1. Overview of the research program:

A specific and targeted response to external stimuli requires the large scale modification of cellular gene expression. The control code for each gene is written into DNA elements either proximal (promoter) or distal (enhancer) to the gene. This code contains the necessary information in the form of DNA binding motifs to recruit the necessary transcription factors. As such, changing the transcription factor (TF) milieu will target different transcriptional codes and gene sets. This suggests that genes within a similar program must contain similar sets of cis DNA control code for transcription factor recruitment. By studying TF targeting we can gain insight in the sets of control codes used for transcriptional activation.

Changes in the nuclear TF milieu are, initially, a property of changes in the localization and post translational modifications of previously expressed TFs1. Considering this, to study changes in TF binding from a null state, we can focus on TFs which are excluded from the nucleus in the basal state. This also allows us to study the different contexts and effects of different signals on global binding patterns. One of the most recognized of these signal dependent TFs is the NF-κB family. *Our focus is to define the differential DNA binding and large scale protein complex recruitment of the NF-κB family members*. Our current work in based on using next generation sequencing (NGS) technologies along with machine learning analysis to infer the DNA binding partners which specify the distinct DNA binding of NF-κB family members by their respective DNA motifs. Over the next five years, we will continue to develop and improve our predictive models of TF targeting to the genome and expand upon the predictive modeling through proteomic approaches. By learning the complimentary tendencies of transcription factors, we may be able to infer the complex and precise binding of proteins at target DNA locations. We also intend to package our bioinformatic pipelines as we have done previously2 to allow others to easily reproduce our analysis on NF-κB in different contexts or on other transcription factors in a simple to use command line tool. Our longer term objectives are to define the NF-κB containing DNA activation complexes by their underlying cis DNA binding motifs and become leaders in the field of NF-κB transcriptional networks.

1. Review of the literature relevant to this program:

TFs target cis regulatory elements through binding of a DNA binding domain to a compatible DNA element2–5. The consensus sequence for these targeted DNA elements are called DNA binding motifs5. However, the presence of a DNA binding motif is not sufficient to recruit the target TF. It is through the specific grouping and arrangement of these motifs which recruits a set of corresponding TFs working together that a region is targeted for transcriptional activation or repression2,6,7. The co-recruited TFs interpret the underlying DNA and coordinate the binding of chromatin modifiers, coactivators and coregulatory factors that must fit together in a large and intricate puzzle that provides the appropriate output on a gene8. The interpretation of DNA is further complicated by TF families which may contain dozens of protein family members capable of binding to similar if not identical DNA binding motifs2. In these circumstances, the surrounding environment is responsible for determining the TF family member specificity2,9,10 (Figure 1). As such, by learning the complimentary tendencies of transcription factors, we may be able to infer the complex and precise binding of proteins at target DNA locations.

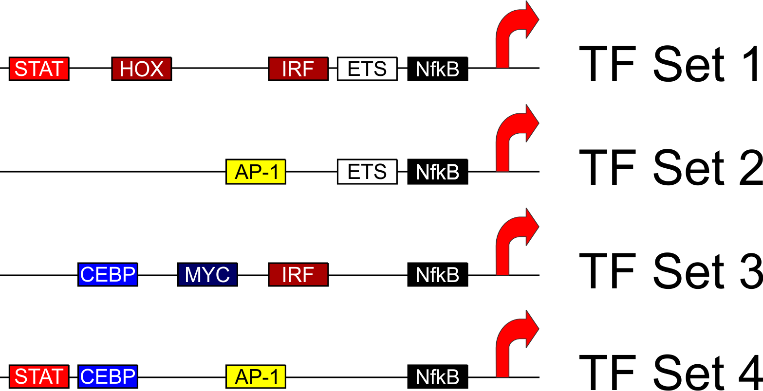


Figure 1. Several different combinations of TFs are necessary for activation. DNA motifs are labelled.

*NF-κB family members are involved in specific and diverse cellular processes*

The Nf-κB family is a TF family which is made up of 5 members including RelA, RelB, cRel, p50 and p52. These members bind as hetero or homo dimers to a consensus DNA motif of GGGRNWYYCC and have been observed in nearly all possible combinations11. The Nf-κB family is defined by the Rel homology domain which contains the DNA binding, dimerization and inhibitory protein binding domains12,13. While all members have a Rel homology domain, the Rel and NfkB subsets within the family differ in other structural domains. The Rel subset (RelA, RelB and cRel) contain a transactivation domain4,12. In contrast, the Nf-κB subset (p50 and p52) do not contain a transactivation domain, and are typically thought of as repressors4,12. The Nf-κB family proteins are all nearly ubiquitously expressed and importantly for this project all 5 are expressed in monocyte derived macrophages in humans and bone marrow derived macrophages in mice. Despite binding to the same consensus motif, these TFs have different transcriptional targets seen by CRISPR knockout experiments in mouse BMDMs (Figure 2). They also have differing phenotypic outcomes suggesting that factors are specifically recruited to consensus binding regions14–18. This is most striking in mouse studies where knockout of RelA is embryonic lethal and myeloid specific knockout results in a reduction of myeloid cells18, while knockout of RelB is viable and results in excessive myeloid cell production and chronic inflammation 19. Knockout of p52 has no known myeloid phenotype. Instead, knockout of p52 results in B-cell production and activation defects16,17. Considering different phenotypes, ubiquitous expression in myeloid cells and similar DNA binding motifs, the specificity of function is likely determined by regions outside of the Rel homology domain.

Figure 2. CRISPR knockout of Nf-κB family members results in differential transcriptional profiles. CRISPR target is labelled at the bottom.

*Macrophages as a model organism for studying transcription.*

As previously stated, macrophages express all 5 members of the Nf-κB family, with no detectable binding in the basal state. Macrophages are also growth arrested, removing the confounding effects of cell cycle and differentiation on DNA targeting and activation. Further, macrophages may be stimulated with a plethora of signaling molecules to resulting in diverse and distinct nuclear environments and targeting contexts. This will allow us in the future to expand our test conditions and define a hierarchy of collaborating TF associations as we can test many possible combinations of TFs in an *ex vivo* state. Lastly, macrophages are an easy cell type to acquire samples from both mouse and humans (in conjunction with Dr. Larry Lands) which allows for the reproducible study of species specific functions.

In this study, we will use this small family of TFs to identify the underlying cis DNA elements which define binding.

1. Our progress in the area:

Preliminary Studies: Previously, we have developed a machine learning model using logistical regression for predicting TF DNA binding2. This model allowed us to tease out the subtle binding differences between AP-1 family members that lead to reproducible and distinct patterns of recruitment to the genome. Our logistic regression model, which we called Transcription factor Binding Analysis (TBA)2, is based on ascribing a positive or negative weight to regions called by HOMER (Hypergeometric Optimization of Motif EnRichment)7. We utilized FIMO (Find Individual Motif Occurrences) to scan each positively and negatively labeled loci for the best matching motif within our motif library, retaining the position, orientation, and log ratio score (a measure of quality from the motif)20. This ensemble of positively and negatively associated classifiers captures the relative importance of each DNA motif in the recruitment of TFs to the genome and paints a broad picture of the larger complexes which form on the DNA for transcriptional activation. We can validate our approach by comparing our sequence based map with a data driven map, which was built using replicate ChIP-sequencing data for all family members. Both of these approaches should partition highly similar clusters. Using mouse bone marrow derived macrophages, we have found that Nf-κB monomers bind distinctly and reproducibly to specific loci on the genome. Importantly, similarly to our previous work on the AP-1 family, binding differences were not simply due to the presence of an Nf-κB DNA MOTIF. Instead, using a TBA on the surrounding DNA MOTIFs, we found that binding of Nf-κB was largely defined by the Nf-κB motif, however, binding of the specific monomers was selected by the surrounding TF landscape (Figure 3). Using this approach, we predict distinct factors involved in the specific recruitment of Nf-κB family members.

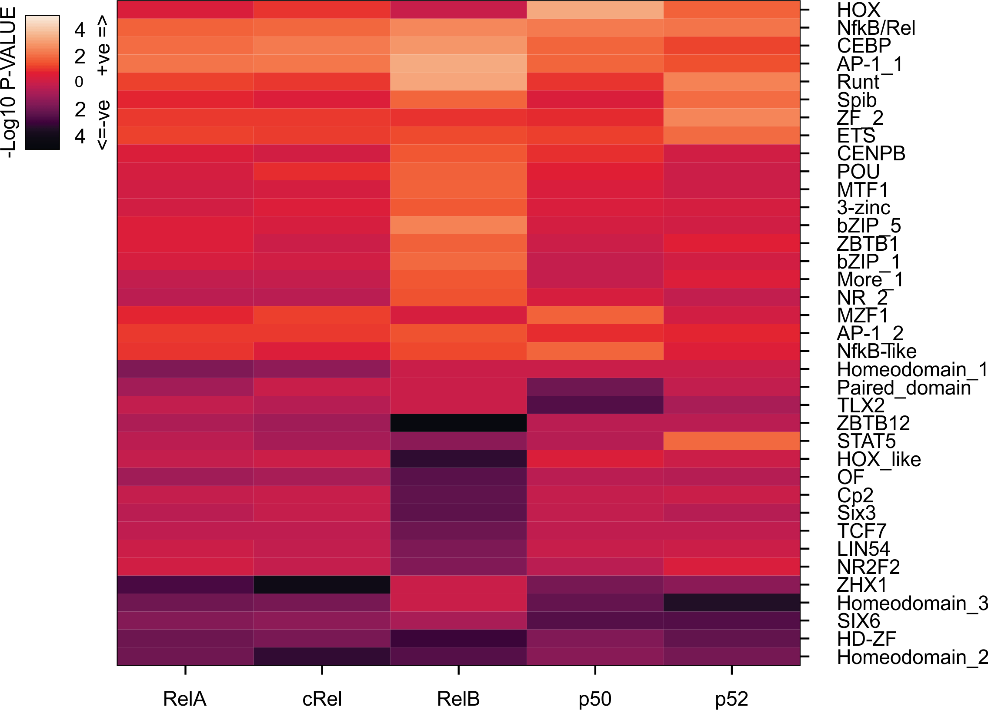


Figure 3. Predicted importance by TBA. Using TBA, we find DNA motifs which predict Nf-κB family member binding to DNA. Importance score is given as the -Log10 p-value with the motifs positively associated with binding in lighter colour and negatively associated with binding in darker colour.

1. Objectives
2. Determining the cis DNA factors which predict each Nf-κB family members specific binding in monocyte derived macrophages.
3. Comprehensive comparison of human and mouse macrophage Nf-κB targeting
4. Using FPLC Mass-spectrometry to identify specific Nf-κB family members complexes.

5. Methodology and program design:

*5.1:* *Determining the cis DNA factors which predict each Nf-κB family members specific binding in human monocyte derived macrophages.* Here, I will use chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) and RNA-seq to study Nf-κB in both human and mouse macrophage in an activated setting (LPS activation). Importantly, both human and mouse macrophages express all Nf-κB family members and our antibodies which we have optimized on mouse Nf-κB family members are predicted to work against human. With help from the lab of Dr. Larry Lands, my neighbour at the Meakins-Christie institute, we have access to large numbers of monocyte derived macrophages, making this dataset a simple and opportunistic way to greatly increase our understanding of TF binding. In order to integrate data for each of the aims, the same protocol for macrophage activation will be used throughout. The time windows of interest will be 4 hours for LPS treatment, when promoters and enhancers exhibit the complete repertoire of activated or repressed histone modifications. For human macrophage isolation, monocytes will be purified using CD-14 magnetic beads and plated into 10% human serum and M-CSF for 4 days to produce macrophages 21,22. For mice, bone marrow derived macrophages will be isolated by flushing bones of 8 week old mice and treated for 7 days with M-CSF.

Previously, I used machine learning to elucidate the cis elements that defined binding of AP-1 2. Similar to the AP-1 family, the Nf-κB family bind to a highly similar DNA motif. Using this approach, I will define the TF environment which specifies each Nf-κB family member binding in mouse and human macrophages. I will then test our model globally by using CRISPR-Cas9 to selectively remove TFs which are predicted to select Nf-κB family members for occupancy at functional loci (Figure 2). I will then use the results of the logistic regression model (Figure 3) to inform the attention neural network model, DEEPBIND. I will use DEEPBIND to define the transcriptional factor motifs sets which are found to collaborate to recruit specific Nf-κB family members. I will also test our model more specifically by CRISPR-Cas9 deletion of motifs at specific sites which our models predict and effect and by the creation of artificial binding elements (sets of motifs with predicted binding specificity to select Nf-κB family members). In the future, we will study how changes in the transcription factor milieu as a response by other signalling molecules such as IL-4 (M2 macrophages) or IL-13 (Wound healing macrophages) can lead to different co-associating transcription factor networks.

*Anticipated results.* Our preliminary data using mice has shown distinct DNA motif preference for the recruitment of Nf-κB family members to chromatin locations. We expect similar results in human macrophages. We also expect to build artificial enhancers with preference for each Nf-κB family members. These will provide a future resource to interrogate Nf-κB family members binding and activation requirements using screening technologies such as CRISPR.

*5.2 Comprehensive comparison of Nf-κB targeting in human and mouse macrophage*. Currently, the level of conservation in Nf-κB activity has not been determined at the level of genomic binding. As well, the conservation of particular Nf-κB family member distribution and co-recruiter requirements have not been studied. To accomplish this, we will measure the similarity in transcriptional effects by comparing the similarity of response in the CRISPR-Cas9 mediated Nf-κB knock out experiments (Figure 2) by hierarchical clustering and principal component analysis. We will also analyze the similarities between the total distribution and make up of the DNA binding of Nf-κB factors and determine overlap in promoter targeting. Unfortunately, enhancers, which make up the majority of Nf-κB target regions, may not be highly conserved between mouse and humans. However, the patterns of co-associating motifs can be used in lieu of location as a global measure of binding similarity (Figure 3). To do this, we will align the predicted co-associating DNA motifs from TBA from all of our mouse and human Nf-κB ChIP-seq samples. We will determine similarity using measures such as hierarchical clustering and principal component analysis. We will also look for similarity in the sets of co-associated DNA motifs as measured by DEEPBIND. With these datasets, we will be able to measure the similarity in function between human and mouse Nf-κB on transcriptional targeting and activation. We will then perform rescue experiments where we will replace each mouse Nf-κB family member with the human equivalent by using a vector which contains both expression of the human Nf-κB family member as well as guide RNAs targeting only the mouse Nf-κB equivalent in mouse BMDMs containing saCas9, which are available through the Meakins-Christie. We will then measure transcriptional response to LPS at 4 and 24 hours and determine the similarity to control infected samples. Finally, we will test the specificity of the artificial enhancers which we developed in 5.1 for human Nf-κB family members for their ability to specifically activate in mouse.

*Anticipated results.* Human and mouse Nf-κB family members share high levels of homology ranging from 76-91% identity. However, these discrete differences likely code for slight variation in species signal response and protein binding motifs. We expect to define the differences in recruitment between these family members. In the future, we will mutational analysis to determine the protein interactions which drive these differences.

*5.3. Using Mass-spectrometry to identify specific Nf-κB family members complexes*. With an understanding of the TF sets which are co-recruited with each Nf-κB family member, we will study the large-scale complexes which are recruited by Nf-κB family members. To do this, we will make use of the state of the art proteomics core of the RI-MUHC. Initially, we will perform immunoprecipitation (IP) on RelA and p52 to cover one of the activating and one of the repressing family members. Nuclei will be extracted from mouse bone marrow derived macrophages and treated with MNase to free DNA bound complexes. Precipitates will be run on an FPLC to separate complexes. Complexes will be sequenced via LC-MS/MS. Sequencing from the complexes will then be aligned. The sequenced complexes will be sorted into the TF sets determined in 5.1 for each factor based on the inclusion of the TFs from these sets. We will then assign these complexes to locations on the genome based on the TF set. Will we test the accuracy of this approach using ChIP-PCR for members of the MASS-spec sequenced complex at the predicted binding sites based on TF set assignment. We will then perform genome run-on sequencing (GRO-seq) to determine the relative level of RNA polymerase genome-wide in both untreated and LPS-4h treated BMDMs. Using this, we can assign a level of activation in LPS treatment or an increase in activation from untreated to LPS treatment for the TF sets and their accompanying complexes at a genome wide scale. This will help us to determine the transcriptional function of TF sets.

6. Significance. In this program, we will thoroughly study the cis DNA elements (DNA binding motifs) which define the specific binding of Nf-κB family members. We will define the similarities and differences in the Nf-κB network in human and mouse macrophages. Lastly, we will use this information to build interacting sets of DNA motifs and build the large scale transcriptional complexes from MASS-spec data. This work will develop bioinformatic pipelines to define differences in DNA binding by the accompanying DNA motifs at a genome-wide scale. This will allow for rapid and in depth analysis of ChIP-seq results using machine learning models. We will release this package as a simple command line program to open this type of research to the widest possible audience. Through this work, we hope to better understand transcriptional activation complexes in LPS treatment. In the future, we hope to expand this research into other treatment and cell contexts. This project will provide training for Ph.D. and undergraduate students in genomics research and will lead to significant publications in the fields of Nf-κB and DNA motif analysis.

7. References

1. Vandromme, M., Gauthier-Rouvière, C., Lamb, N. & Fernandez, A. Regulation of transcription factor localization: fine-tuning of gene expression. *Trends Biochem. Sci.* **21**, 59–64 (1996).

2. Fonseca, G. J. *et al.* Diverse motif ensembles specify non-redundant DNA binding activities of AP-1 family members in macrophages. *Nat. Commun.* **10**, 414 (2019).

3. Halazonetis, T. D., Georgopoulos, K., Greenberg, M. E. & Leder, P. c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. *Cell* **55**, 917–24 (1988).

4. Zhang, Q., Lenardo, M. J. & Baltimore, D. 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. *Cell* **168**, 37–57 (2017).

5. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).

6. Kaikkonen, M. U. *et al.* Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* **51**, 310–25 (2013).

7. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–89 (2010).

8. Liu, Z. *et al.* Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex. *Cell* **159**, 358–373 (2014).

9. Heinz, S. *et al.* Effect of natural genetic variation on enhancer selection and function. *Nature* (2013). doi:10.1038/nature12615

10. Barozzi, I. *et al.* Coregulation of Transcription Factor Binding and Nucleosome Occupancy through DNA Features of Mammalian Enhancers. *Mol. Cell* **54**, 844–857 (2014).

11. Zhang, Q., Lenardo, M. J. & Baltimore, D. 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. *Cell* **168**, 37–57 (2017).

12. Hayden, M. S. & Ghosh, S. NF-κB, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev.* **26**, 203–34 (2012).

13. Zhang, Q., Lenardo, M. J. & Baltimore, D. 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. *Cell* **168**, 37–57 (2017).

14. Wan, F. & Lenardo, M. J. Specification of DNA binding activity of NF-kappaB proteins. *Cold Spring Harb. Perspect. Biol.* **1**, a000067 (2009).

15. Cildir, G., Low, K. C. & Tergaonkar, V. Noncanonical NF-κB Signaling in Health and Disease. *Trends in Molecular Medicine* **22**, 414–429 (2016).

16. Franzoso, G. *et al.* Mice deficient in nuclear factor (NF)-κB/p52 present with defects in humoral responses, germinal center reactions, and splenic microarchitecture. *J. Exp. Med.* **187**, 147–159 (1998).

17. Caamaño, J. H. *et al.* Nuclear factor (NF)-κB2 (p100/p52) is required for normal splenic microarchitecture and B cell-mediated immune responses. *J. Exp. Med.* **187**, 185–196 (1998).

18. Stein, S. J. & Baldwin, A. S. Deletion of the NF-κB subunit p65/RelA in the hematopoietic compartment leads to defects in hematopoietic stem cell function. *Blood* **121**, 5015–5024 (2013).

19. Weih, F. *et al.* Multiorgan inflammation and hematopoietic abnormalities in mice with a targeted disruption of RelB, a member of the NF-κB/Rel family. *Cell* **80**, 331–340 (1995).

20. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

21. Hiebl, B., Fuhrmann, R. & Franke, R. P. Characterization of cryopreserved CD14+-human monocytes after differentiation towards macrophages and stimulation with VEGF-A(165). *Clin. Hemorheol. Microcirc.* **39**, 221–8 (2008).

22. Rey-Giraud, F., Hafner, M. & Ries, C. H. In vitro generation of monocyte-derived macrophages under serum-free conditions improves their tumor promoting functions. *PLoS One* **7**, e42656 (2012).