

The NF-κB Family of Proteins

Mike Schachter
May 31, 2021

[Introduction](#)

[Proteins in the NF-κB Pathway](#)

[Cellular Functions of the NF-κB Pathway](#)

[Structure of the Rel Homology Region](#)

[Protein Sequence Homology Across Species](#)

[Transcription Factor Binding to DNA](#)

[Pathogenic Mutations in NFKB1 Sequence](#)

[Tissue-specific Expression](#)

[Tools Used and Source Code](#)

[References](#)

Introduction

NF-κB is not a single protein, it is a family of transcription factors, and a pathway. The NF-κB pathway is a “rapid responder” when cellular damage or pathogens are sensed within or outside of a cell. NF-κB proteins are found across all mammals and homologues of the proteins and pathways are found in invertebrates such as Drosophila (Gilmore et. al 2012). The most well known function of the pathway is to transcribe genes that produce an inflammatory response. Dysfunction in the proteins or pathway can lead to immune disorders and cancer.

Proteins in the NF-κB Pathway

There are three groups of proteins in the NF-κB pathway (Figure 1):

1. **Rel-homology Domain Proteins:** these 5 proteins contain Rel-homology domains that can bind to DNA, and directly act as transcription activators, in the case of p65/RelA, c-Rel, or RelB, or as repressors, in the case of p105/NFKB1, and p100/NFKB2. These proteins are always transcribed and floating around in the cytoplasm, but are inhibited by IκB proteins.

2. **IκB Proteins:** these proteins attach to Rel proteins and inhibit their action, so they aren't translocated to the nucleus. They detach from Rel proteins when phosphorylated by IKK proteins.
3. **IKK Proteins:** these proteins integrate signals from receptors that sense damage to the cell, or pathogens. When they detect these signals, they phosphoporylate IκB proteins, which detaches them from the NF-κB dimers and allows the dimers to translocate to the nucleus, where they can promote or inhibit transcription.

This report will primarily focus on the most common Rel dimer, which comprised of a cleaved NFKB1 (called p50) and RelA (also called p65). This dimer can be called p50/RelA or p50/p65.

Cellular Functions of the NF-κB Pathway

NF-κB dimers can promote transcription of many genes and have diverse effects on the cell that involve cell survival, proliferation, inflammation, and angiogenesis (Figure 2).

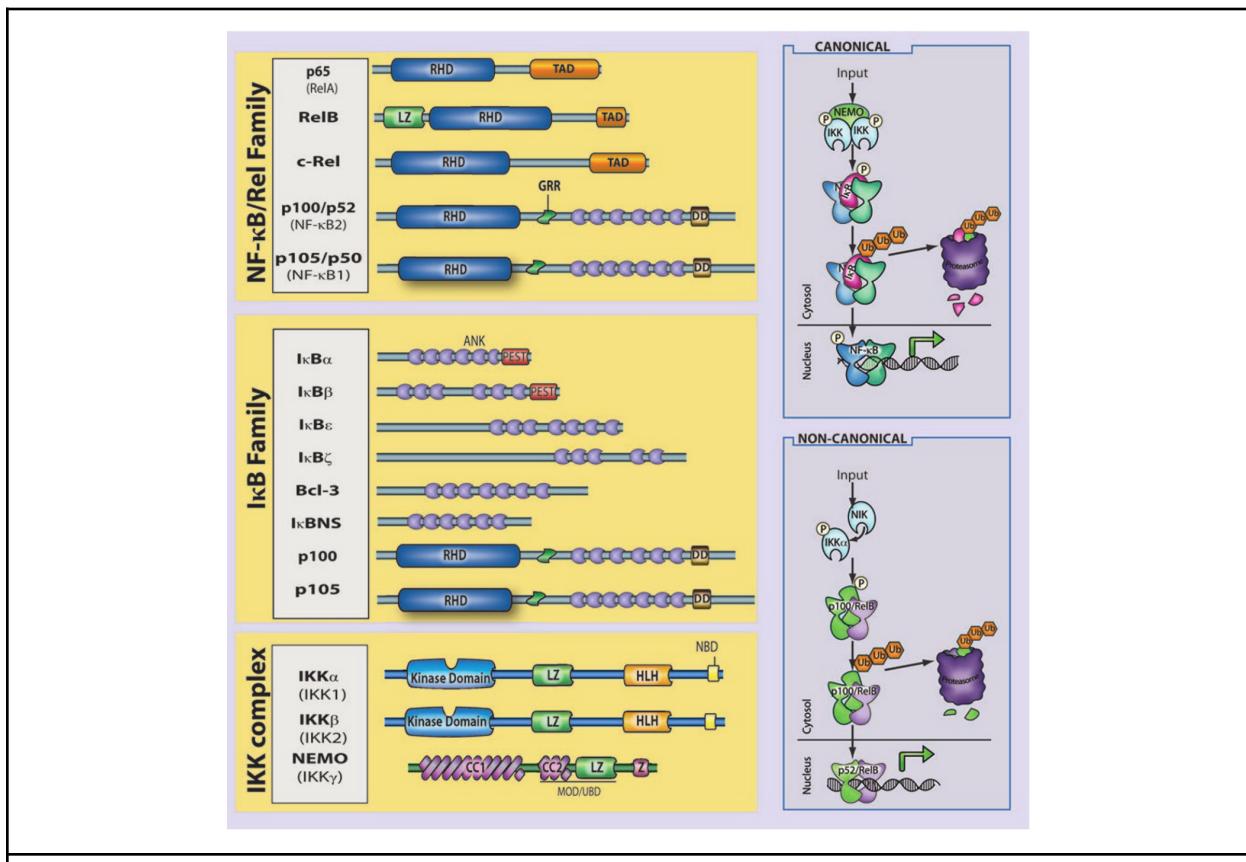


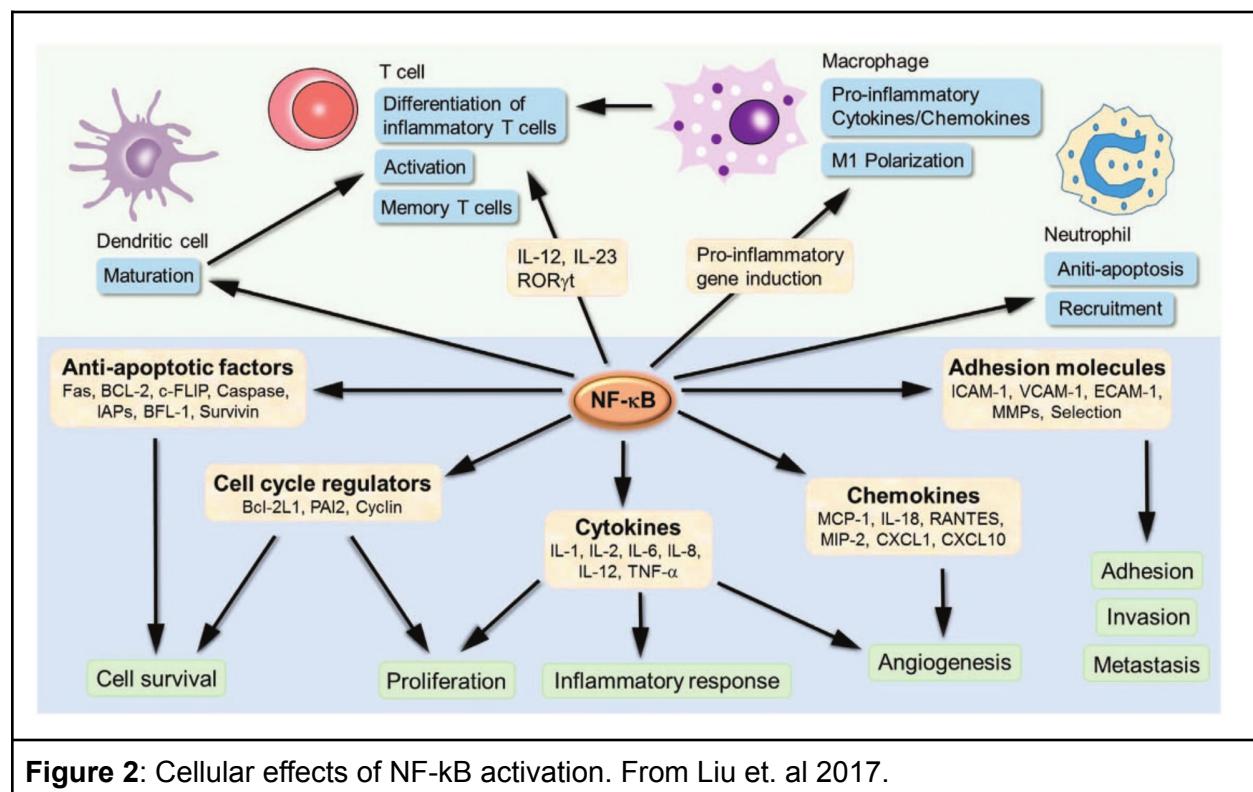
Figure 1: Proteins in the NF-κB pathway. From Hayden et. al 2011.

The NF-κB pathway plays its most prominent role in the mediation of the immune system response to cellular damage and pathogens. Liu et. al (2017) briefly reviews several components of this response. In macrophages, pathogen associated molecular patterns

(PAMPs) activate Toll-like receptors (TLRs). These receptors communicate downstream and eventually activate IKKs, which then dis-inhibit NF- κ B dimers. The dimers translocate to the nucleus and promote transcription of inflammatory cytokines such as IL-1 β . These cytokines promote a strong immune response and are necessary for activating T-cells in the adaptive immune system.

Likewise, activation NF- κ B pathway is a necessary component for T-cell maturation and differentiation. The pathway is also a critical component in the development of inflammasomes, which are protein scaffolds that sense cellular damage, promote inflammation through the release of cytokines, and also promote scavenging of damaged organelles.

Shown in Figure 2 and also detailed in a review by Luo et. al (2005), the NF- κ B pathway promotes cell survival. It does this by promoting transcription of anti-apoptotic proteins such as Bcl-2. TNF-alpha is a ligand which activates receptors that can weakly induce cell death. TNF-alpha also activates the NF- κ B pathway, which then promotes anti-apoptotic proteins and foils the attempts of TNF-alpha. Dysfunction in NF- κ B pathway proteins that produce chronically active NF- κ B are implicated in a variety of cancers.



Structure of the Rel Homology Region

The five proteins in the NF- κ B/Rel family all share a common domain called the “Rel Homology Region”, or “Rel Homology Domain” (Figure 3). The region is comprised of three major modules, including an “Amino-terminal Domain” (NTD), a “Dimerization domain” (DimD), and a “Linker”

(L) (Huxford et. al 2009). The L and part of the NTD domain are the main regions that contact DNA, while the DimD region serves to connect to the DimD region of another Rel protein.

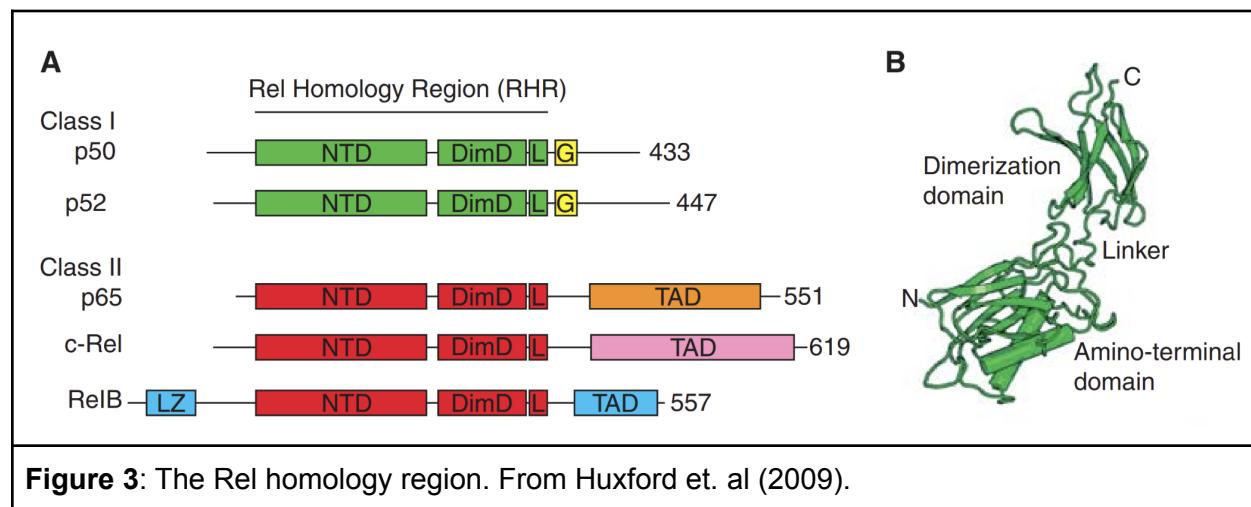


Figure 3: The Rel homology region. From Huxford et. al (2009).

There is some complexity to the NFKB1 protein that must be discussed, which also applies to NFKB2. They have structural and functional differences to RelA, RelB, and c-Rel. NFKB1 is transcribed as a 105kD protein. NFKB1/p105 is cleaved into the p50 protein, which can then pair up with another Rel protein to form a Rel dimer. However, uncleaved NFKB1/p105 is also present in cells and is necessary for cell survival. As shown in Figure 1, p105 contains ankyrin repeats similar to what is found in I κ B family proteins, and this is not a coincidence. p105 proteins can inhibit the activity of NF- κ B dimers, similar to what can be done by I κ B proteins. So the NFKB1 protein serves a dual role - when cleaved into p50, it forms a dimer that can promote transcription, and when uncleaved as p105, it can inhibit translocation of NF- κ B dimers to the nucleus. Also, homodimers of p50/p50 do not contain Transcription Activation Domains (TADs), and they inhibit transcription. It is only when p50 joins with a Rel dimer that contains a TAD, such as RelA, RelB, or c-Rel, that the translocated dimer can promote transcription.

Not discussed, but Figure 4 also shows many sites where NFKB1 can be functionally modified through phosphorylation, acetylation, hydroxylation, or nitrosylation.

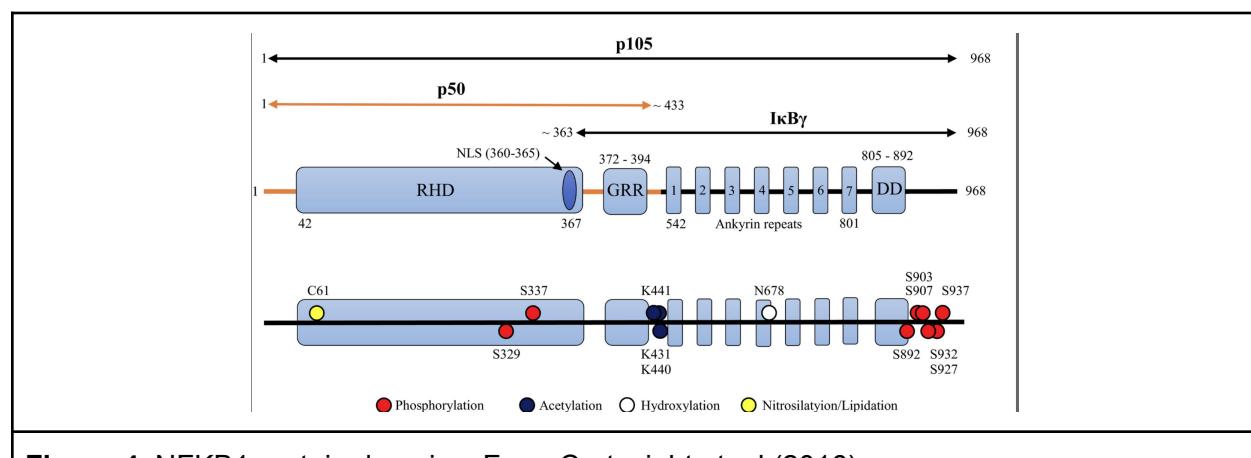


Figure 4: NFKB1 protein domains. From Cartwright et. al (2016).

Protein Sequence Homology Across Species

Three commonly studied organisms were selected for protein sequence comparison for the NFKB1 protein - humans, macaques, and mice. A protein sequence alignment was performed using Clustal Omega (Sievers et. al 2011). Figure 5 shows the percent identity matrix between the organisms. The results show a highly conserved sequence between macaque and human, but less similarity between the primates and mouse.

1: sp P25799 NFKB1_MOUSE	100.00	87.14	86.93
2: sp P19838 NFKB1_HUMAN	87.14	100.00	98.86
3: tr H9Z1D5 NFKB1_MACMU	86.93	98.86	100.00

Figure 5: Percent identity matrix from a multiple sequence alignment of NFKB1.

Figure 6 shows some detailed results from multiple sequence alignment. The largest deviation of mouse protein sequence occurs in the very beginning of the sequence, then the akyrin repeat region (395+), and the death domain (805+). The Rel homology domain (42-376) is highly conserved (not shown).

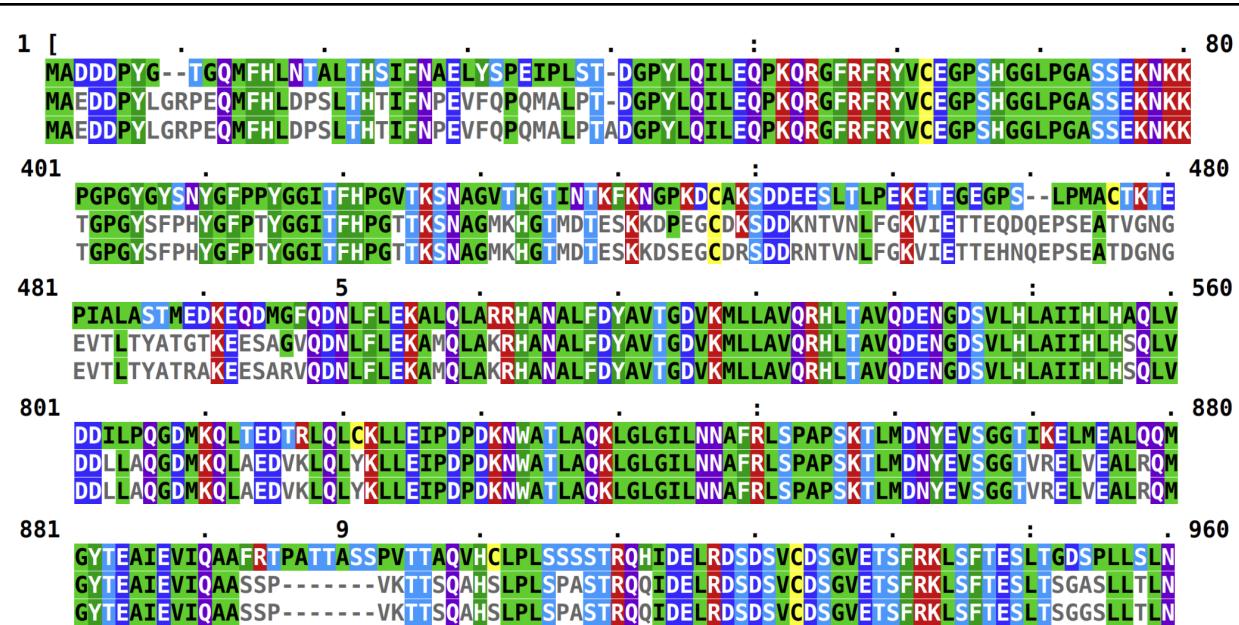


Figure 6: Portions of a multiple sequence alignment for NFKB1 between mouse (top), human (middle), and macaque (bottom). Numbers in the upper left and right of each segment indicate amino acid location.

Transcription Factor Binding to DNA

As transcription factors, NF-kB dimers bind to DNA and promote transcription of a number of genes important to the immune system and cell survival. Figure 7 shows the common binding motifs for NFKB1 (p50) and RelA, retrieved from the tf2dna database (Pujato et. al 2014). These binding motifs were plotted by using the Python library [logomaker](#).

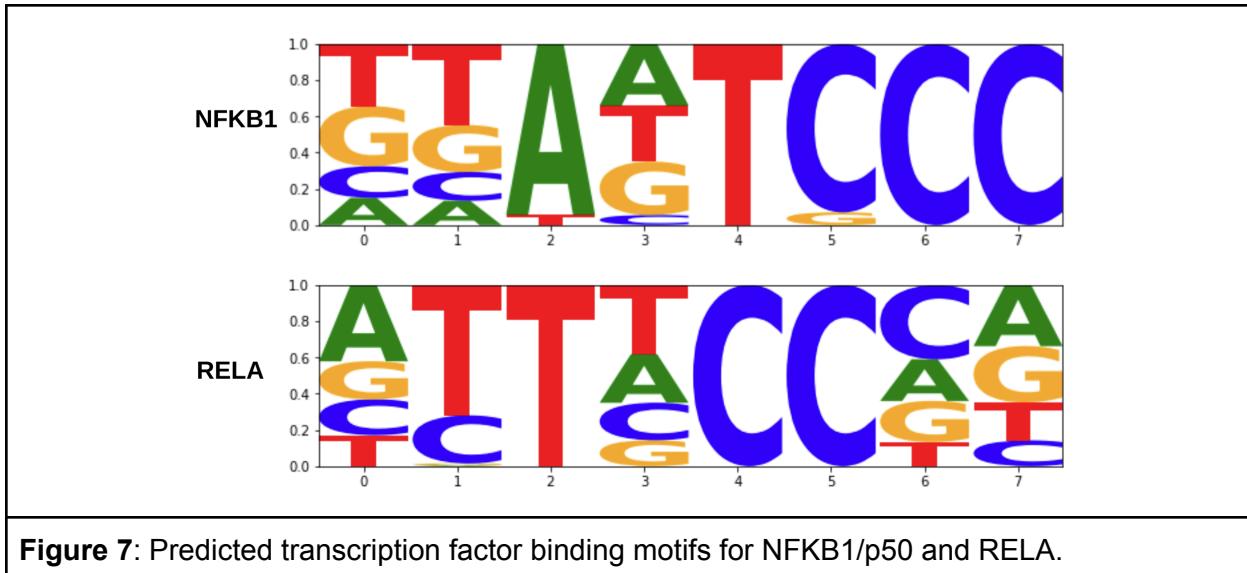


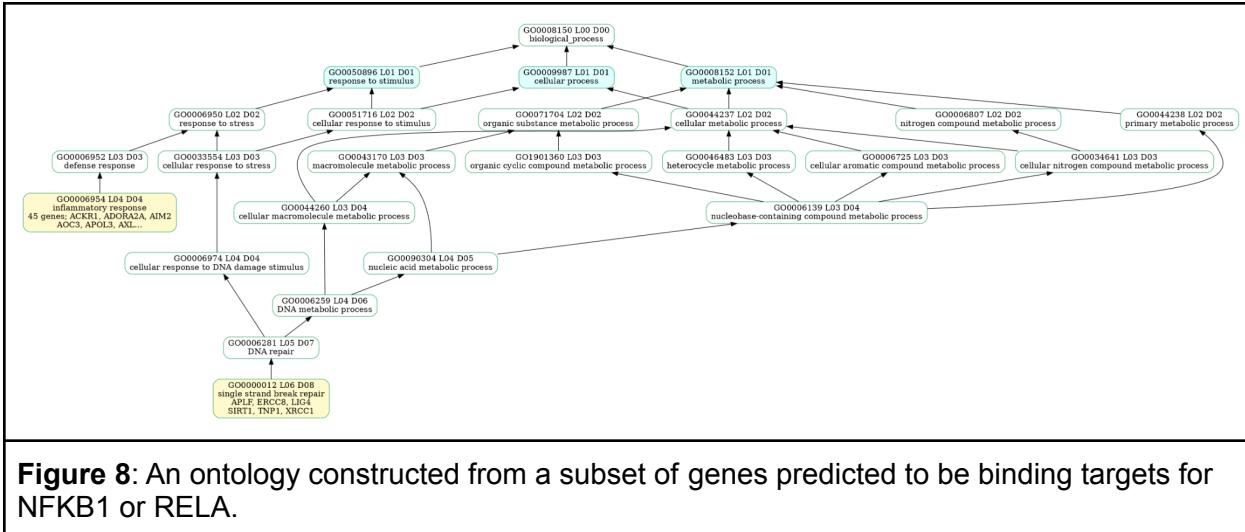
Figure 7: Predicted transcription factor binding motifs for NFKB1/p50 and RELA.

There are many genes with potential upstream binding sites for NF-kB dimers. Methods such as ChIP-seq attempt to isolate protein-DNA pairs and sequence the DNA to obtain binding motifs. In the works of Pujato et. al (2014), a 3D model of transcription factors was used to predict binding affinities to all possible sequences of DNA base pairs. Data about these sequences was added to the tf2dna database, and combined with experimental and theoretical sources.

Complementary to resources such as the tf2dna database are curated lists of genes known to be promoted by NF-kB dimers, such as presented by Pahl (1999). A website listing those genes can be found [here](#).

An attempt to reconcile several data sources about NF-kB targets and understand their function was made in this report. First, all experimental data from the tf2dna database on NFKB1 and RELA was downloaded. The data consisted of .pscan files that contained NFKB1 or RELA targets, their p-values, and binding scores. Next, targets were excluded if they had a p-value of greater than 1e-5. Remaining targets were then combined across experimental datasets and de-duplicated, then intersected between NFKB1 and RELA to produce a final list of around 1000 target genes.

The list of target genes was then run through the gene ontology database, and a graph was created for the ontology using the GOATOOLS Python package. The results can be seen in Figure 8 and Table 1.



The results directly from the Gene Ontology website (Table 1) are much richer than the graph constructed from GOATOOLS. They show what would be expected - target genes of NFKB1 and RELA are involved in immune regulation, inflammation, stress response, cell differentiation DNA repair.

single strand break repair (GO:0000012)
interferon-gamma-mediated signaling pathway (GO:0060333)
cellular response to chemokine (GO:1990869)
regulation of type I interferon production (GO:0032479)
inflammatory response (GO:0006954)
regulation of response to external stimulus (GO:0032101)
regulation of response to stress (GO:0080134)
regulation of immune system process (GO:0002682)
cell differentiation (GO:0030154)

Table 1: The leaf nodes of the gene ontology constructed from a subset of binding targets for NFKB1 or RELA.

Pathogenic Mutations in NFKB1 Sequence

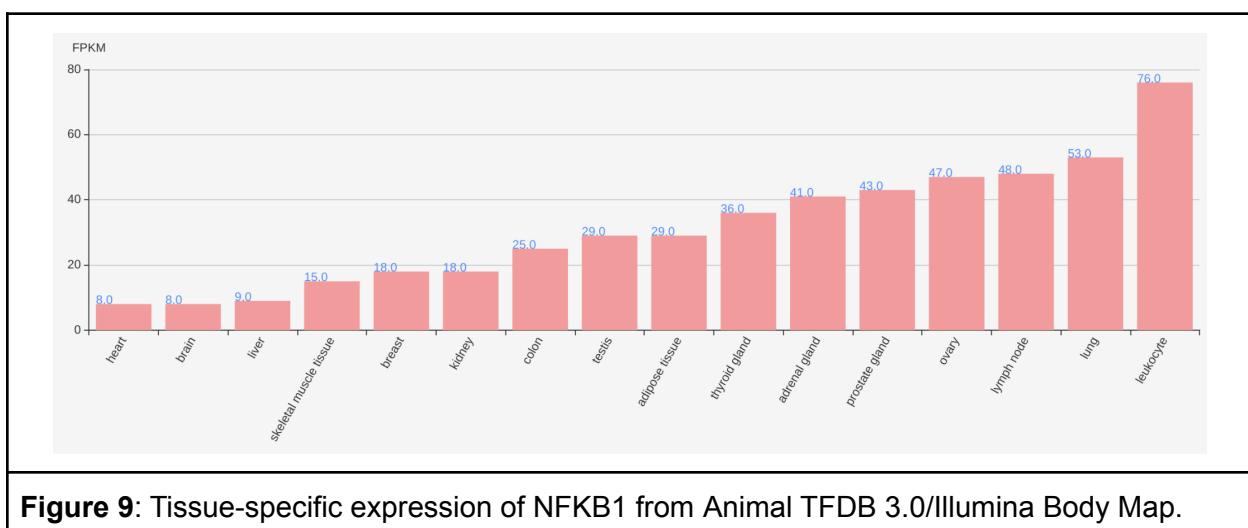
The Clinvar database was consulted to identify pathogenic variants in the DNA sequence of NFKB1 and RELA. To do this, the variant_summary.txt file was downloaded from the Clinvar FTP site and the results were analyzed in Python.

Pathogenic mutations in NFKB1 or RELA were identified. Of the 16 that were identified, 15 were in NFKB1, and analysis was restricted to those variants. The exon annotations for the NFKB1 gene were downloaded using the Python ensembl library, and 13 out of the 15 pathogenic mutations were found to occur in exons. Of the pathogenic variants occurring in exons, 5 were SNVs, 6 were deletions, and 2 were duplications.

A follow-up consultation using the clinvar website directly showed that many of the mutations were implicated in Common Variable Immune Deficiency (CVID), a disease that has a range of manifestations with a common underlying theme of having a weakened immune system.

Tissue-specific Expression

AnimalTFDB is a database that compiles and curates transcription factor information across species (Hu et. al 2019). The interface shows comprehensive information for transcription factors, including tissue-specific expression. Figure 9 shows the tissue-specific expression of NFKB1. Significant expression can be found in leukocytes, white blood cells that include all innate and adaptive immune system cells. Also interesting is high expression found in lung, and ovary, prostate, adrenal, and thyroid glands.



Tools Used and Source Code

The source code can be found in this [Jupyter Notebook](#).

The following tools were used:

- Biopython
- Numpy/Pandas/Matplotlib
- Logomaker
- Pyensembl
- Biothings
- ClustalOmega

The following databases were utilized:

- NCBI
- UniProt
- tf2dna
- JASPAR
- Animal TFDB 3.0
- UCSC Genome Browser
- Clinvar
- GeneOntology
- [Rel/NF- \$\kappa\$ B target genes](#)
- BioGRID
- TRANSFAC

References

Cartwright, T., Perkins, N. D., & Wilson, C. L. (2016). NF κ B1: a suppressor of inflammation, ageing and cancer. *The FEBS journal*, 283(10), 1812-1822.

Gilmore, T. D., & Wolenski, F. S. (2012). Evolution of NF κ B: where did it come from and why. *Immunol. Rev.*, 246, 14-35.

Hayden, M. S., & Ghosh, S. (2011). NF- κ B in immunobiology. *Cell research*, 21(2), 223-244.

Hu, H., Miao, Y. R., Jia, L. H., Yu, Q. Y., Zhang, Q., & Guo, A. Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic acids research*, 47(D1), D33-D38.

Huxford, T., & Ghosh, G. (2009). A structural guide to proteins of the NF- κ B signaling module. *Cold Spring Harbor perspectives in biology*, 1(3), a000075.

Liu, T., Zhang, L., Joo, D., & Sun, S. C (2017) NF- κ B signaling in inflammation. *Signal Transduct Target Ther*, 2, 17023.

Luo, J. L., Kamata, H., & Karin, M. (2005). IKK/NF-κB signaling: balancing life and death—a new approach to cancer therapy. *The Journal of clinical investigation*, 115(10), 2625-2632.

Pahl, H. L. (1999). Activators and target genes of Rel/NF-κB transcription factors. *Oncogene*, 18(49), 6853-6866.

Pujato, M., Kieken, F., Skiles, A. A., Tapinos, N., & Fiser, A. (2014). Prediction of DNA binding motifs from 3D models of transcription factors; identifying TLX3 regulated genes. *Nucleic acids research*, 42(22), 13500-13512.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.