specificity in the 5' half site, which is consistent with previous data sets Jolma:2013fh}. The decreased specificity present in the PBM-derived IRF5 motif corresponds to the ETS portion of the ETS-IRF motif found by Saliba et al. (2014) in IRF5 only ChIP-seq peaks. Given this, IRF5 may be able to use non-ISRE type binding sites as an asymmetric homodimer, or in conjunction with other binding partners. In this context, IRF5 may compete for occupancy of PU.1:IRF4 or IRF8 binding sites. Interestingly, IRF4 and IRF5 have been shown to compete for access to the MyD88 adapter protein down stream of TLR signaling which could provide additional layers of regulation in a cell-type and stimulus-specific manner.

5.2.2 Evolution of the IRFs and IFNα

IRF DNA-binding domains (DBDs) are highly conserved across many species (Figure 5.1). The conserved amino acids include the lysine residue (IRF5 K96) shown to contribute to the specificity of IRF5 seen in our PBM data. In Chapter 3, we note that the exclusion of IRF5 binding at the IFNα promoter sequences on our PBM is evolutionarily conserved between human and mouse, resulting from a nearly complete absence of cytosine bases flanking the IRF 5'-GAAA-3' core that are both highly preferred by IRF5. A more in-depth phylogenetic analysis of the type-I-interferon gene promoters is complicated by a complex evolutionary history involving both gene conversion and gene duplication (Krause and Pestka, 2015; Redmond et al., 2019; Xu et al., 2013), making precise inference about the evolution of IRF sites in these promoters prohibitively difficult. However, there are multiple genes in human and mouse that are known targets of IRF3/5/7 that have 5'-CGAAAC-3' containing ISREs as evident in our PBM data (IFNβ, CXCL10, IL10) and the literature (Koshiba et al., 2013). A single base-pair mutation of a 5'-GAAA-3' flanking cytosine would allow IRF5 binding and maintain IRF3/7 binding to that site. If the immune role of IRF5 is to drive a distinct pro-inflammatory gene program as suggested in the literature (Saliba et al., 2014; Weiss et al., 2015), then maintaining regulatory separation between IRF3/7 (anti-viral response) and IRF5 (inflammatory response) may be evolutionarily advantageous. Our findings that IRF5 homodimers do not bind strongly to the IFNα promoters of human and mice suggests an evolutionarily conserved mechanism to distinguish these homodimers that is not shared across all IRF-dependent genes. Future studies that directly examine the possible alternative IRF5-containing complexes will be important to understand its conserved role in IFNα gene regulation.

5.2.3 Potential mechanisms of affinity-independent ISRE function

As summarized above, we found clear evidence of affinity-independent mechanisms of IRF transcriptional regulation. Understanding the relationship between TF DNA-binding and transcriptional activity is an important pursuit in molecular biology. A complete mechanistic description of IRF-dependent gene regulation is beyond the scope of this dissertation, but there are many avenues for future work.

As mentioned in Chapter 3, mechanisms for DNA-sequence-based allosteric regulation of transcription factors have been described for multiple TFs (Mazumder et al., 2017; Watson et al., 2013). A structural feature of the IRF transcription factors that may play a role in allosteric regulation is a minor-groove contacting loop (L1) of the IRF DBD which contacts DNA bases adjacent to the core GAAA (De Ioannes et al., 2011; Panne et al., 2007) This loop has been shown to change conformation between a DNA bound state and unbound state and a conserved histidine in this loop (hIRF3-H40; hIRF5-H49; hIRF7-H46) makes contacts with the 5' A:T bases in the IRF binding sequence (De Ioannes et al., 2011a; Shukla et al., 2012). Given this important structural feature, changes in flanking sequence could alter IRF DBD conformation impacting the overall structure of a dimeric IRF complex. DNA sequence based allostery described for the glucocorticoid receptor (GR) involves a structural transmission of the DNA sequence readout through the dimerization interface of the GR complex, ultimately leading to differential trans-activation potential (Meijsing et al., 2009; Watson et al., 2013). Although the crystal structure for a full-length IRF protein has not been solved, the IRFs have two well-structured domains (the DBD and the IAD/AID) that are connected together by a linker that is suggested to be semi-structured. Linkers between modular protein domains are capable of transmitting allosteric information to influence a protein's function (Ma et al., 2011). Similar to GR, subtle allosteric perturbations from DNA sequence binding may alter the conformation or dynamics of the IRF C-terminal protein association domains thus impacting the recruitment of transcriptional activators such as CBP and p300.

A speculative model that does not require specific allosteric changes to IRF structural conformation, is a dynamic model where lower-affinity, asymmetric binding to a dimeric IRF site increases the accessibility of an IRF dimer's transactivation domain. In this model, one monomer of the IRF dimer anchors the complex to the DNA while the other facilitates interactions with other proteins via greater binding flexibility due to lower affinity DNA interactions. In our PBM experiments, we found some evidence for asymmetric sequence specificity in IRF binding logos, where the 5' region of the IRF binding site outside of the core had moderately decreased information content. This decrease in specificity could allow more dynamic associations with DNA and cofactors, while not wholly disrupting IRF dimer binding. Given the lack of a full-length IRF crystal structure, it is unclear how the dimerized IRF association domains are oriented in relation to the IRF DBD and how a decrease in binding affinity might influence co-factor recruitment. Determination of a full-length IRF crystal structure would greatly advance IRF-related research on innate immune regulation.

5.2.4 IRF-dependent transcriptional regulation and cofactors

Transcriptional regulation is a multifaceted process that integrates TF DNA-binding, higher order protein-protein interactions, and epigenetic differences in the chromatin landscape. Understanding this complex cellular process is beyond the scope of the work in this dissertation; however, we have made progress by describing key differences in the IRF DNA-binding landscape.

Elucidating how sequence dependent differences in TF-DNA binding may alter cofactor recruitment is an important next step in understanding innate immune regulation. Recruitment of transcriptional activators like CBP, or repressors like

NCoR have been shown to play an important role in IRF-dependent gene regulation (Caillaud et al., 2002; Chen et al., 2008a; Feng et al., 2010). The role of competing inputs, coactivators versus corepressors, in IRF driven gene expression is not well understood. IRF3, IRF5 and IRF7 have been shown to interact with the coactivator CBP (Caillaud et al., 2002; Chen et al., 2008a), with x-ray crystallography evidence for the IRF3-CBP interaction (Qin et al., 2005). IRF5 has shown to interact with the corepressor NCoR/Sin3a/SMRT (Feng et al., 2010) as well as KAP1/Trim28 (Eames et al., 2012). Interestingly, the Trim28-IRF5 interaction described by Eames et al. (2012) was shown to be mediated by the unstructured linker between the IRF5 DBD and IRF association domain (IAD). This provides a potential mechanism for IRFs to function both as repressors or activators of transcription. Speculatively, it is possible that an integration of these activating/inhibiting states is responsible for the affinity-independent regulation of transcription seen in chapter 3.

5.2.5 Outstanding questions and future work

IRF-dependent gene regulation is complex and involves multiple layers of regulation. The research in this thesis set out to characterize intrinsic differences between IRF3/5/7 in order to better understand protein-DNA binding, a foundational layer of gene regulation. Open questions of interest include: 1) How do IRF heterodimers coordinately regulate IRF target-genes? 2) How do IRF3/5/7 differ in their ability to recruit cofactors, and what cofactors do the interact with? 3) What model best explains the relationship between IRF DNA-binding and IRF-dependent gene regulation? 4) How do differences in IRF-DNA binding influence diseases such as Systemic Lupus Erythmatosus, where IRF disregulation is associated with disease risk? These questions may be starting points for future research in the field of IRF-dependent gene regulation.

The MPRA experiments proposed in Chapter 4 provide a platform for investigating the relationship between DNA sequence and transcriptional activity. It is likely that systematic investigation of how variations in IRF binding sites can impact transcription can be integrated with DNA binding data, such as presented here, to define the relationship between binding and activity. By looking for MPRA CREs that violate simple models of transcription factor binding and transcriptional activity we may identify interesting exceptions to explore. Follow up could include DNA-protein pull-down assays to identify members of DNA bound regulatory complexes. MPRAs performed by Grossman et al. (2017) used immunopreciptation to pull-down and sequence MPRA plasmids bound by PPARɣ. This process could be performed with modification to identify other members in the DNA-binding complex after initial immunoprecipitation using a panel of antibodies. After careful, biochemical analysis, high-interest CREs could be further explored using X-ray crystallography or Cryo-EM to resolve in high-order protein complexes (Nogales, Scheres, 2015).

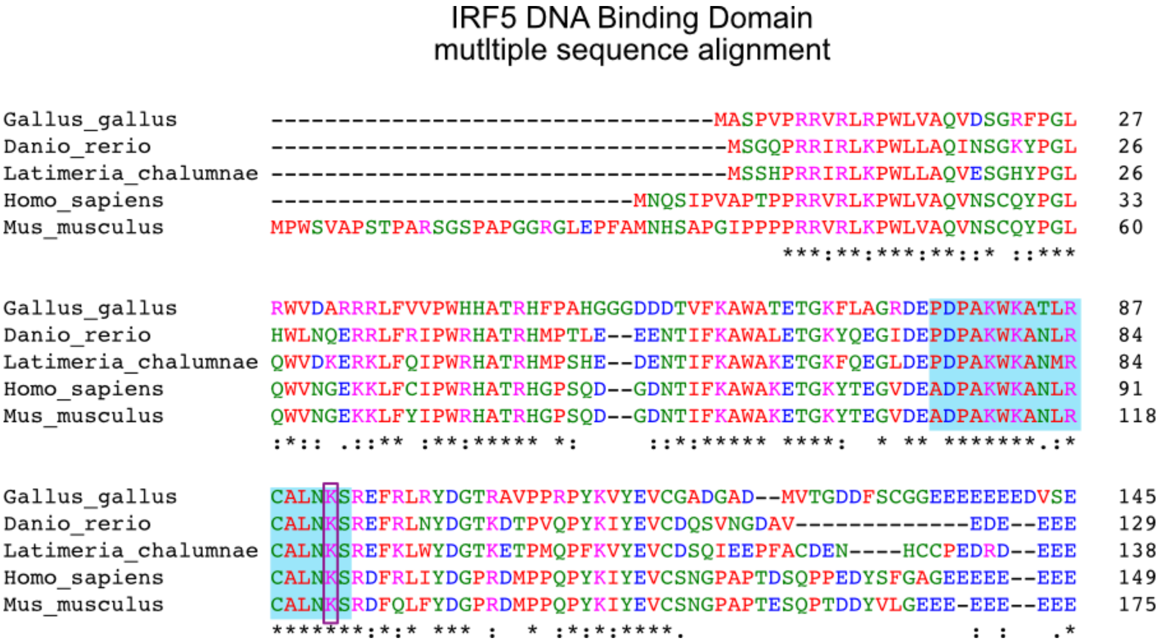
Future work toward understanding IRF3/5/7 transcriptional regulation could also utilize recently developed nuclear extract PBMs (nextPBMs) (Mohaghegh et al., 2019) and cofactor recruitment PBMs (currently in development in the Siggers lab). These exciting new methods provide a platform for investigating the DNA-sequence dependence of higher-order protein-complex assembly. Biochemical studies using IRF3/5/7 constructs with mutated linker regions expressed in mammalian cells could be used to explore potential interactions between DNA-sequence and protein-structure in a systematic way.

5.3 Conclusion

The research in this dissertation was started to better understand the scope of IRF3/5/7 DNA-binding differences and their role in defining dimer-specific target genes. We used protein-binding microarrays (PBMs) to characterize the DNA-binding landscape of IRF3/5/7 dimers and identified key differences in DNA-binding specificity between IRF3, IRF5 and IRF7. We integrated PBM data with reporter assay data and found evidence for affinity-independent mechanisms of IRF-dependent transcriptional regulation. Here we proposed a modified MPRA protocol that incorporates Unique Molecular Identifiers with the potential to improve MPRA data quantification. As part of this work we proposed a multipart IRF-MPRA experiment designed to explore the relationship between IRF DNA-binding affinity and transcriptional regulation. Our results provide new insights into the role and limitations of affinity as a distinguishing mechanism of IRF3/5/7 gene expression and function, as well as providing groundwork for future work in this field.

Figure 5.1 – IRF5 DNA binding domain sequence alignment

The DNA binding domain (DBD) of IRF5 is highly conserved across evolutionary time. The α3 alpha helix (blue box), responsible for recognizing the IRF core 5'-GAAA-3' sequence is nearly identical across chicken, zebra fish, coelacanth, human and mouse. The amino acid residue partially responsible for the exclusion of IRF5 is identically conserved across all species (purple box).



IRF5 DNA Binding Domain
mutltiple sequence alignment

```
Gallus_gallus       ---------------------------------MASPVPRRVRLRPWLVAQVDSGRFPGL     27
Danio_rerio         ----------------------------------MSGQPRRIRLKPWLLAQINSGKYPGL     26
Latimeria_chalumnae ----------------------------------MSSHPRRIRLKPWLLAQVESGHYPGL     26
Homo_sapiens        ---------------------------MNQSIPVAPTPPRRVRLKPWLVAQVNSCQYPGL     33
Mus_musculus        MPWSVAPSTPARSGSPAPGGRGLEPFAMNHSAPGIPPPPRRVRLKPWLVAQVNSCQYPGL     60
                                                          ***:**:***:**::* ::***

Gallus_gallus       RWVDARRRLFVVPWHHATRHFPAHGGGDDDTVFKAWATETGKFLAGRDEPDPAKWKATLR     87
Danio_rerio         HWLNQERRLFRIPWRHATRHMPTLE--EENTIFKAWALETGKYQEGIDEPDPAKWKANLR     84
Latimeria_chalumnae QWVDKERKLFQIPWRHATRHMPSHE--DENTIFKAWAKETGKFQEGLDEPDPAKWKANMR     84
Homo_sapiens        QWVNGEKKLFCIPWRHATRHGPSQD--GDNTIFKAWAKETGKYTEGVDEADPAKWKANLR     91
Mus_musculus        QWVNGEKKLFYIPWRHATRHGPSQD--GDNTIFKAWAKETGKYTEGVDEADPAKWKANLR     118
                    :*:: .::** :**:***** *:     ::*:***** ****:  * ** *******.:*

Gallus_gallus       CALNKSREFRLRYDGTRAVPPRPYKVYEVCGADGAD--MVTGDDFSCGGEEEEEEEDVSE     145
Danio_rerio         CALNKSREFRLNYDGTKDTPVQPYKIYEVCDQSVNGDAV-------------EDE--EEE     129
Latimeria_chalumnae CALNKSREFKLWYDGTKETPMQPFKVYEVCDSQIEEPFACDEN----HCCPEDRD--EEE     138
Homo_sapiens        CALNKSRDFRLIYDGPRDMPPQPYKIYEVCSNGPAPTDSQPPEDYSFGAGEEEEE--EEE     149
Mus_musculus        CALNKSRDFQLFYDGPRDMPPQPYKIYEVCSNGPAPTESQPTDDYVLGEEE-EEE--EEE     175
                    *******:*:* *** :  * :*:*:****.                     : :   .*
```

BIBLIOGRAPHY

Alon, S., Vigneault, F., Eminaga, S., Christodoulou, D.C., Seidman, J.G., Church, G.M., Eisenberg, E., 2011. Barcoding bias in high-throughput multiplex sequencing of miRNA. Genome Research 21, 1506 1511. https://doi.org/10.1101/gr.121715.111

Andreou, A.I., Nakayama, N., 2018. Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. PLoS ONE 13, e0189892. https://doi.org/10.1371/journal.pone.0189892

Andrilenas, K.K., Penvose, A., Siggers, T., 2015. Using protein-binding microarrays to study transcription factor specificity: homologs, isoforms and complexes. Briefings in Functional Genomics 14, 17–29. https://doi.org/10.1093/bfgp/elu046

Andrilenas, K.K., Ramlall, V., Kurland, J., Leung, B., Harbaugh, A.G., Siggers, T., 2018. DNA-binding landscape of IRF3, IRF5 and IRF7 dimers: implications for dimer-specific gene regulation. Nucleic Acids Research 6, 644. https://doi.org/10.1093/nar/gky002

Antonczyk, A., Krist, B., Sajek, M., Michalska, A., Piaszyk-Borychowska, A., Plens-Galaska, M., Wesoly, J., Bluyssen, H.A., 2019. Direct Inhibition of IRF-Dependent Transcriptional Regulatory Mechanisms Associated With Disease. Frontiers in immunology 10, 3397. https://doi.org/10.3389/fimmu.2019.01176

Auwerx, J., 1991. The human leukemia cell line, THP-1: A multifacetted model for the study of monocyte-macrophage differentiation. Experientia 47, 22 31. https://doi.org/10.1007/bf02041244

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.-F., Coburn, D., Newburger, D.E., Morris, Q., Hughes, T.R., Bulyk, M.L., 2009. Diversity and complexity in DNA recognition by transcription factors. Science 324, 1720 1723. https://doi.org/10.1126/science.1162327

Badis, G., Chan, E.T., van Bakel, H., Peña-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., Gebbia, M., Talukder, S., Yang, A., Mnaimneh, S., Terterov, D., Coburn, D., Yeo, A., Yeo, Z., Clarke, N.D., Lieb, J.D., Ansari, A.Z., Nislow, C., Hughes, T.R., 2008. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Molecular cell 32, 878 887. https://doi.org/10.1016/j.molcel.2008.11.020

Bageritz, J., Raddi, G., 2019. Single-Cell RNA Sequencing with Drop-Seq. Methods in molecular biology (Clifton, N.J.) 1979, 73 85. https://doi.org/10.1007/978-1-4939-9240-9_6

Barnes, B., Field, A., Pitha-Rowe, P.M., 2003. Virus-induced Heterodimer Formation between IRF-5 and IRF-7 Modulates Assembly of the IFNA Enhanceosome in Vivo and Transcriptional Activity of IFNA Genes. Journal of Biological Chemistry 278, 16630 16641. https://doi.org/10.1074/jbc.m212609200

Barnes, B., Kellum, M., Field, A., Pitha-Rowe, P.M., 2002a. Multiple Regulatory Domains of IRF-5 Control Activation, Cellular Localization, and Induction of Chemokines That Mediate Recruitment of T Lymphocytes. Molecular and Cellular Biology 22, 5721 5740. https://doi.org/10.1128/mcb.22.16.5721-5740.2002

Barnes, B., Lubyova, B., Pitha, P.M., 2002b. Review: On the Role of IRF in Host Defense. Journal of Interferon & Cytokine Research 22, 59–71. https://doi.org/10.1089/107999002753452665

Barnes, B., Moore, P., Pitha-Rowe, P.M., 2001. Virus-specific activation of a novel interferon regulatory factor, IRF-5, results in the induction of distinct interferon alpha genes. Journal of Biological Chemistry 276, 23382 23390. https://doi.org/10.1074/jbc.m101216200

Barnes, B., Richards, J., Mancl, M.E., Hanash, S., Beretta, L., Pitha-Rowe, P.M., 2004. Global and distinct targets of IRF-5 and IRF-7 during innate response to viral infection. Journal of Biological Chemistry 279, 45194 45207. https://doi.org/10.1074/jbc.m400726200

Barnes, B.J., Field, A.E., Pitha-Rowe, P.M., 2003. Virus-induced Heterodimer Formation between IRF-5 and IRF-7 Modulates Assembly of theIFNA Enhanceosome in Vivo and Transcriptional Activity of IFNA Genes. Journal of Biological Chemistry 278, 16630–16641. https://doi.org/10.1074/jbc.m212609200

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S.A., Morris, Q.D., Bulyk, M.L., Hughes, T.R., 2008. Variation in Homeodomain DNA Binding Revealed by High-Resolution

Analysis of Sequence Preferences. Cell 133, 1266 1276. https://doi.org/10.1016/j.cell.2008.05.024

Berger, M.F., Bulyk, M.L., 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nature Protocols 4, 393 411. https://doi.org/10.1038/nprot.2008.195

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., Bulyk, M.L., 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nature Biotechnology 24, 1429 1435. https://doi.org/10.1038/nbt1246

Bernard, P., Gabarit, P., Bahassi, E., Couturier, M., 1994. Positive-selection vectors using the F plasmid ccdB killer gene. Gene 148, 71 74. https://doi.org/10.1016/0378-1119(94)90235-6

Best, K., Oakes, T., Heather, J.M., Shawe-Taylor, J., Chain, B., 2015. Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. Scientific Reports 5, 14629. https://doi.org/10.1038/srep14629

Bolotin, E., Liao, H., Ta, T., Yang, C., Hwang-Verslues, W., Evans, J.R., Jiang, T., adek, F., 2010. Integrated approach for the identification of human hepatocyte nuclear factor 4alpha target genes using protein binding microarrays. Hepatology 51, 642 653. https://doi.org/10.1002/hep.23357

Bonetta, L., 2010. Whole-Genome Sequencing Breaks the Cost Barrier. Cell 141, 917 919. https://doi.org/10.1016/j.cell.2010.05.034

Brown, C.D., Mangravite, L.M., Engelhardt, B.E., 2013. Integrative Modeling of eQTLs and Cis-Regulatory Elements Suggests Mechanisms Underlying Cell Type Specificity of eQTLs. PLoS genetics 9, e1003649. https://doi.org/10.1371/journal.pgen.1003649

Brownell, J., Bruckner, J., Wagoner, J., Thomas, E., Loo, Y.-M., Gale, M., Liang, J.T., Polyak, S.J., 2014. Direct, interferon-independent activation of the CXCL10 promoter by NF-κB and interferon regulatory factor 3 during hepatitis C virus infection. Journal of Virology 88, 1582 1590. https://doi.org/10.1128/jvi.02007-13

Bulyk, M.L., Gentalen, E., Lockhart, D., Church, G., 1999. Quantifying DNA-protein interactions by double-stranded DNA arrays. Nature Biotechnology 17, 573 577. https://doi.org/10.1038/9878

Bushnell, B., Rood, J., Singer, E., 2017. BBMerge - Accurate paired shotgun read merging via overlap. PLoS ONE 12, e0185056. https://doi.org/10.1371/journal.pone.0185056

Caillaud, A., Hovanessian, A., Levy, D., Marie, I., 2005. Regulatory Serine Residues Mediate Phosphorylation-dependent and Phosphorylation-independent Activation of Interferon Regulatory Factor 7. Journal of Biological Chemistry 280, 17671 17677. https://doi.org/10.1074/jbc.m411389200

Caillaud, A., Prakash, A., Smith, E., Masumi, A., Hovanessian, A.G., Levy, D.E., Marié, I., 2002. Acetylation of Interferon Regulatory Factor-7 by p300/CREB-binding Protein (CBP)-associated Factor (PCAF) Impairs its DNA Binding. Journal of Biological Chemistry 277, 49417–49421. https://doi.org/10.1074/jbc.m207484200

Chan, Y., Gack, M.U., 2016. Viral evasion of intracellular DNA and RNA sensing. Nature Reviews Microbiology 14, 360 373. https://doi.org/10.1038/nrmicro.2016.45

Chattopadhyay, S., Fensterl, V., Zhang, Y., Veleeparambil, M., Yamashita, M., Sen, G.C., 2013. Role of interferon regulatory factor 3-mediated apoptosis in the establishment and maintenance of persistent infection by Sendai virus. Journal of Virology 87, 16 24. https://doi.org/10.1128/jvi.01853-12

Chen, W., Lam, S.S., nath, H., Jiang, Z., Correia, J.J., Schiffer, C.A., Fitzgerald, K.A., Lin, K., Royer, W.E., 2008a. Insights into interferon regulatory factor activation from the crystal structure of dimeric IRF5. Nature Structural & Molecular Biology 15, 1213 1220. https://doi.org/10.1038/nsmb.1496

Chen, W., nath, H., Lam, S.S., Schiffer, C.A., Jr., W.E., Lin, K., 2008b. Contribution of Ser386 and Ser396 to Activation of Interferon Regulatory Factor 3. Journal of Molecular Biology 379, 251 260. https://doi.org/10.1016/j.jmb.2008.03.050

Chen, W., Royer, W.E., 2010. Structural insights into interferon regulatory factor activation. Cellular signaling 22, 883 887. https://doi.org/10.1016/j.cellsig.2009.12.005

Cheng, T.-F., Brzostek, S., Ando, O., Scoy, S., Kumar, P.K., Reich, N.C., 2006. Differential activation of IFN regulatory factor (IRF)-3 and IRF-5 transcription factors during viral infection. The Journal of Immunology 176, 7462 7470.

https://doi.org/10.4049/jimmunol.176.12.7462

Cheon, H., Holvey-Bates, E.G., Schoggins, J.W., Forster, S., Hertzog, P., Imanaka, N., Rice, C.M., Jackson, M.W., Junk, D.J., Stark, G.R., 2013. IFNβ-dependent increases in STAT1, STAT2, and IRF9 mediate resistance to viruses and DNA damage. The EMBO Journal 32, 2751 2763. https://doi.org/10.1038/emboj.2013.203

Chow, K.T., Wilkins, C., Narita, M., Green, R., Knoll, M., Loo, Y.-M., Gale, M., 2018. Differential and Overlapping Immune Programs Regulated by IRF3 and IRF5 in Plasmacytoid Dendritic Cells. The Journal of Immunology 201, 3036 3050. https://doi.org/10.4049/jimmunol.1800221

Civas, A., Génin, P., Morin, P., Lin, R., Hiscott, J., 2006. Promoter organization of the interferon-A genes differentially affects virus-induced expression and responsiveness to TBK1 and IKKepsilon. Journal of Biological Chemistry 281, 4856 4866. https://doi.org/10.1074/jbc.m506812200

Civas, A., Island, M.-L., Génin, P., Morin, P., Navarro, S., 2002. Regulation of virus-induced interferon-A genes. Biochimie 84, 643 654. https://doi.org/10.1016/s0300-9084(02)01431-1

Clement, J., Bibeau-Poirier, A., Gravel, S., Grandvaux, N., Bonneil, E., Thibault, P., Meloche, S., Servant, M., 2008. Phosphorylation of IRF-3 on Ser 339 Generates a Hyperactive Form of IRF-3 through Regulation of Dimerization and CBP Association. Journal of Virology 82, 3984 3996. https://doi.org/10.1128/jvi.02526-07

Cock, P., Antao, T., Chang, J., Chapman, Cox, C., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422 1423. https://doi.org/10.1093/bioinformatics/btp163

Creagh, E.M., O'Neill, L.A., 2006. TLRs, NLRs and RLRs: a trinity of pathogen sensors that co-operate in innate immunity. Trends in Immunology 27, 352 357. https://doi.org/10.1016/j.it.2006.06.003

Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R.S., Stern, D.L., 2015. Low affinity binding site clusters confer hox specificity and regulatory robustness. Cell 160, 191 203. https://doi.org/10.1016/j.cell.2014.11.041

Czerkies, M., Korwek, Z., Prus, W., Kochańczyk, M., Jaruszewicz-Błońska, J., Tudelska, K., Błoński, S., Kimmel, M., Brasier, A.R., Lipniacki, T., 2018. Cell fate in antiviral response arises in the crosstalk of IRF, NF-κB and JAK/STAT pathways. Nature Communications 9, 267. https://doi.org/10.1038/s41467-017-02640-8

Dao-Thi, M.-H., Melderen, L., Genst, E., Afif, H., Buts, L., Wyns, L., Loris, R., 2005. Molecular Basis of Gyrase Poisoning by the Addiction Toxin CcdB. Journal of Molecular Biology 348, 1091 1102. https://doi.org/10.1016/j.jmb.2005.03.049

De Ioannes, P., Escalante, C.R., Aggarwal, A.K., 2011. Structures of apo IRF-3 and IRF-7 DNA binding domains: effect of loop L1 on DNA binding. Nucleic Acids Research 39, 7300 7307. https://doi.org/10.1093/nar/gkr325

de Wet, J., Wood, K., Helinski, D., Luca, 1985. Cloning of firefly luciferase cDNA and the expression of active luciferase in Escherichia coli. Proceedings of the National Academy of Sciences 82, 7870 7873. https://doi.org/10.1073/pnas.82.23.7870

del Fresno, C., Soulat, D., Roth, S., Blazek, K., Udalova, I., Sancho, D., Ruland, J., Ardavín, C., 2013. Interferon-β production via Dectin-1-Syk-IRF5 signaling in dendritic cells is crucial for immunity to C. albicans. Immunity 38, 1176 1186. https://doi.org/10.1016/j.immuni.2013.05.010

Diamond, M.S., Farzan, M., 2012. The broad-spectrum antiviral functions of IFIT and IFITM proteins. Nature Reviews Immunology 13, 46 57. https://doi.org/10.1038/nri3344

Dragan, A.I., Hargreaves, V.V., Makeyeva, E.N., Privalov, P.L., 2007. Mechanisms of activation of interferon regulator factor 3: the role of C-terminal domain phosphorylation in IRF-3 dimerization and DNA binding. Nucleic Acids Research 35, 3525 3534. https://doi.org/10.1093/nar/gkm142

Eames, H., Saliba, D., Krausgruber, T., Lanfrancotti, A., Ryzhakov, G., Udalova, I., 2012. KAP1/TRIM28: An inhibitor of IRF5 function in inflammatory macrophages. Immunobiology 217, 1315 1324. https://doi.org/10.1016/j.imbio.2012.07.026

Engler, C., Marillonnet, S., 2013. Combinatorial DNA Assembly Using Golden Gate Cloning. Synthetic Biology 1073, 141 156. https://doi.org/10.1007/978-1-62703-625-2_12

Enoch, T., Zinn, K., Maniatis, T., 1986. Activation of the human beta-interferon gene requires an interferon-inducible factor. Molecular and Cellular Biology 6, 801 810. https://doi.org/10.1128/mcb.6.3.801

Escalante, C.R., Nistal-Villán, E., Shen, L., García-Sastre, A., Aggarwal, A.K., 2007. Structure of IRF-3 Bound to the PRDIII-I Regulatory Element of the Human Interferon-β Enhancer. Molecular cell 26, 703 716. https://doi.org/10.1016/j.molcel.2007.04.022

Fang, B., Mane-Padros, D., Bolotin, E., Jiang, T., adek, F., 2012. Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. Nucleic Acids Research 40, 5343 5356. https://doi.org/10.1093/nar/gks190

Feng, D., Sangster-Guity, N., Stone, R., Korczeniewska, J., Mancl, M.E., Fitzgerald-Bocarsly, P., Barnes, B., 2010. Differential requirement of histone acetylase and deacetylase activities for IRF5-mediated proinflammatory cytokine expression. Journal of immunology (Baltimore, Md. : 1950) 185, 6003 6012. https://doi.org/10.4049/jimmunol.1000482

Fensterl, V., Chattopadhyay, S., Sen, G.C., 2015. No Love Lost Between Viruses and Interferons. Annual Review of Virology 2, 549 572. https://doi.org/10.1146/annurev-virology-100114-055249

Field, S., Udalova, I., Ragoussis, J., 2006. Accuracy and Reproducibility of Protein–DNA Microarray Technology. Analytics of Protein–DNA Interactions 104, 87 110. https://doi.org/10.1007/10_2006_035

Fink, K., Grandvaux, N., 2013. STAT2 and IRF9: Beyond ISGF3. JAK-STAT 2, e27521. https://doi.org/10.4161/jkst.27521

Fiore, C., Cohen, B.A., 2016. Interactions between pluripotency factors specify cis-regulation in embryonic stem cells. Genome Research 26, 778 786. https://doi.org/10.1101/gr.200733.115

Foreman, H.-C., Scoy, S., Cheng, T.-F., Reich, N.C., 2012. Activation of Interferon Regulatory Factor 5 by Site Specific Phosphorylation. PLoS ONE 7, e33098. https://doi.org/10.1371/journal.pone.0033098

Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., Solano, R., 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. Proceedings of the National Academy of

Sciences of the United States of America 111, 2367 2372. https://doi.org/10.1073/pnas.1316278111

Freaney, J.E., Kim, R., Mandhana, R., Horvath, C.M., 2013. Extensive Cooperation of Immune Master Regulators IRF3 and NFκB in RNA Pol II Recruitment and Pause Release in Human Innate Antiviral Transcription. Cell Reports 4, 959 973. https://doi.org/10.1016/j.celrep.2013.07.043

Gad, H., Dellgren, C., Hamming, O.J., Vends, S., Paludan, S.R., Hartmann, R., 2009. Interferon-lambda is functionally an interferon but structurally related to the interleukin-10 family. The Journal of biological chemistry 284, 20869 20875. https://doi.org/10.1074/jbc.m109.002923

García-Nafría, J., Watson, J.F., Greger, I.H., 2016. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. Scientific Reports 6, 27459. https://doi.org/10.1038/srep27459

Geissmann, Q., 2013. OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects. PLoS ONE 8, e54072. https://doi.org/10.1371/journal.pone.0054072

Génin, P., Lin, R., Hiscott, J., Civas, A., 2009a. Differential Regulation of Human Interferon A Gene Expression by Interferon Regulatory Factors 3 and 7. Molecular and Cellular Biology 29, 3435 3450. https://doi.org/10.1128/mcb.01805-08

Génin, P., Vaccaro, A., Civas, A., 2009b. The role of differential expression of human interferon-A genes in antiviral immunity. Cytokine & growth factor reviews 20, 283 295. https://doi.org/10.1016/j.cytogfr.2009.07.005

Ghodke-Puranik, Y., Niewold, T.B., 2015. Immunogenetics of systemic lupus erythematosus: A comprehensive review. Journal of Autoimmunity 64, 125 136. https://doi.org/10.1016/j.jaut.2015.08.004

Gordân, R., Murphy, K.F., McCord, R.P., Zhu, C., Vedenko, A., Bulyk, M.L., 2011. Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. Genome Biology 12, R125. https://doi.org/10.1186/gb-2011-12-12-r125

Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., Bulyk, M.L., 2013. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. CellReports 3, 1093 1104.

https://doi.org/10.1016/j.celrep.2013.03.014

Grossman, S.R., Zhang, X., Wang, L., Engreitz, J., Melnikov, A., Rogov, P., Tewhey, R., Isakova, A., Deplancke, B., Bernstein, B.E., Mikkelsen, T.S., Lander, E.S., 2017. Systematic dissection of genomic features determining transcription factor binding and enhancer function. Proceedings of the National Academy of Sciences 114, E1291 E1300. https://doi.org/10.1073/pnas.1621150114

Grove, C.A., Masi, F., Barrasa, I.M., Newburger, D.E., Alkema, M.J., Bulyk, M.L., Walhout, A.J., 2009. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. Cell 138, 314 327. https://doi.org/10.1016/j.cell.2009.04.058

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., Dor, Y., Regev, A., Yanai, I., 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biology 17, 892. https://doi.org/10.1186/s13059-016-0938-8

Hellman, L.M., Fried, M.G., 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nature Protocols 2, 1849 1861. https://doi.org/10.1038/nprot.2007.249

Hillyer, P., Mane, V.P., hramm, L., Puig, M., Verthelyi, D., Chen, A., Zhao, Z., Navarro, M.B., Kirschman, K.D., kant Bykadi, Jubin, R.G., Rabin, R.L., 2012. Expression profiles of human interferon-alpha and interferon-lambda subtypes are ligand- and cell-dependent. Immunology and Cell Biology 90, 774 783. https://doi.org/10.1038/icb.2011.109

Hiscott, J., 2007. Convergence of the NF-κB and IRF pathways in the regulation of the innate antiviral response. Cytokine & growth factor reviews 18, 483 490. https://doi.org/10.1016/j.cytogfr.2007.06.002

Honda, K., Takaoka, A., Taniguchi, T., 2006. Type I Inteferon Gene Induction by the Interferon Regulatory Factor Family of Transcription Factors. Immunity 25, 349–360. https://doi.org/10.1016/j.immuni.2006.08.009

Honda, K., Taniguchi, T., 2006. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. Nature Reviews Immunology 6, 644 658. https://doi.org/10.1038/nri1900

Honda, K., Yanai, H., Negishi, H., Asagiri, M., Sato, M., Mizutani, T., Shimada, N., Ohba, Y., Takaoka, A., Yoshida, N., Taniguchi, T., 2005. IRF-7 is the master regulator of type-I interferon-dependent immune responses. Nature 434, 772 777. https://doi.org/10.1038/nature03464

Hong, J., Gresham, D., 2017. Incorporation of unique molecular identifiers in TruSeq adapters improves the accuracy of quantitative sequencing. BioTechniques 63. https://doi.org/10.2144/000114608

Horspool, D.R., Coope, R.J., Holt, R.A., 2010. Efficient assembly of very short oligonucleotides using T4 DNA Ligase. BMC Research Notes 3, 291. https://doi.org/10.1186/1756-0500-3-291

Hu, L.-L., Zhang, S.-S., Li, X.-X., Wang, B.-L., 2010. The use of the ccdB lethal gene for constructing a zero background vector in order to clone blunt-end PCR products. Molekuliarnaia biologiia 44, 174 176.

Hughes, A.E., Myers, C.A., Corbo, J.C., 2018. A massively parallel reporter assay reveals context-dependent activity of homeodomain binding sites in vivo. Genome Research 28, 1520 1531. https://doi.org/10.1101/gr.231886.117

Ingraham, C.R., Kinoshita, A., Kondo, S., Yang, B., Sajan, S., Trout, K.J., Malik, M.I., Dunnwald, M., Goudy, S.L., Lovett, M., Murray, J.C., Schutte, B.C., 2006. Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). Nature Genetics 38, 1335 1340. https://doi.org/10.1038/ng1903

Islam, S., Zeisel, A., Joost, S., Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., Linnarsson, S., 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. Nature Methods 11, 163 166. https://doi.org/10.1038/nmeth.2772

Iversen, M.B., Paludan, S.R., 2010. Mechanisms of Type III Interferon Expression. Journal of interferon & cytokine research 30, 573 578. https://doi.org/10.1089/jir.2010.0063

Iverson, S.V., Haddock, T.L., Beal, J., nsmore, D., 2016. CIDAR MoClo: Improved MoClo Assembly Standard and New E. coliPart Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. ACS Synthetic Biology 5, 99 103. https://doi.org/10.1021/acssynbio.5b00124

Jaitin, D., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., Amit, I., 2014. Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. Science 343, 776 779. https://doi.org/10.1126/science.1247651

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J.M., Vincentelli, R., Luscombe, N.M., Hughes, T.R., Lemaire, P., Ukkonen, E., Kivioja, T., Taipale, J., 2013. DNA-Binding Specificities of Human Transcription Factors. Cell 152, 327 339. https://doi.org/10.1016/j.cell.2012.12.009

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E., Hong, M.-Y., Karczewski, K.J., Huber, W., Weissman, S.M., Gerstein, M.B., Korbel, J.O., Snyder, M., 2010. Variation in transcription factor binding among humans. Science 328, 232 235. https://doi.org/10.1126/science.1183621

Kawai, T., Akira, S., 2011. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. Immunity 34, 637 650. https://doi.org/10.1016/j.immuni.2011.05.006

Kawai, T., Akira, S., 2006. Innate immune recognition of viral infection. Nature Immunology 7, 131 137. https://doi.org/10.1038/ni1303

Kebschull, J.M., Zador, A.M., 2015. Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Research 43, e143 e143. https://doi.org/10.1093/nar/gkv717

Kimura, H., Yoshizumi, M., Ishii, H., Oishi, K., Ryo, A., 2013. Cytokine production and signaling pathways in respiratory virus infection. Frontiers in Microbiology 4. https://doi.org/10.3389/fmicb.2013.00276

King, D.M., Maricque, B.B., Cohen, B.A., 2018. Synthetic and genomic regulatory elements reveal aspects of cis regulatory grammar in Mouse Embryonic Stem Cells. bioRxiv 398107. https://doi.org/10.1101/398107

Kinney, J.B., McCandlish, D.M., 2019. Massively Parallel Assays and Quantitative Sequence–Function Relationships. Annual review of genomics and human genetics 20, annurev-genom-083118-014845. https://doi.org/10.1146/annurev-genom-083118-014845

Koshiba, R., Yanai, H., Matsuda, A., Goto, A., Nakajima, A., Negishi, H., Nishio, J., Smale, S.T., Taniguchi, T., 2013a. Regulation of cooperative function of the Il12b enhancer and promoter by the interferon regulatory factors 3 and 5. Biochemical and Biophysical Research Communications 430, 95 100. https://doi.org/10.1016/j.bbrc.2012.11.006

Koshiba, R., Yanai, H., Matsuda, A., Goto, A., Nakajima, A., Negishi, H., Nishio, J., Smale, S.T., Taniguchi, T., 2013b. Regulation of cooperative function of the Il12b enhancer and promoter by the interferon regulatory factors 3 and 5. Biochemical and Biophysical Research Communications 430, 95 100. https://doi.org/10.1016/j.bbrc.2012.11.006

Krause, C.D., Pestka, S., 2015. Cut, copy, move, delete: The study of human interferon genes reveal multiple mechanisms underlying their evolution in amniotes. - PubMed - NCBI. Cytokine 76, 480 495. https://doi.org/10.1016/j.cyto.2015.07.019

Krausgruber, T., Blazek, K., Smallie, T., Alzabin, S., 2011. IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. Nature 12. https://doi.org/10.1038/ni.1990

Krausgruber, T., Saliba, D., Ryzhakov, G., Lanfrancotti, A., Blazek, K., Udalova, I., 2010. IRF5 is required for late-phase TNF secretion by human dendritic cells. Blood 115, 4421 4430. https://doi.org/10.1182/blood-2010-01-263020

Krueger, F., Andrews, S.R., Osborne, C.S., 2011. Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by deferred cluster calling. PLoS ONE 6, e16607. https://doi.org/10.1371/journal.pone.0016607

Ku, C., Yang, K., Bustamante, J., Puel, A., Bernuth, H., Santos, O., Lawrence, T., Chang, H., Mousa, H., Picard, C., Casanova, J.-L., 2005. Inherited disorders of human Toll-like receptor signaling: immunological implications. Immunological reviews 203, 10 20. https://doi.org/10.1111/j.0105-2896.2005.00235.x

Kumar, H., Kawai, T., Akira, S., 2011. Pathogen Recognition by the Innate Immune System. International Reviews of Immunology 30, 16 34. https://doi.org/10.3109/08830185.2010.529976

Kwasnieski, J., Mogno, I., Myers, C., Corbo, J., Cohen, B.A., 2012. Complex effects of nucleotide variants in a mammalian cis-regulatory element. Proceedings of the National Academy of Sciences 109, 19498 19503. https://doi.org/10.1073/pnas.1210678109

Lazear, H.M., Lancaster, A., Wilkins, C., Suthar, M.S., Huang, A., Vick, S.C., Clepper, L., Thackray, L., Brassil, M.M., Virgin, H.W., Nikolich-Zugich, J., Moses, A.V., Gale, M., Früh, K., Diamond, M.S., 2013. IRF-3, IRF-5, and IRF-7 Coordinately Regulate the Type I IFN Response in Myeloid Dendritic Cells

Downstream of MAVS Signaling. PLoS Pathogens 9, e1003118. https://doi.org/10.1371/journal.ppat.1003118

Lazear, H.M., Nice, T.J., Diamond, M.S., 2015. Interferon-λ: Immune Functions at Barrier Surfaces and Beyond. Immunity 43, 15 28. https://doi.org/10.1016/j.immuni.2015.07.001

Lee, M., Kim, Y.-J., 2007. Signaling pathways downstream of pattern-recognition receptors and their cross talk. Annual Review of Biochemistry 76, 447 480. https://doi.org/10.1146/annurev.biochem.76.060605.122847

LeProust, E.M., Peck, B.J., Spirin, K., McCuen, H., Moore, B., Namsaraev, E., Caruthers, M.H., 2010. Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. Nucleic Acids Research 38, 2522 2540. https://doi.org/10.1093/nar/gkq163

Li, D., De, S., Li, D., Song, S., Matta, B., Barnes, B.J., 2016. Specific detection of interferon regulatory factor 5 (IRF5): A case of antibody inequality. Scientific Reports 6, 31002. https://doi.org/10.1038/srep31002

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589 595. https://doi.org/10.1093/bioinformatics/btp698

Lin, R., Mamane, Y., Hiscott, J., 1999. Structural and functional analysis of interferon regulatory factor 3: localization of the transactivation and autoinhibitory domains. Molecular and Cellular Biology 19, 2465 2474. https://doi.org/10.1128/mcb.19.4.2465

Linnell, J., Mott, R., Field, S., Kwiatkowski, D.P., Ragoussis, J., Udalova, I., 2004. Quantitative high-throughput analysis of transcription factor binding specificities. Nucleic Acids Research 32, e44. https://doi.org/10.1093/nar/gnh042

Lis, M., Walther, D., 2016. The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? BMC Genomics 17, 185. https://doi.org/10.1186/s12864-016-2549-x

Liu, Y., Zhang, Y.-B., Liu, T.-K., Gui, J.-F., 2013. Lineage-Specific Expansion of IFIT Gene Family: An Insight into Coevolution with IFN Gene Family. PLoS ONE 8, e66859. https://doi.org/10.1371/journal.pone.0066859

Loo, Y.-M., Jr, M., 2011. Immune Signaling by RIG-I-like Receptors. Immunity 34, 680 692. https://doi.org/10.1016/j.immuni.2011.05.003

Ma, B., Tsai, C.-J., Haliloğlu, T., Nussinov, R., 2011. Dynamic Allostery: Linkers Are Not Merely Flexible. Structure (London, England : 1993) 19, 907 917. https://doi.org/10.1016/j.str.2011.06.002

MacConaill, L.E., Burns, R.T., Nag, A., Coleman, H.A., Slevin, M.K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M.S., Ducar, M.D., Meyerson, M., Thorner, A.R., 2018. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics 19, 30. https://doi.org/10.1186/s12864-017-4428-5

Majewski, J., Pastinen, T., 2011. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends in Genetics 27, 72 79. https://doi.org/10.1016/j.tig.2010.10.006

Mangalam, H., 2002. The Bio* toolkits -- a brief overview. Briefings in bioinformatics 3, 296 302. https://doi.org/10.1093/bib/3.3.296

Marié, I., Smith, E., Prakash, A., Levy, D., 2000. Phosphorylation-induced dimerization of interferon regulatory factor 7 unmasks DNA binding and a bipartite transactivation domain. Molecular and Cellular Biology 20, 8803 8814. https://doi.org/10.1128/mcb.20.23.8803-8814.2000

Marx, V., 2017. How to deduplicate PCR. Nature Methods 14, 473 476. https://doi.org/10.1038/nmeth.4268

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., Zhang, A.W., Parcy, F., Lenhard, B., Sandelin, A., Wasserman, W.W., 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Research 44, D110 D115. https://doi.org/10.1093/nar/gkv1176

Matta, B., Barnes, B., 2019. Coordination between innate immune cells, type I IFNs and IRF5 drives SLE pathogenesis. Cytokine 154731. https://doi.org/10.1016/j.cyto.2019.05.018

Mazumder, A., Batabyal, S., Mondal, M., Mondol, T., Choudhury, S., Ghosh, R., Chatterjee, T., Bhattacharyya, D., Pal, S., Roy, S., 2017. Specific DNA sequences allosterically enhance protein–protein interaction in a transcription factor through modulation of protein dynamics: implications for specificity of gene regulation. Physical Chemistry Chemical Physics 19, 14781 14792. https://doi.org/10.1039/c7cp01193h

McNab, F., Mayer-Barber, K., Sher, A., Wack, A., O'Garra, A., 2015. Type I interferons in infectious disease. Nature Reviews Immunology 15, 87 103. https://doi.org/10.1038/nri3787

Medzhitov, R., 2001. Toll-like receptors and innate immunity. Nature Reviews Immunology 1, 135 145. https://doi.org/10.1038/35100529

Medzhitov, R., Horng, T., 2009. Transcriptional control of the inflammatory response. Nature Reviews Immunology 9, 692 703. https://doi.org/10.1038/nri2634

Meijsing, S., Pufall, M., So, A., Bates, D., Chen, L., Yamamoto, K., 2009. DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity. Science 324, 407 410. https://doi.org/10.1126/science.1164265

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Kinney, J.B., Kellis, M., Lander, E.S., Mikkelsen, T.S., 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nature Biotechnology 30, 271 +. https://doi.org/10.1038/nbt.2137

Miyamoto, M., Fujita, T., Kimura, Y., Maruyama, M., Harada, H., Sudo, Y., Miyata, T., Taniguchi, T., 1988. Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to IFN-β gene regulatory elements. Cell 54, 903 913. https://doi.org/10.1016/s0092-8674(88)91307-4

Mohaghegh, N., Bray, D., Keenan, J., Penvose, A., Andrilenas, K.K., Ramlall, V., Siggers, T., 2019. NextPBM: a platform to study cell-specific transcription factor binding and cooperativity. Nucleic Acids Research 47, e31 e31. https://doi.org/10.1093/nar/gkz020

Mori, M., Yoneyama, M., Ito, T., Takahashi, K., Inagaki, F., Fujita, T., 2004. Identification of Ser-386 of Interferon Regulatory Factor 3 as Critical Target for Inducible Phosphorylation That Determines Activation. Journal of Biological Chemistry 279, 9698 9702. https://doi.org/10.1074/jbc.m310616200

Mukherjee, S., Berger, M.F., Jona, G., Wang, X.S., Muzzey, D., Snyder, M., Young, R.A., Bulyk, M.L., 2004. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nature Genetics 36, 1331 1339.

https://doi.org/10.1038/ng1473

Mulero, M., Huang, D.-B., Nguyen, T.H., Wang, V., Li, Y., Biswas, T., Ghosh, G., 2017. DNA-binding affinity and transcriptional activity of the RelA homodimer of nuclear factor kappa B are not correlated. Journal of Biological Chemistry 292, jbc.M117.813980 18830. https://doi.org/10.1074/jbc.m117.813980

Negishi, H., Miki, S., Sarashina, H., Taguchi-Atarashi, N., Nakajima, A., Matsuki, K., Endo, N., Yanai, H., Nishio, J., Honda, K., Taniguchi, T., 2012. Essential contribution of IRF3 to intestinal homeostasis and microbiota-mediated Tslp gene induction. Proceedings of the National Academy of Sciences of the United States of America 109, 21016 21021. https://doi.org/10.1073/pnas.1219482110

Nishigaki, K., Kaneko, Y., Wakuda, H., Husimi, Y., Tanaka, T., 1985. Type II restriction endonucleases cleave single-stranded DNAs in general. Nucleic Acids Research 13, 5747 5760. https://doi.org/10.1093/nar/13.16.5747

Nogales, E., Scheres, S.H., 2015. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. Molecular cell 58, 677 689. https://doi.org/10.1016/j.molcel.2015.02.019

Ogawa, T., Kryukov, K., Imanishi, T., Shiroguchi, K., 2017. The efficacy and further functional advantages of random-base molecular barcodes for absolute and digital quantification of nucleic acid molecules. Scientific Reports 7, 13576. https://doi.org/10.1038/s41598-017-13529-3

Oliphant, A., Nussbaum, A., Struhl, K., 1986. Cloning of random-sequence oligodeoxynucleotides. Gene 44, 177 183. https://doi.org/10.1016/0378-1119(86)90180-0

Ortigão, F.J., Rösch, H., Selter, H., Fröhlich, A., Lorenz, A., Montenarh, M., Seliger, H., 1992. Antisense Effect of Oligodeoxynucleotides with Inverted Terminal Internucleotidic Linkages: A Minimal Modification Protecting against Nucleolytic Degradation. Antisense Research and Development 2, 129 146. https://doi.org/10.1089/ard.1992.2.129

Ott, J., Eckstein, F., 1987. Protection of oligonucleotide primers against degradation by DNA polymerase I. Biochemistry 26, 8237 8241. https://doi.org/10.1021/bi00399a032

Panne, D., Maniatis, T., Harrison, S.C., 2007. An atomic model of the interferon-beta enhanceosome. Cell 129, 1111 1123.

https://doi.org/10.1016/j.cell.2007.05.019

Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D., Shendure, J., 2009. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nature Biotechnology 27, 1173 1175. https://doi.org/10.1038/nbt.1589

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., Liuni, S., 2001. Structural and functional features of eukaryotic mRNA untranslated regions. Gene 276, 73 81. https://doi.org/10.1016/s0378-1119(01)00674-6

Pflug, F.G., von Haeseler, A., 2018. TRUmiCount: correctly counting absolute numbers of molecules using unique molecular identifiers. Bioinformatics 34, 3137 3144. https://doi.org/10.1093/bioinformatics/bty283

Pingoud, A., Fuxreiter, M., Pingoud, V., Wende, W., 2005. Type II restriction endonucleases: structure and mechanism. Cellular and molecular life sciences : CMLS 62, 685 707. https://doi.org/10.1007/s00018-004-4513-1

Pingoud, A., Wilson, G.G., Wende, W., 2014. Type II restriction endonucleases— a historical perspective and more. Nucleic Acids Research 42, 7489 7527. https://doi.org/10.1093/nar/gku447

Pope, S.D., Medzhitov, R., 2018. Emerging Principles of Gene Expression Programs and Their Regulation. Molecular cell 71, 389 397. https://doi.org/10.1016/j.molcel.2018.07.017

Prakash, A., Levy, D.E., 2006. Regulation of IRF7 through cell type-specific protein stability. Biochemical and Biophysical Research Communications 342, 50 56. https://doi.org/10.1016/j.bbrc.2006.01.122

Qin, B.Y., Liu, C., Lam, S.S., nath, H., Delston, R., Correia, J.J., Derynck, R., Lin, K., 2003. Crystal structure of IRF-3 reveals mechanism of autoinhibition and virus-induced phosphoactivation. Nature Structural Biology 10, 913 921. https://doi.org/10.1038/nsb1002

Qin, B.Y., Liu, C., nath, H., Lam, S.S., Correia, J.J., Derynck, R., Lin, K., 2005. Crystal Structure of IRF-3 in Complex with CBP. Structure 13, 1269–1277. https://doi.org/10.1016/j.str.2005.06.011

Rabani, M., Pieper, L., Chew, G.-L., Schier, A.F., 2017. A Massively Parallel Reporter Assay of 3′ UTR Sequences Identifies In Vivo Rules for mRNA

Degradation. Molecular cell 68, 1083 1094.e5. https://doi.org/10.1016/j.molcel.2017.11.014

Redmond, A.K., Zou, J., Secombes, C.J., Macqueen, D.J., Dooley, H., 2019. Discovery of All Three Types in Cartilaginous Fishes Enables Phylogenetic Resolution of the Origins and Evolution of Interferons. Frontiers in immunology 10, 1918. https://doi.org/10.3389/fimmu.2019.01558

Ren, J., Chen, X., Chen, Z.J., 2014. IKKβ is an IRF5 kinase that instigates inflammation. Proceedings of the National Academy of Sciences of the United States of America 111, 17438 17443. https://doi.org/10.1073/pnas.1418516111

Roberts, R., 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Research 31, 1805 1812. https://doi.org/10.1093/nar/gkg274

Ryzhakov, G., Eames, H.L., Udalova, I., 2015. Activation and Function of Interferon Regulatory Factor 5. Journal of interferon & cytokine research 35, 71 78. https://doi.org/10.1089/jir.2014.0023

Saliba, D.G., Heger, A., Eames, H.L., Oikonomopoulos, S., Teixeira, A., Blazek, K., Androulidaki, A., Wong, D., Goh, F.G., Weiss, M., Byrne, A., Pasparakis, M., Ragoussis, J., Udalova, I., 2014a. IRF5:RelA Interaction Targets Inflammatory Genes in Macrophages. CellReports 8, 1308 1317. https://doi.org/10.1016/j.celrep.2014.07.034

Sboner, A., Mu, X., Greenbaum, D., Auerbach, R.K., Gerstein, M.B., 2011. The real cost of sequencing: higher than you think! Genome Biology 12, 125. https://doi.org/10.1186/gb-2011-12-8-125

Schnoor, M., Buers, I., Sietmann, A., Brodde, M.F., Hofnagel, O., Robenek, H., Lorkowski, S., 2009. Efficient non-viral transfection of THP-1 cells. Journal of Immunological Methods 344, 109 115. https://doi.org/10.1016/j.jim.2009.03.014

Shagin, D.A., Turchaninova, M.A., Shagina, I.A., Shugay, M., Zaretsky, A.R., Zueva, O.I., Bolotin, D.A., Lukyanov, S., Chudakov, D.M., 2017. Application of nonsense-mediated primer exclusion (NOPE) for preparation of unique molecular barcoded libraries. BMC Genomics 18, e81. https://doi.org/10.1186/s12864-017-3815-2

Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A.V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., Chakraborty, A.K., 2019.

Enhancer Features that Drive Formation of Transcriptional Condensates. Molecular cell 75, 549 561.e7. https://doi.org/10.1016/j.molcel.2019.07.009

Shukla, H., Vaitiekunas, P., Majumdar, A.K., Dragan, A.I., Dimitriadis, E.K., Kotova, S., Crane-Robinson, C., Privalov, P.L., 2012. The Linker of the Interferon Response Factor 3 Transcription Factor Is Not Unfolded. Biochemistry 51, 6320 6327. https://doi.org/10.1021/bi300260s

Siggers, T., Chang, A.B., Teixeira, A., Wong, D., Williams, K.J., Ahmed, B., Ragoussis, J., Udalova, I., Smale, S.T., Bulyk, M.L., 2012a. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. Nature Immunology 13, 95 102. https://doi.org/10.1038/ni.2151

Siggers, T., Chang, A.B., Teixeira, A., Wong, D., Williams, K.J., Ahmed, B., Ragoussis, J., Udalova, I., Smale, S.T., Bulyk, M.L., 2012b. Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. Nature Immunology 13, 95 102. https://doi.org/10.1038/ni.2151

Siggers, T., Duyzend, M.H., Reddy, J., Khan, S., Bulyk, M.L., 2011. Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. Molecular Systems Biology 7, 555 555. https://doi.org/10.1038/msb.2011.89

Smith, E.N., Jepsen, K., Khosroheidari, M., Rassenti, L.Z., D’Antonio, M., Ghia, E.M., Carson, D.A., Jamieson, C.H., Kipps, T.J., Frazer, K.A., 2014. Biased estimates of clonal evolution and subclonal heterogeneity can arise from PCR duplicates in deep sequencing experiments. Genome Biology 15, 420. https://doi.org/10.1186/s13059-014-0420-4

Smith, T., Heger, A., Sudbery, I., 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Research 27, 491 499. https://doi.org/10.1101/gr.209601.116

Stetson, D.B., Medzhitov, R., 2006a. Type I Interferons in Host Defense. Immunity 25, 373 381. https://doi.org/10.1016/j.immuni.2006.08.007

Stetson, D.B., Medzhitov, R., 2006b. Recognition of Cytosolic DNA Activates an IRF3-Dependent Innate Immune Response. Immunity 24, 93 103. https://doi.org/10.1016/j.immuni.2005.12.003

Suhara, W., Yoneyama, M., Iwamura, T., Yoshimura, S., Tamura, K., Namiki, H., Aimoto, S., Fujita, T., 2000. Analyses of virus-induced homomeric and heteromeric protein associations between IRF-3 and coactivator CBP/p300. Journal of Biochemistry 128, 301 307.

Tailor, P., Tamura, T., Ozato, K., 2006. IRF family proteins and type I interferon induction in dendritic cells. Cell Research 16, 7310018. https://doi.org/10.1038/sj.cr.7310018

Takahasi, K., Horiuchi, M., Fujii, K., Nakamura, S., Noda, N.N., Yoneyama, M., Fujita, T., Inagaki, F., 2010. Ser386 phosphorylation of transcription factor IRF-3 induces dimerization and association with CBP/p300 without overall conformational change. Genes to Cells 15, 901 910. https://doi.org/10.1111/j.1365-2443.2010.01427.x

Takaoka, A., Yanai, H., Kondo, S., Duncan, G., Negishi, H., Mizutani, T., Kano, S., Honda, K., Ohba, Y., Taniguchi, T., 2005. Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. Nature 434, 243 249. https://doi.org/10.1038/nature03308

Takeuchi, O., Akira, S., 2010. Pattern recognition receptors and inflammation. Cell 140, 805 820. https://doi.org/10.1016/j.cell.2010.01.022

Tamura, T., Yanai, H., Savitsky, D., Taniguchi, T., 2008. The IRF Family Transcription Factors in Immunity and Oncogenesis. Annual Review of Immunology 26, 535 584. https://doi.org/10.1146/annurev.immunol.26.021607.090400

Taniguchi, T., Ogasawara, K., Takaoka, A., Tanaka, N., 2001. IRF family of transcription factors as regulators of host defense. Annual Review of Immunology 19, 623 655. https://doi.org/10.1146/annurev.immunol.19.1.623

Taniguchi, T., Takaoka, A., 2001. A weak signal for strong responses: interferon-alpha/beta revisited. Nature Reviews Molecular Cell Biology 2, 378 386. https://doi.org/10.1038/35073080

Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander, E.S., Schaffner, S.F., Sabeti, P.C., 2016. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell 165, 1519 1529. https://doi.org/10.1016/j.cell.2016.04.027

Thielecke, L., Aranyossy, T., Dahl, A., Tiwari, R., Roeder, I., Geiger, H., Fehse, B., Glauche, I., Cornils, K., 2017. Limitations and challenges of genetic barcode quantification. Scientific Reports 7, 43249. https://doi.org/10.1038/srep43249

Udalova, I., Mott, R., Field, D., Kwiatkowski, D., 2002. Quantitative prediction of NF-kappa B DNA-protein interactions. Proceedings of the National Academy of Sciences 99, 8167 8172. https://doi.org/10.1073/pnas.102674699

van Arensbergen, J., Pagie, L., FitzPatrick, V.D., de Haas, M., Baltissen, M.P., Comoglio, F., van der Weide, R.H., Teunissen, H., Võsa, U., Franke, L., de Wit, E., Vermeulen, M., Bussemaker, H.J., van Steensel, B., 2019. High-throughput identification of human SNPs affecting regulatory element activity. Nature Genetics 51, 1160 1169. https://doi.org/10.1038/s41588-019-0455-2

van de Vosse, E., van Dissel, J.T., Ottenhoff, T.H., 2009. Genetic deficiencies of innate immune signalling in human infectious disease. The Lancet Infectious Diseases 9, 688 698. https://doi.org/10.1016/s1473-3099(09)70255-5

van Dijk, E.L., Jaszczyszyn, Y., Thermes, C., 2014. Library preparation methods for next-generation sequencing: Tone down the bias. Experimental Cell Research 322, 12 20. https://doi.org/10.1016/j.yexcr.2014.01.008

Vogel, S.N., Fitzgerald, K.A., Fenton, M.J., 2003. TLRs: differential adapter utilization by toll-like receptors mediates TLR-specific patterns of gene expression. Molecular interventions 3, 466 477. https://doi.org/10.1124/mi.3.8.466

Wang, V., Huang, W., Asagiri, M., Spann, N., Hoffmann, A., Glass, C., Ghosh, G., 2012. The Transcriptional Specificity of NF-κB Dimers Is Coded within the κB DNA Response Elements. CellReports 2, 824 839. https://doi.org/10.1016/j.celrep.2012.08.042

Wathelet, M.G., Lin, C.H., Parekh, B.S., Ronco, L.V., Howley, P.M., Maniatis, T., 1998. Virus Infection Induces the Assembly of Coordinately Activated Transcription Factors on the IFN-β Enhancer In Vivo. Molecular cell 1, 507 518. https://doi.org/10.1016/s1097-2765(00)80051-9

Watson, L.C., Kuchenbecker, K.M., Schiller, B.J., Gross, J.D., Pufall, M.A., Yamamoto, K.R., 2013. The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. Nature Structural & Molecular Biology 20, 876 883. https://doi.org/10.1038/nsmb.2595

Wei, G.-H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R., Yan, J., Talukder, S., Turunen, M., Taipale, M., Stunnenberg, H.G., Ukkonen, E., Hughes, T.R., Bulyk, M.L., Taipale, J., 2010. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. The EMBO Journal 29, 2147 2160. https://doi.org/10.1038/emboj.2010.106

Weiss, M., Byrne, A.J., Blazek, K., Saliba, D.G., Pease, J.E., Perocheau, D., Feldmann, M., Udalova, I., 2015. IRF5 controls both acute and chronic inflammation. Proceedings of the National Academy of Sciences of the United States of America 112, 11001 11006. https://doi.org/10.1073/pnas.1506254112

Werner, S., Engler, C., Weber, E., Gruetzner, R., Marillonnet, S., 2014. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. Bioengineered 3, 38 43. https://doi.org/10.4161/bbug.3.1.18223

White, M.A., 2015. Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. Genomics 106, 165 170. https://doi.org/10.1016/j.ygeno.2015.06.003

White, M.A., Kwasnieski, J.C., Myers, C.A., Shen, S.Q., Corbo, J.C., Cohen, B.A., 2016. A Simple Grammar Defines Activating and Repressing cis-Regulatory Elements in Photoreceptors. CellReports 17, 1247 1254. https://doi.org/10.1016/j.celrep.2016.09.066

Wolenski, F.S., Chandani, S., Stefanik, D.J., Jiang, N., Chu, E., Finnerty, J.R., Gilmore, T.D., 2011. Two Polymorphic Residues Account for the Differences in DNA Binding and Transcriptional Activation by NF-κB Proteins Encoded by Naturally Occurring Alleles in Nematostella vectensis. Journal of Molecular Evolution 73, 325–336. https://doi.org/10.1007/s00239-011-9479-7

Wong, D., Teixeira, A., Oikonomopoulos, S., Humburg, P., Lone, I.N., Saliba, D., Siggers, T., Bulyk, M.L., Angelov, D., Dimitrov, S., Udalova, I., Ragoussis, J., 2011. Extensive characterization of NF-κB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. Genome Biology 12, R70. https://doi.org/10.1186/gb-2011-12-7-r70

Workman, C.T., Yin, Y., Corcoran, D.L., Ideker, T., Stormo, G.D., Benos, P.V., 2005. enoLOGOS: a versatile web tool for energy normalized sequence logos. Nucleic Acids Research 33, W389 92. https://doi.org/10.1093/nar/gki439

Xu, L., Yang, L., Liu, W., 2013. Distinct evolution process among type I interferon in mammals. Protein & cell 4, 383 392. https://doi.org/10.1007/s13238-013-3021-1

Yanai, H., Chen, H.-M., Inuzuka, T., Kondo, S., Mak, T.W., Takaoka, A., Honda, K., Taniguchi, T., 2007. Role of IFN regulatory factor 5 transcription factor in antiviral immunity and tumor suppression. Proceedings of the National Academy of Sciences 104, 3402 3407. https://doi.org/10.1073/pnas.0611559104

Yeow, W., Au, W., Juang, Y., Fields, C., Dent, C., Gewert, D., Pitha-Rowe, P.M., 2000. Reconstitution of virus-mediated expression of interferon alpha genes in human fibroblast cells by ectopic interferon regulatory factor-7. Journal of Biological Chemistry 275, 6313 6320.

Zhang, X., Li, T., Liu, F., Chen, Y., Yao, J., Li, Z., Huang, Y., Wang, J., 2019. Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. Molecular cell 73, 130 142.e5. https://doi.org/10.1016/j.molcel.2018.10.020

Zheng, W., Chung, L.M., Zhao, H., 2011. Bias detection and correction in RNA-Sequencing data. BMC Bioinformatics 12, 290. https://doi.org/10.1186/1471-2105-12-290

Zhou, X., Michal, J.J., Zhang, L., Ding, B., Lunney, J.K., Liu, B., Jiang, Z., 2013. Interferon Induced IFITFamily Genes in Host Antiviral Defense. International Journal of Biological Sciences 9, 200 208. https://doi.org/10.7150/ijbs.5613

Zhu, C., Byers, K.J., McCord, R., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M., Philippakis, A.A., Hu, Y., Masi, F., Pacek, M., Rolfs, A., Murthy, T., Labaer, J., Bulyk, M.L., 2009. High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Research 19, 556 566. https://doi.org/10.1101/gr.090233.108

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., Enard, W., 2017. Comparative Analysis of Single-Cell RNA Sequencing Methods. Molecular cell 65, 631 643.e4. https://doi.org/10.1016/j.molcel.2017.01.023

Zinn, K., Maniatis, T., 1986. Detection of factors that interact with the human beta-interferon regulatory region in vivo by DNAase I footprinting. Cell 45, 611 618. https://doi.org/10.1016/0092-8674(86)90293-x

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]

[redacted]