Final report – PD 124843

Highly integrated genome level examination of transcription factor cascades during alternative macrophage polarization using next-generation sequencing methods

Gergely Nagy, Ph.D.

Introduction

As a bioinformatician, I am working on next-generation sequencing (NGS) data generated by experimental biologists of the research group of Prof. László Nagy, and also publicly available NGS data. During my 3-year fellowship, 8 papers – 7 research articles and 1 review – could be published, and most were tightly connected to my initial research plan, which aimed at the transcription factor (TF) cascade turned on during alternative macrophage polarization, resulting in so-called healing macrophages.

Discovering the TF cascade in alternatively polarized macrophages

Retinoid X receptor (RXR) is an obligate heterodimeric partner of Class II nuclear receptors (NRs), and as such a central component of NR signaling regulates the immune and metabolic phenotype of macrophages. Recognition motifs of RXR heterodimers are enriched at tissue-selective gene regulatory sites of resident macrophages, suggesting roles in subtype specification. Recent genome-wide studies revealed that RXR binds to thousands of sites in the genome, but the mechanistic details of how the whole set of binding sites (cistrome) is established and how it serves ligand-induced transcriptional activity remained elusive.

In our investigations, we showed that interleukin-4 (IL-4)-mediated macrophage plasticity resulted in a greatly extended RXR cistrome via both direct and indirect actions of the signal transducer and activator of transcription 6 (STAT6). Activation of STAT6 led to chromatin remodeling and RXR recruitment to already active and *de novo* enhancers, and this could also be confirmed by Stat6 *knock-out* (KO) experiments. The dynamics of both chromatin opening (measured by Assay for Transposase-Accessible Chromatin sequencing – ATAC-seq) and the recruitment of the master regulator PU.1, the coactivator P300, and the Cohesin component RAD21 (measured by Chromatin Immunoprecipitation sequencing – ChIP-seq) at the *de novo* RXR binding sites without the presence of STAT6 was delayed as compared to those sites with STAT6 binding. In addition, we could identify the motifs of a STAT6-triggered secondary TF wave, including that of peroxisome proliferator (PP)-activated receptor gamma (PPARγ) from Class II NRs. PPARγ appeared to be indispensable for the development of RXR-bound *de novo* enhancers, and the activity of some of these enhancers could be modulated by the ligands of the two NRs conferring ligand selective cellular responses. However, according to our ATAC-seq data from *Pparg* and *Rxr* KO cells, these enhancers still showed some chromatin opening in the lack of PPARγ or RXR upon IL-4 treatment, which suggested that

there were TFs other than PPARγ/RXR playing roles at these enhancers. Importantly, PPARγ and RXR ChIP density correlated well with ligand responsiveness, and also the PP response elements (PPREs) showed different compositions (Dániel and Nagy et al. NAR. 2018).

Importantly, only a small fraction of the extended RXR cistrome was ligand-induced or repressed and regulated genes accordingly; and most sites were unaffected by the used PPARγ agonist or antagonist, which latter suggested that PPARγ was able to function without endogenous ligand(s). Then, we found that upon repeated stimulation with IL-4, macrophages gained new functions, which were absent in *Pparg* KO cells. Namely, ligand-insensitive PPARγ binding controlled the expression of an extracellular matrix remodeling-related gene network. Expression of these genes increased during muscle regeneration in a mouse model of injury, and this increase coincided with the detection of IL-4 and PPARγ in the affected tissue. In total, we proposed an epigenomic ratchet mechanism of transcriptional memory, which supports a progressive macrophage polarization (Dániel and Nagy et al. Immunity. 2018).

During mapping the early (STAT6-induced) and late (PPARγ-driven) epigenomic changes of IL-4-induced alternative macrophage polarization, we identified the TF, early growth response 2 (EGR2), bridging the early transient and late stable gene expression program of polarization. EGR2 is a direct target of IL-4-activated STAT6, having broad action indispensable for 77% of the genes induced during alternative polarization, including its autoregulation and a robust, downstream TF cascade involving PPARγ. Mechanistically, EGR2 binding results in chromatin opening and the recruitment of chromatin remodelers and RNA Polymerase II. We also showed that *Egr2* induction is evolutionarily conserved during alternative polarization of mouse and human macrophages. In the context of tissue-resident macrophages, *Egr2* expression is most prominent in the lung of a variety of species. Thus, EGR2 is an example of an essential and evolutionarily conserved broad-acting factor, linking transient polarization signals to stable epigenomic and transcriptional changes in macrophages (Dániel and Czimmerer et al. Genes Dev. 2020).

Examination of PPARγ binding sites and the enhancer grammar of NRs and their collaborating partners

PPARγ is essential not only in alternatively polarized macrophages but also in adipocytes and their differentiation, so we investigated its DNA-protein interactions in these two models available in our laboratory. PPRE is composed of two directly repeated NR half-sites – one for PPAR, and one for RXR – and it is spaced by one nucleotide, so it has a co-called direct repeat 1 (DR1) structure. However, previously, it has not been analyzed systematically and genome-wide how *cis* factors such as the sequence of DR1s and adjacent sequences and *trans* factors such as co-binding lineage-determining TFs (LDTFs) contribute to the direct binding of PPARγ/RXR in different cellular contexts. We developed a novel motif optimization approach using sequence composition and (PPARγ and RXR) ChIP-seq densities from macrophages and preadipocytes to complement *de novo* motif enrichment analysis and to define and classify

high-affinity binding sites. We found that approximately half of the PPARγ cistrome represents direct DNA binding; both half-sites can be extended upstream providing an additional minor groove binding site leading to higher binding affinity, and PPAR and RXR half-sites are typically not of equal strength within a DR1. Importantly, strategically positioned LDTFs have a greater impact on PPARγ binding than the quality of PPRE, and the presence of the extension of DR1 provides a remarkable synergy with LDTFs. This approach of considering not only nucleotide frequencies but also their contribution to protein binding in a cellular context is applicable to other TFs (Nagy et al. MCB. 2020).

Emboldened by the term enhancer grammar fitting also the sequence requirements of direct PPARγ/RXR binding, we extended our interest to the binding sites of other NRs, basic leucine zipper (bZIP) proteins – several of which are key factors in macrophages –, and any further TFs with known, complex sequence needs, and collected them from the literature to a review article.

Collaboration of TFs and their recognition motifs in DNA is the result of coevolution and forms the basis of gene regulation. However, the way how these short genomic sequences contribute to setting the level of gene products is not understood in sufficient detail. The biological problem to be solved by the cell is complex because each gene requires a unique regulatory network in each cellular condition using the same genome. Thus far, only some components of these networks have been uncovered. In our work, we compiled the features and principles of the motif grammar, which dictates the characteristics and thus the likelihood of the interactions of the binding TFs and their co-regulators. We presented how sequence features provide specificity using the NRs and bZIP proteins. We also discussed the phenomenon of "weak" (low-affinity) binding sites, which appear to be components of several important genomic regulatory regions, but paradoxically are barely detectable by the currently used approaches. Assembling the complete set of regulatory regions composed of both weak and strong binding sites will allow one to get more comprehensive lists of factors playing roles in gene regulation, thus making possible the deeper understanding of regulatory networks (Gergely Nagy and László Nagy. CSBJ. 2020).

Then, I took part in a collaboration, in which we investigated the tip of the iceberg in gene regulatory networks, the so-called super-enhancers (SEs). SEs are clusters of highly active enhancers, regulating cell type-specific and disease-related genes, including oncogenes. The individual regulatory regions within SEs might be simultaneously bound by different TFs and co-regulators, which together establish a chromatin environment conducting effective transcription. While cells with distinct TF profiles can have different functions, how different cells control overlapping genetic programs remains a question. In this study, we showed that the construction of estrogen receptor alpha (ERα)-driven SEs is tissue-specific, both collaborating TFs and the active SE components greatly differ between human breast cancer-derived MCF-7 and endometrial cancer-derived Ishikawa cells; nonetheless, SEs common to both cell lines have similar transcriptional outputs. These results delineate that despite the

existence of a combinatorial code allowing alternative SE construction, a single master regulator – like ERα in these cells – might be able to determine the overall activity of SEs. ERα is one of the most known and most significant NRs, and the used cell lines are broadly used in this field (Bojcsuk, Nagy and Bálint. IJMS. 2020).

Examining the role of the "honorary nuclear receptor" BACH1 (bZIP) in macrophages of regenerating skeletal muscle

The infiltration and subsequent *in situ* subtype specification of monocytes to effector/ inflammatory and repair (healing) macrophages are indispensable for tissue regeneration upon acute sterile injury. However, the chromatin-level mediators and regulatory events controlling this highly dynamic macrophage phenotype switch are not well-known. We used a murine acute muscle injury model to assess global chromatin accessibility (ATAC-seq) and gene expression (RNA-seq) dynamics in infiltrating macrophages during sterile physiological inflammation and tissue regeneration. Based on our gene expression data and motif enrichment analyses of differential open chromatin regions – resulting in activation protein 1 (AP-1, bZIP) motifs and their specific extended forms, the so-called Maf response elements (MAREs) –, we identified an AP-1-related heme-binding transcriptional repressor, the BTB and CNC homology 1 (BACH1) as a novel regulator of this process. BACH1 competes with the nuclear factor erythroid 2 (NFE2) activators for both their heterodimerizing partners, the small Maf proteins and MAREs in a heme-dependent manner. *Bach1* KO mice displayed impaired muscle regeneration, altered dynamics of the macrophage phenotype transition, and transcriptional deregulation of key inflammatory and repair-related genes. We also found that BACH1 directly binds to and regulates distal regulatory elements of these genes, suggesting a novel role for BACH1 in controlling a broad spectrum of the repair response genes in macrophages upon injury. Inactivation of heme oxygenase-1 (*Hmox1*), one of the most stringently deregulated genes in the *Bach1* KO in macrophages, impairs muscle regeneration by changing the dynamics of the macrophage phenotype switch. Collectively, our data suggest the existence of a heme-BACH1-HMOX1 regulatory axis that controls the phenotype and function of the infiltrating myeloid cells upon tissue damage, shaping the overall tissue repair kinetics (Patsalos et al. JI. 2019).

Examination of the enhancer selection by IRF proteins in murine dendritic cells

Dendritic cells (DCs), similarly to macrophages, are professional antigen-presenting cells and share similar morphology and function, in general. Except for CCAAT-enhancer-binding proteins (C/EBPs) specific only for macrophages, these immune cells express the very same LDTFs like PU.1, interferon regulatory factor 4, or 8 (IRF4/8), and AP-1-related proteins. In our models, PU.1 and IRF8 tightly collaborate, mostly as heterodimers with lineage-determining roles, while the other expressed IRFs have distinct roles in DCs. IRF3, IRF5, and

IRF9 play distinct roles in the regulation of antiviral and inflammatory responses, but the determinants that mediate IRF-specific enhancer selection are not fully understood. In order to uncover regions occupied predominantly by IRF3, IRF5, or IRF9, we performed ChIP-seq experiments in activated murine DCs. The identified regions were analyzed with respect to the enrichment of DNA motifs, the interferon-stimulated response element (ISRE) and ISRE half-site variants, and chromatin accessibility (ATAC-seq). Using a machine learning method, we investigated the predictability of IRF dominance. We found that IRF5-dominant regions differed fundamentally from the IRF3- and IRF9-dominant regions: ISREs were rare, while the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) motif and special ISRE half-sites, such as 5'-GAGA-3' and 5'-GACA-3', were enriched. IRF3- and IRF9-dominant regions were characterized by the enriched ISRE motif and lower frequency of accessible chromatin. Motif enrichment analysis and the machine learning method uncovered the features that favor IRF3 or IRF9 dominancy (e.g., a tripartite form of ISRE and NF-κB motifs for IRF3, and the gamma interferon activation site [GAS] motif and certain ISRE variants for IRF9). This study contributes to our understanding of how IRF members, which bind overlapping sets of DNA sequences, can initiate signal-dependent responses without activating superfluous or harmful programs (Csumita et al. NAR. 2020).