The Many Faces of Gene Regulation: Extrinsic Control of Cell Fate and Function

Audrey Sporrij

The Many Faces of Gene Regulation: Extrinsic Control of Cell Fate and Function
A doctoral dissertation by Audrey Sporrij

All of the work presented in this thesis was performed at the Department of Stem Cell and Regenerative Biology of Harvard University, Cambridge, USA and was supported by a PhD fellowship from the Boehringer Ingelheim Fonds.

For reasons of consistency within this thesis, text may differ in respect to the articles that have been published.

ISBN: 978-94-6416-443-5
DOI: https://doi.org/10.33540/558
Design and layout by: Audrey Sporrij
Printed by: Ridderprint | www.ridderprint.nl

The Many Faces of Gene Regulation: Extrinsic Control of Cell Fate and Function

De vele gezichten van genregulatie:
Extrinsieke controle over het lot en de functie van cellen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

dinsdag 30 maart 2021
des middags te 4.15 uur

door

Audrey Sporrij

geboren op 4 september 1992
te Eindhoven

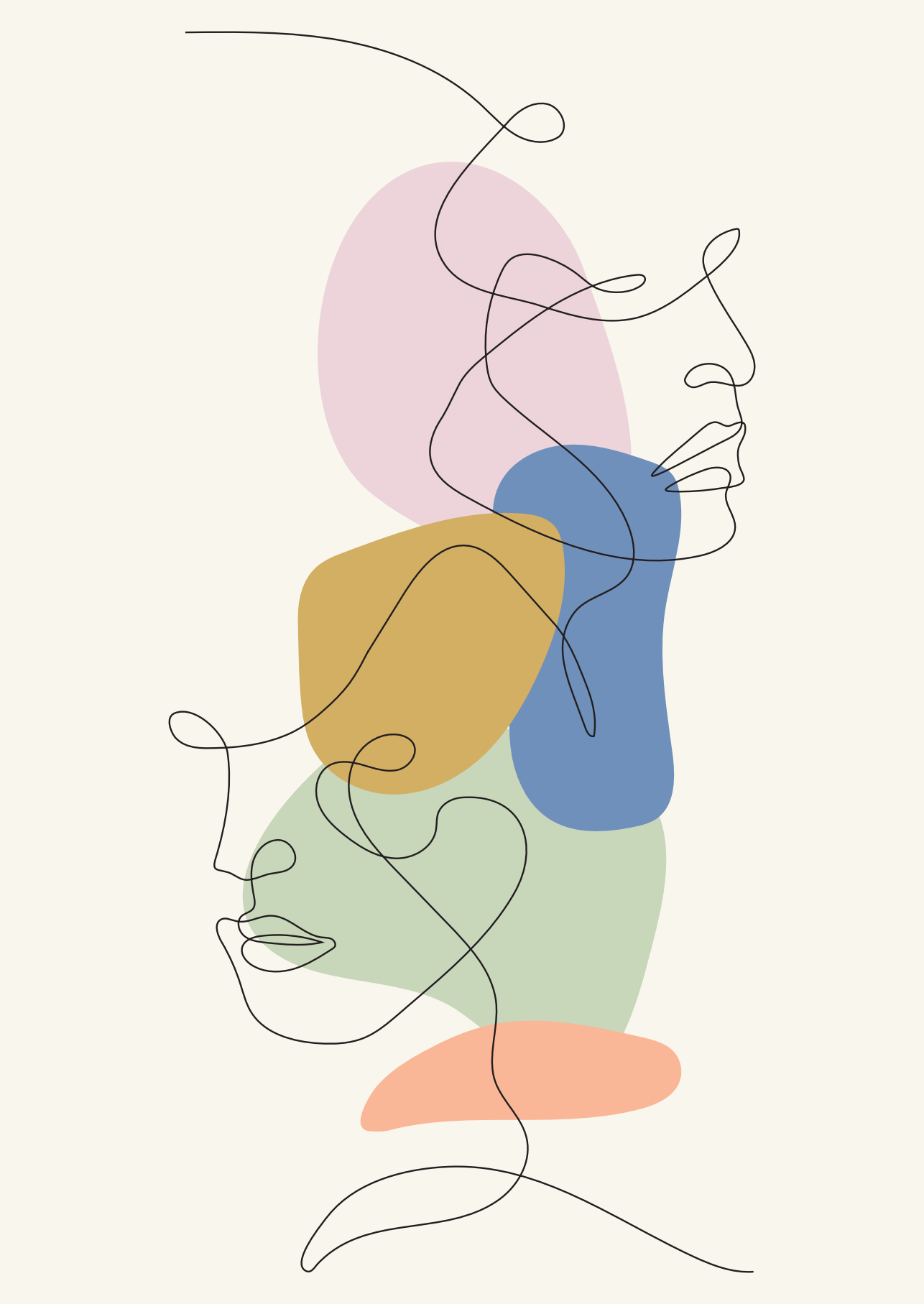
Promotoren: Prof. dr. J.C. Clevers
Prof. dr. L.I. Zon

Voor mama
Omdat ik je elke dag lief heb

En papa
Omdat ik je elke dag mis

Dissertation committee:	Prof. dr. F.C.P. Holstege
	Prof. dr. W.L. de Laat
	Prof. dr. ir. A. van Oudenaarden
	Dr. C. Robin
	Prof. dr. E. van Rooij
	Prof. dr. M. Vermeulen
Paranymphs:	Dr. D.A.C. Heesterbeek
	Drs. T. Bekkering

Contents



1

Acute transcriptional induction in response to stress signals

General introduction

Chapter 1

Acute transcriptional induction in response to stress signals

Audrey Sporrij, Leonard I. Zon

Cell fate and function is regulated by the integrated response to intracellular and extracellular (*i.e.* environmental) signals. Extracellular mediators of cell fate include growth factors and small molecules such as hormones and inflammatory cytokines. However, these signals come in many additional forms. Cells have developed extraordinary mechanisms to sense, transduce and adapt to environmental factors. Rapid adaptation is crucial for maximal cell survival in changing conditions[1]. Ultimately, the adaptive capacity of a cell determines its fate.

Cellular adaptation depends on the type and level of the stressor as well as the cell type affected. One of the main adaptive mechanisms is the rapid and specific modification of gene expression through activation of intracellular signaling pathways[2]. The transcriptional response to extracellular stressors is regulated at many different levels. Here, we discuss our current understanding of stress-induced gene expression dynamics, from chromatin architecture to mRNA synthesis and processing. While adaption to extracellular signals forms one of the main drivers of evolutionary success, it also contributes to diseases such as aging and cancer. Uncovering the exact mechanisms that define transcriptional stress responses to different environmental signals holds novel therapeutic opportunities.

Sensing and signaling of environmental factors

Sensing of environmental signals is one of the most fundamental biological processes. It is implicated in cell growth, proliferation, differentiation, and cell death. To interact with, and survive in, a changing environment, cells need to coordinate their behavior in response to a diverse set of extracellular stressors. This requires timely and accurate sensing of environmental signals. Sensors are responsible for the direct activation of autonomous signaling pathways via membrane-bound or intracellular receptors[3]. Signaling pathways will communicate the environmentally imposed stress to effector molecules in the cytosol and nucleus (Figure 1). Here, stress-specific transactivators will induce responses that form the basis of adaptation[4]. The sensing systems and signal transduction link diverse environmental inputs to rapid and coordinated cellular responses that promote fitness in changing extracellular conditions.

Organisms have evolved a plethora of sensors that are specific to particular stresses. For instance, inflammatory stress is sensed in the form of cytokines. Cytokines includes interleukins, interferons, chemokines, and growth factors such as G-CSF. These molecules are mainly sensed by receptors that converge on the JAK-STAT signal transduction pathway[5] (Figure 1A). Binding of cytokines to their cognate receptors activates receptor-associated Janus kinases (JAKs) that in turn phosphorylate tyrosine residues in the receptor cytoplasmic domain. These phosphorylated residues provide docking sites for members of the signal transducer and activator of transcription (STAT) family of transcription factors (TFs). Upon binding of STATs to the intracellular domain

of the cytokine receptors, JAK-mediated phosphorylation activates STATs which leads to their dissociation from the receptor. Activated STATs dimerize and translocate to the nucleus, where they regulate gene expression[6]. Other signaling mechanisms involve activation of the Nuclear Factor kappa B (NF-κB) pathway[7]. In this cascade, cytokine binding to receptors relieves IκBα-mediated inhibition of NF-κB dimers. The TF dimers subsequently translocate to the nucleus where they initiate transcription of genes controlling cell growth, proliferation and survival.

Immune mediators such as prostanoids employ yet different sensing mechanisms. Prostanoids are bioactive lipid mediators that play pivotal roles in inflammatory responses. They exert their effects in autocrine and paracrine manners via binding to G protein-coupled receptors (GPCRs)[8]. Prostanoid receptors mediate signaling events by induction of phospholipase C (PLC) and the increase of inositol triphosphate (IP3) and diacylglycerol (DAG) (Figure 1B). These signaling intermediates enhance cytosolic calcium levels and regulate protein kinase C (PKC) activity. A second signaling mechanism is linked to the stimulation of adenylyl cyclase and cAMP/protein kinase A (PKA). Transcriptional activation mediated by PKC and PKA is achieved through direct activation of downstream transcription factors, including the cAMP-Response Element Binding protein (CREB), NF-κB and STATs[9,10]. Binding of activated TFs to DNA induces expression of genes required to help reestablish homeostasis in an altered environment.

Steroid hormones (e.g. estrogen, androgen, progesterone) signal by binding to intracellular nuclear receptors. Hormone binding causes conformational changes of the receptor, inducing translocation to the nucleus where they interact with specific sequences of DNA known as hormone response elements (HREs) to regulate transcription[11] (Figure 1C). Alternately, steroid hormones can signal through plasma membrane localized nuclear receptors or hormone-responsive GPCRs, which activate pathways including MAPK/ERK and PI3K/AKT to increase transcription of genes underlying diverse biological processes[12]. Thus, steroid hormones can directly and indirectly control gene expression through activation of nuclear receptors or modification of transcriptional regulators through signaling networks, respectively.

Cellular metabolism is determined by both nutrient availability and exogenous signals from growth factors, cytokines, and hormones. All organisms have the ability to sense nutrients required for energy (catabolism) and macromolecule synthesis (anabolism). Specialized transmembrane proteins serve as metabolite sensors enabling environmental cross-talk and/or uptake of nutrients such as glucose, essential amino acids, and free fatty acids. A key integrator and transducer of these signals is the protein kinase mTOR, which controls conversion of nutrients into macromolecules such as proteins and nucleic acids[13] (Figure 1D). mTOR activates the ribosomal protein S6 kinase (S6K) and inhibits the eIF4E binding protein (4E-BP) to increase translation of metabolic enzymes and metabolic TFs. mTOR impacts transcription of genes involved in cell growth and survival processes through activation of specific TFs. TFs affected by mTOR are the hypoxia-

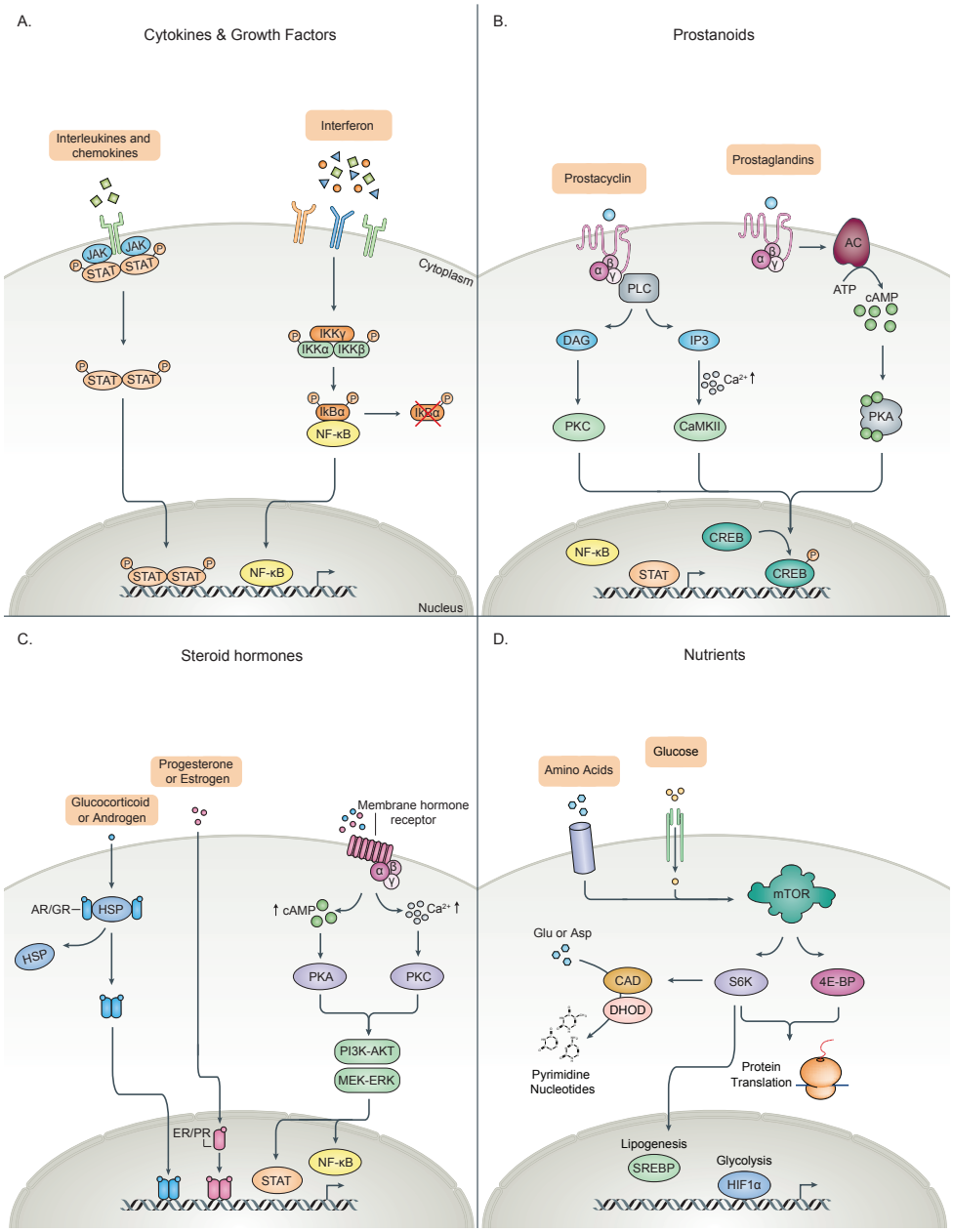


Figure 1. Stress sensing and signaling pathways activate transcription factors. Different environmental stresses, such as inflammatory cytokines, growth factors, hormones, and nutrient availability, can activate signaling pathways. Downstream targets of these pathways include several kinases (PKA, PKC, and CamKII) and transcription factors (CREB, NF-κB, and STATs). Activation of the transcription factors and their binding to stress-responsive genomic regions allows for the recruitment of RNA polymerase II and initiation of transcription.

inducible factors (HIFs) that mediate glycolysis, sterol regulatory element binding protein 1 (SREBP1) which controls lipid synthesis, STATs, and NF-κB, among others[14,15]. In addition, mTOR and S6K directly regulate the activity of numerous metabolic enzymes. This includes de nucleotide synthesis enzyme CAD, for example. Thus, mTOR regulates cellular adaptation to the microenvironment.

Sensing mechanisms emerged early in the history of life and played a fundamental role in the evolutionary success of multicellular organisms. Eukaryotic cells have evolved intricate, and specific, sensors systems to obtain and transmit a variety of environmental signals to the cytosol and nucleus. Signal transduction pathways connect sensors to effector molecules that, in turn, regulate expression of genes involved in adaptation to fluctuating environmental conditions[16].

Transcriptional activation in response to stress

Transcriptional changes play a prominent role in the physiological response to environmental signals. Modulation of gene expression directly underlies the alterations in cell behavior and cell fate that are essential for adaption. Dynamic and differential gene expression is one of the first signs of adaptation to stress[17]. Within minutes after the initial exposure, major changes in the transcriptional pattern of cells are observed. While the exact expression changes depend on the type of environmental signal, general principles for transcription activation upon stress exist.

Many stress sensing and signaling cascades end with the activation of specific signal-responsive transcription factors (STFs) and their binding to DNA recognition sites located within gene regulatory elements. TF dynamics are defined by the localization and affinity of binding sites[18]. The affinity of exposed DNA binding motifs determines the TF abundance threshold that is required for induction of the regulatory region. STFs tend to localize to genomic sites present within inducible promoters and enhancers that are co-occupied by lineage-specific master TFs (MTFs)[19,20]. Many STFs bind DNA as homodimers and can heterodimerize with other TFs. Heterodimerization creates a large combinatorial complexity that facilitates fine regulation of gene expression responses during stress. Upon DNA binding, STFs activated by signaling pathways recruit co-factors that in turn create opportunities for the transcription machinery to engage[21,22]. TF binding partners include chromatin remodeling complexes that alter local DNA accessibility and co-activators that interact with RNA polymerase II (Pol II) and other general transcription factors (GTFs). Hereby, STFs facilitate the initiation or progression of transcription during stress.

Stress responsive genes are highly dynamic to ensure both rapid activation and a quick return to baseline when stimulation ceases. Production of stress-specific gene products is achieved through the coordinated regulation of all aspects of mRNA biosynthesis. TF binding, chromatin restructuring, transcription initiation, elongation, and mRNA maturation

are all aligned to facilitate a controlled transcriptional response within minutes of exposure to stress. This fine regulation of transcription ensures adequate adaptation to stress and maximizes cell survival and fitness.

Chromatin landscape of stress response elements

Gene induction upon stress is not solely defined by the sequence properties of regulatory elements but also by the architecture of the chromatin surrounding TF binding sites. Inducible regulatory regions often have at least one STF motif exposed in the DNA between nucleosomes or contain a partly accessible motif located on the nucleosome edge. Additional accessibility may be required for binding of co-activators and efficient gene induction. Stress responses are concomitant with major changes in chromatin organization necessary for activation of promoters and enhancers[3,23].

Determinants of chromatin accessibility

Physical access to DNA is a key determinant of cell fate. Genomic accessibility is shaped by environmental and developmental signals. Chromatin openness enables the interactions between distant regulatory elements that drive the cooperative regulation of gene expression by promoters and enhancers. Cell-type specific usage of regulatory elements, in combination with lineage-restricted expression of MTFs, controls and maintains cell identity.

Arguably, one of the main determinants of chromatin accessibility is the local nucleosome organization. Nucleosomes are the fundamental structural elements of chromatin. The nucleosome core particle consists of an octamer of positively charged histone proteins (two of each H2A, H2B, H3 and H4) wrapped by approximately 147 base pairs of DNA or 1.7 turns[24] (Figure 2A). Linker histones (H1 variants) bind an additional 20 base pairs of DNA at the entry and exit site of the nucleosome core particle[25]. The nucleosome core complex and linker histone form the chromatosome, which enables folding into higher-order structures. Nucleosome packaging impedes the binding of chromatin factors. Local modification of nucleosome organization alters DNA accessibility and provides opportunities for transcriptional activity.

Chromatin structure at transcribed regions are most likely the result of sequence determinants and chromatin modifiers. The sequence determinants of nucleosome occupancy remain a standing question in chromatin biology. DNA analysis of well-positioned nucleosomes revealed an enrichment for A/T dinucleotides at 10bp intervals. It is suggested that the periodic appearance of A/T creates DNA bending properties that require minimum energy to bend DNA on the nucleosome[26,27]. With limited evidence on the precise DNA sequences dictating nucleosome positions, the currently accepted view attributes precise nucleosome localization to trans-acting proteins such as TFs and

chromatin remodelers. Nucleosome organization is the primary determinant of chromatin accessibility. Access to nucleosomal DNA is facilitated by a complex interplay of nucleosome occupancy, composition, and modifications that is mediated by a variety of chromatin-modifying enzymes (Figure 2B).

Nucleosome occupancy and positioning

Nucleosome dynamics are governed by ATP-dependent remodeling complexes that can move, evict, or modify nucleosomes (Figure 2B). Chromatin remodelers are recruited to regulatory elements during stress responses and play a crucial role in transcriptional induction. Remodelers can be classified into four subfamilies: SWI/SNF (switch/sucrose non-fermentable), ISWI (imitation SWI), CHD (chromodomain helicase DNA-binding), and INO80 (inositol-requiring mutant 80)[28]. All known classes of chromatin remodelers are recruited to specific genomic locations by direct interaction with sequence-specific TFs and can happen independently of RNA polymerase II (Pol II)[29,30]. The relationship between TFs and chromatin state creates a link through which environmental signals alter DNA accessibility to facilitate expression of stress-response genes.

A conserved phenomenon in all eukaryotes is that transcriptionally active promoters consist of a nucleosome-depleted region (NDR) immediately upstream of the transcription start site (TSS) that is flanked by two highly positioned nucleosomes[24]. This local nucleosome depletion ensures that DNA is accessible for the transcriptional machinery. Many inactive, stimuli-inducible promoters show a nucleosome organization similar to active promoters, that is a structure of patterned nucleosomes flanking a NDR. These inactive, inducible promoters are often strongly enriched in promoter-bound, paused Pol II prior to stimulation (Figure 3A). This organization at inducible promoters prior to activation may poise genes for transcriptional induction. The direct accessibility and pre-loading of Pol II allows for rapid activation upon stimulation[31]. Recent studies also revealed that a non-trivial number of stimuli-responsive genes do not present a stable NDR at their TSS prior to stimulation (Figure 3A). Acute activation of transcription in response to stress therefore requires drastic nucleosome reorganization to open these promoters. Nucleosomes within such promoters are transiently displaced by the SWI/SNF remodeling complex[32]. Transient displacement is suggestive of nucleosome re-assembly between cycles of activation and is proposed as a mechanism to achieve highly regulated gene expression.

While nucleosome positioning patterns at promoters are well described, much less is known about the nucleosome organization of enhancers. The prevailing view is that active enhancers are essentially nucleosome-free (Figure 3B). Many inducible enhancers reside in a primed state with a pre-existing NDR prior to activation. However, a large majority of inducible enhancers are nucleosomal during non-stress conditions[33]. Stimulation leads to a shift in nucleosome occupancy away from the enhancer, with nucleosomes

being evicted or repositioned through SWI/SNF-mediated remodeling[34]. The subsequent open chromatin structure correlates with binding of TFs and co-activators. Evidence for the absence of nucleosomes at active enhancers is conflicting. Several studies found TF-bound, active enhancers with a high density of precisely positioned nucleosomes[35-37] (Figure 3B). This indicates that enhancer nucleosomes are not exclusively repressive to gene expression. Removal of the linker histone H1 by the ISWI-remodeling complex NURF and SWI/SNF-mediated H2A/H2B dimer displacement, rather than eviction of the entire nucleosome core particle, achieves an accessible chromatin landscape in these regulatory regions. H1 eviction and/or H2A/H2B displacement have been described at steroid hormone response enhancers and may hold true for a larger number of inducible enhancers[38-40]. It has suggested that binding of TFs to DNA organized in nucleosomes is stabilized via interaction of TF-associated co-factors with histone tails. Retention of the H3/H4 tetramer is in line with a suggested active role for histone particles in facilitating TF

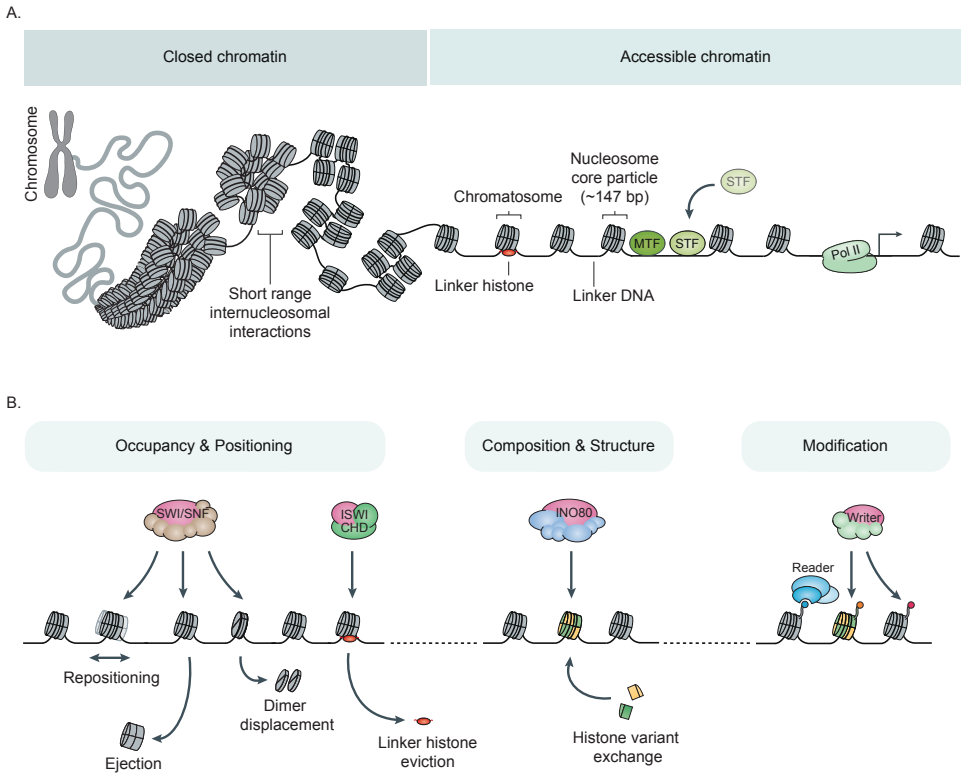


Figure 2. Regulation of chromatin accessibility. (A) DNA compaction within the nucleus occurs through a hierarchy of histone-dependent interactions, including the formation of the nucleosome core particle, the chromatosome and internucleosomal interactions that produces tertiary structures. (B) Chromatin accessibility is controlled by coordinated processes that affect by nucleosome density, composition and histone-tail modifications. ATP-dependent chromatin remodelers can provide access to DNA sequences by altering the nucleosome occupancy and position, or through incorporation of histone variants. Chromatin modifiers (writers, such as histone acetyltransferases) can deposit chemical modifications on histone tails, which affect electrostatic interactions between histones and DNA and directly influence DNA accessibility.

cooperativity and transcriptional induction[41]. The nucleosomes that undergo remodeling are flanked by immobile nucleosomes which themselves are resistant to displacement[42]. This nucleosome organization of enhancers provides robustness against stress-induced chromatin deviations that may otherwise affect the cis-regulatory repertoire and impact cell fate under homeostatic conditions.
The combined changes in nucleosome occupancy and positioning at promoters and enhancers that is observed following acute stress increases chromatin accessibility and correlates directly to gene expression changes.

Nucleosome composition and structure

A cis-regulatory landscape where nucleosomes are retained yet DNA is accessible can, in part, be explained by additional mechanisms that underlie chromatin openness. One such mechanism is the incorporation of histone variants. Histone variants are encoded on separate genes and can replace canonical core histones (H2A, H2B, H3, and H4). In contrast to core histones, histone variants can be incorporated in replication-independent manners. Histone variants can drastically alter nucleosome properties and chromatin packaging. Two histone variants linked to chromatin accessibility and transcription are H2A.Z and H3.3[43] (Figure 2B).
H2A.Z shares only 60% sequence identity with H2A, giving the histone variant unique functional properties. Major structural differences make the H2A.Z-H2B dimer unstable compared to dimers containing H2A. The structural variations also weaken dimer interactions with the H3/H4 tetrasome, leading to H2A.Z-mediated destabilization of the nucleosome core particle. In addition, H2A.Z results in resistance to linker histone H1 binding, thereby inhibiting chromatin condensation[44]. Thus, H2A.Z creates a chromatin structure permissive of co-activator recruitment, hereby facilitating transcriptional activation[45-47]. H2A.Z is incorporated through exchange of a H2A-H2B dimer for a H2A.Z-H2B dimers. This is mediated by SRCAP and p400, members of the INO80 family of chromatin remodelers (Figure 2B). H2A.Z-containing nucleosomes are commonly found flanking the NDR of promoters[48]. Incorporation of H2A.Z at inactive promoters poises genes for active transcription[49,50] (Figure 3A). Direct interactions between H2A.Z and components of the transcriptional machinery facilitate Pol II recruitment while the destabilized nucleosome particle simultaneously promotes Pol II penetration[51]. Similar to promoters, H2A.Z is enriched at active enhancers. At enhancers, H2A.Z increases chromatin accessibility[47], mediates Pol II recruitment and facilitates enhancer-promoter interactions[52]. H2A.Z plays an important role in the transcriptional response to stress. At many stress-responsive genes H2A.Z is deposited at corresponding promoters and enhancers prior to induction, with few changes in promoter H2A.Z levels upon activation[33,53] (Figure 3B). This highlights the role of H2A.Z in shaping the local chromatin structure into states poised for activation. Loss of H2A.Z at promoters or enhancers

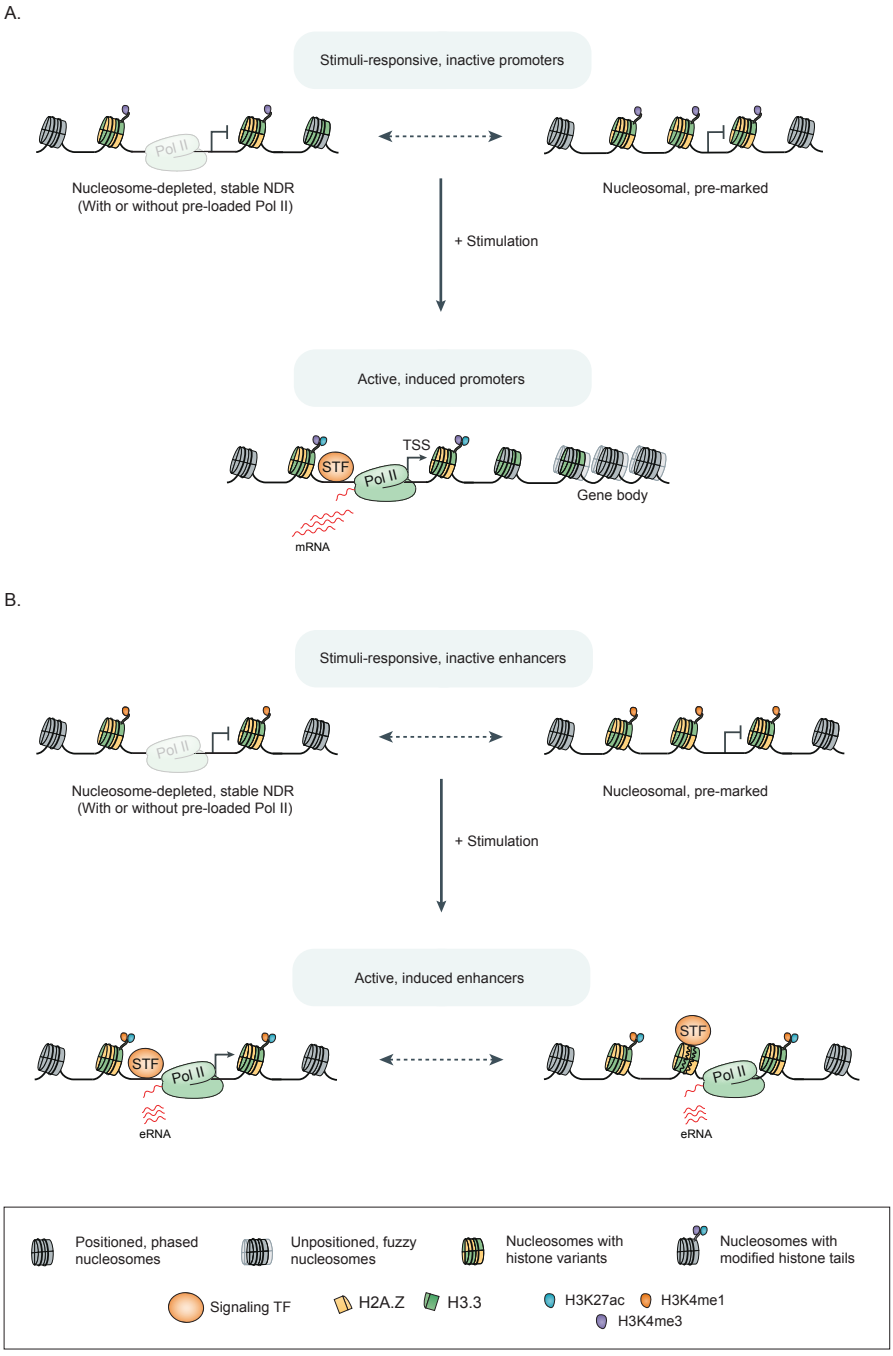


Figure 3. Chromatin organization at stimuli-inducible promoters and enhancers. (A, B) Nucleosome-depleted or nucleosomal chromatin architectures at inducible promoters and enhancers prior to stimulation (A, B: upper panel). A nucleosome-depleted region flanked by two well positioned nucleosomes characterizes active promoters (A: lower panel) while active enhancers have been observed in both nucleosome-depleted and nucleosome-dense conformations (B: lower panel).

diminishes transcription activation of stress-response genes[47,54].

H3.3 is an evolutionarily conserved histone variant that shares nearly 95% sequence identity with canonical histone H3. Despite differing by just 4-5 amino acids from H3, these substituted amino acids provide specific characteristics to histone H3.3[55]. Altered residues localize to a region that contributes to H3-H4 tetramer stability. H3.3-variant nucleosomes are significantly less stable than canonical nucleosomes, independently of H2A.Z incorporation. As a result, nucleosome disruption through loss of H2A-H2B/H2A.Z-H2B dimers and H3-H4 tetramer splitting are associated with H3.3[56,57]. The structural differences not only affect intra-nucleosomal bonds, they also affect higher-order interactions. H3.3 impedes linker histone H1 binding and oligonucleosome formation hereby inhibiting chromatin condensation[58]. Incorporation of H3.3 provides a mechanism to create an open chromatin conformation. H3.3 is deposited at gene regulatory regions by the histone chaperone complex Hira through incorporation of a H3.3-H4 dimer rather than a tetrameric unit[59,60]. H3.3-enriched nucleosomes are commonly found at the promoters, enhancers and gene bodies of active genes[61]. H3.3 is present in the +1 nucleosome at the TSS of both active and repressed genes[62]. In active genes, H3.3 incorporation extends up to the +3 nucleosomes and further into the coding region[63] (Figure 3A). Acute transcriptional induction of stress-response genes results in incorporation of H3.3 at promoters to facilitate binding of TFs, co-factors, and the transcription machinery[64,65]. H3.3-variant nucleosomes also create an open chromatin structure at distal enhancers. H3.3 is significantly enriched at accessible enhancers[63]. Enhancers that control stress-response genes accumulate H3.3 prior to stimulation. At stress enhancers, H3.3 pre-deposition allows for a chromatin structure that is permissive to STFs that translate environmental signals to the genome[64] (Figure 3B).

Nucleosomes containing both H2A.Z and H3.3 are particularly unstable[43,66]. H2A.Z/H3.3-variant nucleosomes may mark sites that traditionally are described as 'nucleosome-free'. These two histone variants appear to be coordinately inserted at promoters and enhancers[67]. Replication-independent incorporation of histone variants plays an important role in transcriptional response to stress.

Histone post-translational modifications

Histone post-translational modifications (PTMs) add to the complexity of chromatin accessibility. Distinct chemical modifications are deposited on histone tails by enzymes known are epigenetic 'writers' (Figure 2B). Histone-modifying enzymes are not sequence specific but rather recruited to particular genomic regions via interactions with TFs, including STFs bound to stress-inducible regions. The modifications function as docking sites for specific 'readers' that bind histones when certain modifications are present. Reader proteins includes histone chaperones and other remodeling enzymes that further alter the local chromatin architecture to enable transcription[68]. Acetylation

and phosphorylation are PTMs that reduce the positive charge on histones, hereby disrupting electrostatic interactions between histone particles and DNA. The modifications counteract chromatin compactions by decreasing nucleosome stability and increasing local DNA accessibility[69,70]. Histone methylation does not alter the charge but rather affects turnover rate and the recruitment of effector molecules that further alter the chromatin structure[71,72].

Especially the N-terminal tails of histone H3 and H4 variants are routinely subjected to modifications. A variety of chromatin modifications characterize active promoter and enhancer regions. A modifications enriched at active promoters is histone H3 lysine 4 trimethylation (H3K4me3)[73,74]. H3K4me3 interacts with the IWSI-chromatin remodeling complex NURF and subunits of transcription initiation and elongation complexes[75]. While H3K4me3 pre-exist at many stress-inducible promoters prior to stimulation, secondary response genes gain H3K4me3 quickly upon stimulation[42,76] (Figure 3A).

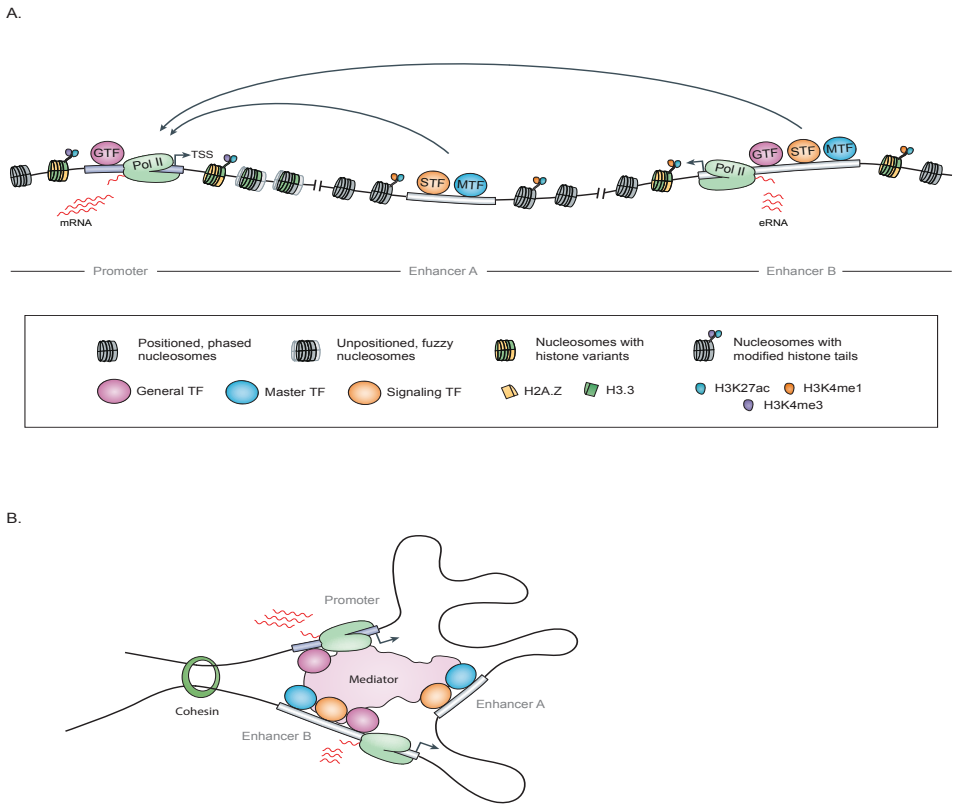


Figure 4. Characteristics of active chromatin structures. (A) Active genomic elements are bound by transcription factors and marked by various epigenetic features. Active promoters are characterized by a depletion of nucleosomes. Nucleosomes that flank active promoters and enhancers contain specific histone variants and histone-tail modifications, such as H3K27ac, H3K4me1, and H3K4me3. The organization of chromatin arranged via activity of nucleosome remodelers, histone chaperones, and modifiers enables transcription at both promoters and enhancers. (B) Distal enhancers contain binding sequences for transcription factors and can upregulate target gene transcription through long-range interaction. Enhancers are brought into proximity of their respective target promoters through looping mediated by cohesion and the mediator complex.

Active enhancers are associated with monomethylation of H3 at lysine 4 (H3K4me1)[77,78]. H3K4me1 is recognized by the BAF complex, a member of the SWI/SNF family of chromatin remodelers. H3K4me1-marked nucleosomes are more efficiently remodeled by BAF than those lacking the modification, revealing a link between histone modifications and the activity of chromatin remodelers[79]. Several stressors allow for the surfacing of so-called latent enhancers that gain both H3K4me1 upon activation. These enhancers can retain residual H3K4me1 when stress ceases, allowing for a faster response during re-exposure. This direct adaptation provides an epigenetic memory to external stressors[80].
Active promoters and enhancers are furthermore characterized by acetylation of H3 lysine 27 (H3K27ac)[81]. H3K27ac not only affects chromatin compaction due to changes in local electrostatic charge but also specifically recruits chromatin factors known as bromodomain proteins[82]. Bromodomain proteins are present in a wide variety of chromatin complexes, including ATP-dependent chromatin remodelers and transcription initiation factors[83]. Histone tails play an essential role in establishing the long-range interactions between promoters and enhancers through internucleosomal electrostatic interactions. Therefore, H3K27ac likely affects promoter-enhancer dynamics[84]. Interactions between Histone Acetyltransferases (HATs) and STFs bound to stress-response regions lead to rapid deposition of H3K27ac at promoters and enhancers (Figure 3A, 3B). Changes in H3K27ac are associated with gene expression changes in response to stress signaling, including inflammatory stimuli[85]. Attenuation of signaling results in rapid loss of enhancer H3K27 acetylation at stress enhancers to ensure inducibility upon re-exposure.
Non-canonical histone variants are subjected to the same modification as their canonical counterparts[86]. Both K27 acetylation and K4 methylation are enriched on H3.3 compared to H3 at active regulatory loci. H2A.Z can be acetylated at lysine 4, 7 and 11. H2A.Z acetylation destabilizes nucleosomes and is associated with chromatin accessibility and gene activation at both promoters and enhancers[87-90]. While H2A.Z acetylation is important for transcriptional induction in response to stress[91], the exact dynamics of H2A.Zac during acute stimulation are currently poorly understood.
The collective modulation of chromatin accessibility through coordinated actions by nucleosome remodelers, histone chaperones, and modifiers are essential for appropriate transcriptional adaptations to stress (Figure 4).

Regulation of mRNA synthesis during stress

Changes in chromatin architecture lead to exposure of sequences that are critical for engagement of Pol II. Activation of Pol II-mediated transcription consist of two phases: transcription initiation and elongation (Figure 5A). At least two mechanisms of stress-induced transcriptional activation are known to exist. First, GTFs bind and position Pol II at the TSS of accessible promoters. This leads to ordered assembly of the pre-initiation complex (PIC) and subsequent transcription. Second, genes are pre-loaded with the

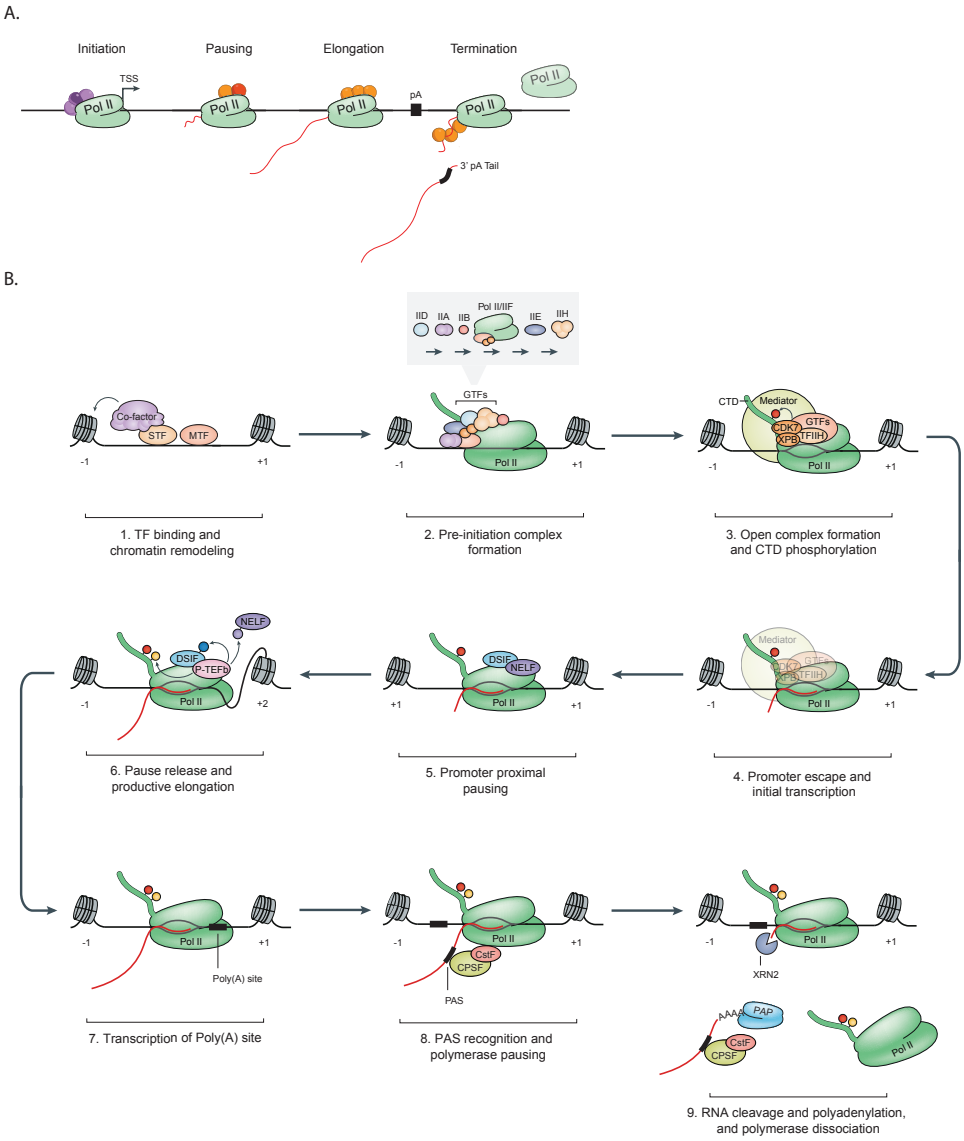


Figure 5. Regulation of transcription initiation, elongation, and termination. (A) Simplified overview RNA polymerase II progressing through different steps of the transcription cycle. (B) Transcription activation starts with the binding of transcription factors at regulatory regions, which then recruit various co-activator complexes. Co-activators include chromatin remodelers that alter the local chromatin structure and factors that directly affect assembly of the preinitiation complex (PIC). The PIC is comprised of Pol II and general transcription factors. CDK7 phosphorylates the carboxy-terminal domain (CTD) of Pol II at Ser5 (in red), which enables Pol II to escape from the promoter and to transition into initial elongation. After promoter escape and synthesis of a short nascent RNA, Pol II is paused by DSIF and NELF. Pause–release is mediated by P- TEFb, which phosphorylates DSIF, NELF and Ser2 (in yellow) of the Pol II CTD. This leads to dissociation of NELF and progression into productive elongation. Termination occurs when transcription progresses through the poly(A) site. Recognition of the transcribed Poly(A) signal by CstF and CPSF initiates RNA cleavage and processing while Pol II dissociates from the DNA.

basic transcriptional machinery during non-stress conditions and stress signaling leads to the release of paused Pol II and transition into productive elongation[92]. Whether regulated at the level of Pol II recruitment or by release of Pol II pausing, stress-response genes must be activated quickly to ensure fast adaptation. In addition, the transcriptional induction is coupled to stress-induced changes in transcription termination, processing and nuclear export that ensure efficient translation of mRNA products into functional proteins. Together, the modification of initiation, elongation, and termination enables a rapid and coordinated response to stress.

Transcription initiation model

After chromatin opening exposes binding sites for the core transcription machinery, the PIC assembles on the promoter. The PIC consists of class II GTFs (TFIID, -A, -B, -E, -F, and -H) and Pol II. The process begins when TATA-box binding protein (TBP), a subunit of TFIID, recognizes the TATA box upstream of the TSS. Subsequently, TFIIA and TFIIB bind the TBP-promoter complex[93]. TFIIB directly interacts with the Pol II-TFIIF complex thus bringing Pol II to the promoter. Finally, TFIIE and TFIIH bind sequentially and complete PIC assembly (Figure 5B). A key function of the PIC is formation of an open complex. DNA opening is carried out by the ATP-dependent DNA translocase XPB, a TFIIH subunit that unwinds DNA and inserts it into the active site of Pol II. A crucial step for transcription to commence is phosphorylation of the carboxy-terminal domain (CTD) of Pol II. CDK7, part of TFIIH, phosphorylates serine (Ser) 5 of the Pol II CTD[94]. This is essential for Pol II to escape the promoter and initiate transcription of the first 20-40 bases of a gene, marking the transition into the early elongation phase[95]

Many promoters lack well described DNA sequence elements such as a TATA-box. TFIID can selectively bind H3K4me3, a PTM highly enriched at active promoters[75]. This suggests that transcription initiation factors recognize the +1 nucleosome surrounding the TSS and highlights the contribution of nucleosomes and histone modifications to gene expression. Stress-activated STF facilitate PIC assembly through interaction with Pol II and TFIIH[96]. Additionally, a wide variety of STF interact with components of mediator[97]. Mediator is an essential large co-factor complex that stabilizes the PIC and stimulates CDK7-mediated CTD phosphorylation[98]. It exerts a structural role by acting as a scaffold through contacts with Pol II, TFIIB, TFIIH and other TFs. Thus, mediator connects the general transcription machinery to STFs and other activators.

Transcription initiation plays an especially important role at secondary stress-response genes, which lack Pol II during basal conditions and require chromatin remodeling at their promoters[32,99]. At these genes, Pol II recruitment is the key regulatory step to transcriptional induction.

STFs do not solely bind promoters but also engage with inducible enhancers. At enhancers, they facilitate recruitment of co-factors such as mediator. Mediator

co-occupies both active promoters and enhancers hereby acting as an important transducer of activating signals between these genomic elements[100]. Thus, enhancers promote PIC assembly at promoters during transcriptional stress responses.
Stress-activated enhancers themselves also display increased levels of transcriptionally engaged Pol II and actively produces enhancer RNAs (eRNAs) upon stimulation[101-103]. Prior to stress, little enrichment of Pol II is found at these enhancers. Rather, the stress enhancers are bound by lineage-specific MTFs. Stress leads to recruitment of Pol II and co-activators at enhancers that display high transcriptional induction. Enhancer transcription can proceed promoter activity, making it one of the earliest events of the transcriptional stress response. Stress signals initiate stable transcription at promoters and enhancers of genes involved in adaptation[104].

Pause release and elongation model

Promoter escape after Ser5 phosphorylation of the Pol II CTD is insufficient for effective transcription of genes. Rather, transcription is paused after the first 20-40 nucleotides. At the promoter proximal pause site, Pol II is halted by the negative elongation factors (NELF) and DRB-sensitivity inducing factor (DSIF), which is composed of SPT4 and SPT5 (Figure 5B). NELF and DSIF impair mobility of the Pol II complex. Productive transcription elongation commences only upon phosphorylation of Ser2 of the Pol II CTD by P-TEFb. Ser2 phosphorylation creates binding sites for additional elongation factors and mRNA processing proteins. P-TEFb, a complex containing CDK9 and Cyclin T1, also phosphorylates NELF and DSIF. The phosphorylation leads to dissociation of NELF and transforms DSIF into a positive elongation factor that associates with Pol II into productive elongation[94].
Many stress-inducible promoters are already bound to Pol II in absence of the stressor and exist in a transcriptionally paused state[105-107]. Pol II stalling has been described at genes involved in hormone and inflammatory responses[108,109]. Upon stress, the loci recruit factors that mediate the regulated release of Pol II pausing. STFs can facilitate the pause-release via direct interaction with P-TEFb, as is the case for NF-κB, or through recruitment of co-factors that bind P-TEFb and promote its activity, such as BRD4 and mediator[110-113]. The rapid switch into productive elongation at target genes bypasses the need for PIC assembly. Specific stressors also inhibit effective elongation of genes required for growth and proliferation[114]. This is accomplished through induction of Hexim, a transcription elongation regulator that forms an inhibitory complex with P-TEFb. Hexim is part of the 7SK small nuclear ribonucleoprotein particle (snRNP) which negatively regulates elongation by sequestering and inactivating P-TEFb[115].
Enhancers affect Pol II pause-release and transcription elongation at their corresponding promoters[116,117]. This is mediated by enhancer-bound TFs and co-activators that can play roles in p-TEFb recruitment and activity. eRNA transcription itself is also regulated

by Pol II pause release[118]. Pol II pauses at a location similar to promoters, that is 20-40 nucleotides downstream of the enhancer TSS. After pause-release, however, Pol II typically transcribes no more than 100 nucleotides from the enhancer TSS. The exact function of the resulting short transcripts remains to be determined, but supporting roles for eRNAs in Pol II binding and transition to elongation have been described (Recently reviewed in ref[119]). Rather than changes in Pol II levels, many enhancers show altered distribution of engaged Pol II during stress. This indicates that eRNA synthesis is due to pause-release rather than Pol II recruitment at specific stress-responsive enhancers[104].
Promoter- and enhancer-mediated regulation of Pol II pausing and elongation allows for finely coordinated transcriptional responses to stress. Polymerase pausing not only allows for accelerated activation but also for synchronization of transcriptional induction through coordinated pause-release[120]. Once Pol II transitions from the paused state into productive elongation, it progresses through the gene body.

Termination and mRNA processing during stress

Transcription termination plays an important role in gene expression by affecting the stability and localization of nascent RNA transcripts. Termination releases the RNA transcript and frees Pol II for reinitiation. The 3′ termination occurs after transcription through the poly(A) site[121]. The Cleavage and Polyadenylation Specificity Factor (CPSF) recognizes the transcribed polyadenylation sequence (PAS; AAUAAA) on the nascent transcript and induces Pol II pausing. Subsequently, Cleavage stimulation factor (CstF) binds to the U/GU-rich region downstream of the PAS and interacts with CPSF (Figure 5B). This stimulates CPSF-mediated cleavage at the nascent RNA right after the PAS and release of Pol II. The cleaved RNA is then polyadenylated by Poly(A) polymerase (PAP) at the 3′ end, which facilitates nuclear export and translation. Following cleavage, the 5'-3' exoribonuclease XRN2 degrades the downstream RNA, which disrupts and releases Pol II from the DNA[122].
Stress signals impact transcriptional termination in several ways. During stress conditions, Pol II transcription can extend kilobases beyond the poly(A) site[123,124]. The transcriptional readthrough is predominantly observed at genes not directly implicated in the stress response. It varies across different stress stimuli, indicating that this is not an arbitrary defect but a deliberate process. In addition, extended transcripts display decreased polyadenylation due to stress-induced inhibition of Poly(A) polymerase activity[125]. mRNA splicing is also altered during stress[126,127]. Widespread inhibition of post-transcriptional splicing leads to the retention of introns. The non-spliced transcripts are not translated into proteins. Long RNAs arising from failed termination and splicing remains in the nucleus[123,128].
mRNAs of genes implicated in the stress response escape from transcriptional readthrough and splicing inhibition. Rather, transcripts of stress genes undergo correct

termination and are effectively spliced through co-transcriptional splicing mechanisms. As a result, there is a transient but selective decrease in the production of growth-related proteins, whereas expression of proteins involved in stress management increases[129,130]. Stress-induced transcriptional readthrough and co-transcriptional splicing provide a mechanism to selectively prioritize translation of mRNA transcripts associated with adequate adaptation.

Current state of the field and knowledge gaps

Rapid adaptation to stress is crucial to maintain cell viability. Cells have developed intricate sensing, signaling, and effector mechanisms to generate the appropriate adaptive response. One of the key strategies for adaptation is alteration of gene expression. Many signaling pathways control a variety of downstream TF and effector molecules that allow for acute transcriptional changes within minutes after exposure to stress.

These gene expression changes are the result of coordinated regulation of chromatin accessibility and mRNA synthesis. Acute stress can lead to genomic binding of TFs and induction of promoters and enhancers regulating stress-response genes. At some stress-inducible regions, the key step is recruitment of transcription machinery and assembly of the PIC. At others, Pol II is already engaged and regulation occurs at the level of transcription elongation. Regulation of Pol II pause-release allows for synchronized activation of genes. Additional coordination is in place at the level of transcription termination, mRNA splicing, and mRNA processing. Transcriptional readthrough and co-transcriptional splicing facilitate prioritized synthesis of proteins essential for the immediate stress response. The ability to regulate different stages of transcription provides the opportunity to tightly control gene expression changes and adapt specifically to different stressors present in the environment.

While our knowledge on regulation of gene expression following stress has advanced tremendously, many questions are left unanswered. Our current understanding of the relationship between specific chromatin structures and different stress-responsive TFs is minimal. Insights into the different functional properties of TFs relevant for both homeostasis and stress are still incomplete. It also remains unclear how regulation of gene expression at the level of transcription initiation versus pausing and elongation is defined, controlled and maintained at different genes. Elucidating the mechanisms that underlie these coordinated processes holds great promise for organisms in both health and disease.

Outline of this thesis

To identify extrinsic regulators of cell fate that offer potential for new clinical therapies, a precise understanding of transcriptional responses to distinct environmental signals is crucial. This thesis aims to decipher the molecular mechanisms of acute transcriptional induction to stress in cancer and stem cells. Our work focuses on the transcriptional control of cells from the two most regenerative systems of the body, those being the skin and the blood.

Extrinsic signals control neural crest cell fate

Neural crest cells (NCCs) are a transient, migratory group of cells that arise from the embryonic ectoderm during vertebrate development[131]. These multipotent progenitor cells give rise to a plethora of cell types. This includes the pigment-producing cells of the skin, also known as melanocytes. Melanocytes are present in the epidermis and hair follicles of the human skin and serve an important role in the protection against ultraviolet light-induced DNA damage. Melanin is one of the most potent free radicals that protects us from mutagenic reactive oxygen species that can otherwise induce structural damage to our DNA[132]. On any given day, we generate millions of new epidermal skin cells and new hairs. Both are pigmented by the transfer of melanin from melanocytes. Therefore, the demand for melanocytes is high and a tight regulation of melanocyte differentiation is required to ensure pigmentation throughout life.

Commitment of NCCs to melanocyte fate occurs via gradual lineage restriction. First, multipotent NCCs become glial-melanocyte lineage restricted cells[133]. These bipotent progenitors become further restricted to the melanocyte lineage as committed, unpigmented melanocyte precursors termed melanoblasts. Melanoblasts can subsequently terminally differentiate into melanized melanocytes[114]. Melanocyte fate commitment is instructed by inductive, extrinsic signals that converge on transcription. Wnt signaling promotes melanocyte lineage commitment over glial cell fate, hereby playing a crucial role in the specification of melanoblasts[134]. Wnt induces expression of microphthalmia-associated transcription factor (MITF), the master regulator of melanocyte identity[135]. MITF acts as a transcriptional activator on several pigment-related genes and is the earliest known marker of NCC commitment to the melanocyte lineage[136-139]. Other environmental signals on which the gradual lineage restriction towards melanoblasts depends include stem cell factor (SCF, also known as KIT ligand) and endothelin B. These factors control the proliferation and survival of melanoblasts.

There is increasing evidence that the transcriptional programs that govern melanocyte differentiation during development also contribute to tumor formation and progression[114,140]. Dysregulation of the expression of genes that regulate and maintain melanocyte fate can transform normal melanocytes into malignant derivatives.

Transformed melanocytes give rise to melanoma, a particularly aggressive and highly metastatic type of cancer with a poor prognosis. Previous studies revealed that hallmark mutations of melanoma, such as mutations in the BRAF oncogene, interact with the transcriptional programs that control NCC fate[114]. The neural crest specifiers that are required for proper development of melanocytes also play a role in the initiation, growth, and dissemination of melanoma[141]. Melanomas are enriched for genes that mark the neural crest, implying that malignant cells have undergone transcriptional reprogramming and adopted an NCC-like fate. The reemergence of this progenitor cell state correlates with melanoma initiation *in vivo*[114]. Thorough understanding of the developmental pathways and environmental signals that control lineage commitment during development and cell fate conversion during tumor initiation aids new therapeutic opportunities. Drugs that disrupt neural crest development during embryogenesis may hold potential as novel treatments for melanoma.

A high degree of conservation in development between vertebrate species makes zebrafish (*Danio rerio*) a valuable model to study melanocyte biology and melanoma[142]. Similar to mice and humans, zebrafish melanocytes originate from migratory NCCs during development and differentiate into specialized melanin-producing pigment cells[143]. Zebrafish melanomas also undergo fate conversion during tumor initiation, with malignant cells adopting a transcriptional signature similar to embryonic NCCs[114]. Previous studies in zebrafish identified the DHODH inhibitor leflunomide as a potent suppressor of the transcriptional programs that regulate NCC fate. DHODH plays an important role in nucleotide synthesis. The enzyme catalyzes the fourth and rate limiting step of de novo pyrimidine (*i.e.* cytosine, thymidine, and uracil) synthesis. Consequently, DHODH inhibition by leflunomide leads to a depletion of nucleotide pools. Low nucleotide levels impede effective transcriptional elongation and lead to defective expression of the genes that govern NCC development and melanoma formation[114,144]. The exact means through which leflunomide inhibits transcriptional elongation are currently unknown. Patients undergoing chemotherapy regimens that target nucleotide metabolism often develop resistance. Insights on the molecular mechanism of leflunomide-mediated transcriptional abrogation can therefore have a direct impact on cancer treatment strategies.

Gene regulation by environmental signals in hematopoietic cells

Cells of the hematopoietic system undergo rapid turnover. Each day, humans require the production of around one hundred billion new blood cells for proper function. Hematopoietic stem cells (HSCs) are rare cells that reside in specialized niches and are required throughout life to produce specific progenitor cells that will replenish all blood lineages. There is, however, an incomplete understanding of the signals and molecular properties that control HSC fate.

As embryonic development progresses, the site and cellular makeup of the HSC niche changes, and thus also the extrinsic signals that affect HSC behavior[145]. Blood development occurs in two waves, known as primitive and definitive hematopoiesis. During the primitive wave of hematopoiesis, blood cells, including erythrocytes necessary for oxygenating rapidly growing tissues, are produced in the yolk sac in mammals. Although this branch of hematopoiesis is transient, yolk sac-derived cells populate the adult stem cell niche and contribute to multilineage hematopoiesis[146-148]. Definitive hematopoiesis begins when HSCs bud off from specialized hemogenic endothelial cells in the aorta-gonad mesonephros (AGM)[149-152]. These HSCs can both self-renew and give rise to more specialized progenitors. As development continues, HSCs sequentially migrate to, and colonize, the placenta, fetal liver, spleen, and finally, the bone marrow[145]. The bone marrow is the adult hematopoietic niche and forms an important extrinsic regulator of HSCs. It not only anchors the stem cells but also ensures an appropriate balance between proliferation, self-renewal, differentiation, and migration of HSCs.

The bone marrow niche is made up of cells from both hematopoietic and non-hematopoietic origin. Key cell types found in this microenvironment are macrophages, stromal cells, and endothelial cells, amongst others. The niche cells produce stimuli in the form of membrane-bound and secreted factors (e.g. growth factors and chemokines) that are important determinants of HSC fate[153-156]. Two niche signals that are crucial for HSC maintenance and retention are SCF and CXC-chemokine ligand (CXCL12, also known as stromal derived factor 1 or SDF1). Niche signals ensure that the majority of HSCs remain in a quiescent state during homeostasis[157,158]. However, HSCs can become rapidly activated to proliferate and differentiate following stress. Soluble signals that are released by niche cells in response to injury and infection, such as interferons, prostaglandins, and G-CSF, were shown to affect HSC behavior. Thus, environmental signals play a central role in the regulation of HSC function.

Given their potential to reconstitute the entire hematopoietic system, HSCs are clinically used for transplantation in patients with a variety of blood and immune disorders[159]. Understanding precisely how extrinsic signals affect HSC fate can lay the foundation for novel cell-based therapies.

In **Chapter 2**, we describe how transcriptional defects from nucleotide stress can be overcome by activation of progesterone signaling in multipotent, migratory neural crest cells. Limited availability of nucleotides, induced by the dihydroorotate dehydrogenase (DHODH) inhibitor leflunomide, leads to near complete abrogation of neural crest development in zebrafish and to a reduction in self-renewal of mammalian neural crest cells. Leflunomide abrogates effective transcription elongation of genes required for neural crest development and melanoma growth *in vivo*. We found that alterations in progesterone and progesterone receptor signaling strongly suppress leflunomide-mediated neural crest effects through bypassing of the elongation block. We identified

the RNA helicase DDX21 to be an interaction partner of the progesterone receptor. DDX21 acts as a sensor of nucleotide stress and holds dual functions in the control of gene expression. In homeostatic conditions, DDX21 binds promoters and facilitates effective transcription elongation. Low nucleotide availability results in loss of DDX21 at chromatin to prevent elongation during nucleotide stress and enforces Pol II pausing. DDX21 binding shifts to mRNA, where it is thought to play a role in transcript stability and processing. This work reveals that progesterone signaling and DDX21 act as important mediators of the nucleotide stress response. Modulating DDX21 levels could be an attractive therapeutic strategy to delay melanoma progression.

In **Chapter 3**, we study the cellular stress response to 16,16-dimethyl-prostaglandin E_2 ($dmPGE_2$), a stable derivative of the inflammatory lipid mediator PGE_2. We assessed the acute metabolic, transcriptional, translational and post-translational effects in primary human $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) and U937 myeloid leukemia cells exposed to $dmPGE_2$. We found acute, transient induction of migration and cell cycle genes, combined with sustained upregulation of factors that prime the bone marrow niche for HSPC colonization, underlie altered homing behavior and proliferation after $dmPGE_2$ stimulation. The acute changes in mRNA levels are directly reflected on protein expression levels as determined by quantitative mass spectrometry-based proteomics. Using phospho-proteome and metabolomic profiling, we furthermore found that transcriptional changes are concomitant with stress-mediated post-translational changes of proteins involved in chromatin structure, which likely facilitate gene induction. While homing and proliferation is mediated on a transcriptional level, enhanced survival of HSPCs after pulse exposure to $dmPGE_2$ is controlled on a post-translational level. Our integrated multiomics approach provides the first comprehensive dissection of acute transcriptome-, proteome- and metabolome-wide changes in HSPCs in response to the inflammatory stressor $dmPGE_2$.

In **Chapter 4**, we perform comprehensive profiling of HSPC *in vivo* during normal homeostasis and during the acute stress responses to extracellular perturbagens. We assessed single-cell transcriptional and chromatin dynamics after *in vivo* stimulation with the inflammatory mediators $dmPGE_2$, Poly I:C, and G-CSF. We observed that during normal homeostasis, HSCs exist in a gradient of transcriptional and epigenetic states rather than in distinct populations. Baseline chromatin state indicates cell autonomous heterogeneity, which likely affects distinct transcriptional responses. Stimulation with stress signals leads to rapid transcriptional changes within HSPC subpopulations. Environmentally-induced changes in the distribution of HSPC states provide evidence that HSPC states are highly dynamic *in vivo*. This research shows that distinct transcriptomic and epigenetic landscapes may underlie heterogeneity in stress responses of HSCs and that this may relate to distinct functional properties.

In **Chapter 5**, To understand the molecular mechanism of stress-induced transcriptional changes, we investigated chromatin reorganization in primary human $CD34^+$ HSPCs after acute stimulation with $dmPGE_2$. Using next generation sequencing (NGS) technologies, we investigated chromatin states prior to and after exposure to stress. We identified genomic loci bound by stress-responsive transcription factors, assessed changes in nucleosomes occupancy and structure, and mapped alterations in chromatin accessibility. We found that $dmPGE_2$ activates the transcription factor CREB, which specifically engages enhancers upon stimulation. $dmPGE_2$-activated enhancers gain chromatin openness while maintaining a nucleosome dense organization. Enhancer nucleosomes are enriched in histone variant H2A.Z and we observe a direct correlation between stress-mediated H2A.Z acetylation and gene induction. CREB binding to H2A.Z-variant nucleosomes may facilitate H2A.Z acetylation through recruitment of the histone-acetylate transferases p300. We propose that H2A.Z acetylation creates accessibility at nucleosome-dense inducible enhancers. This study reveals a mechanism of stress-responsive TF-directed chromatin reorganization at enhancers that facilitates enhancer activation and transcriptional induction.

Finally, in **Chapter 6**, we summarize the work presented in this thesis, provide a critical view of the implications of these findings, and discuss the questions that remain to be answered.

References

1. Badeaux, A. I. & Shi, Y. Emerging roles for chromatin as a signal integration and storage platform. Nature Reviews Molecular Cell Biology 14, 211–224 (2013).
2. Weake, V. M. & Workman, J. L. Inducible gene expression: diverse regulatory mechanisms. Nat. Rev. Genet. 11, 426–437 (2010).
3. Vihervaara, A., Duarte, F. M. & Lis, J. T. Molecular mechanisms driving transcriptional stress responses. Nat. Rev. Genet. 19, 385–397 (2018).
4. de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. Nat. Rev. Genet. 12, 833–845 (2011).
5. Yu, H., Lee, H., Herrmann, A., Buettner, R. & Jove, R. Revisiting STAT3 signalling in cancer: new and unexpected biological functions. Nat Rev Cancer 14, 736–746 (2014).
6. Shuai, K. & Liu, B. Regulation of JAK-STAT signalling in the immune system. Nat Rev Immunol 3, 900–911 (2003).
7. Taniguchi, K. & Karin, M. NF-κB, inflammation, immunity and cancer: coming of age. Nat Rev Immunol 18, 309–324 (2018).
8. Malbon, C. C. G proteins in development. Nature Reviews Molecular Cell Biology 6, 689–701 (2005).
9. Garg, R. et al. Activation of nuclear factor κB (NF-κB) in prostate cancer is mediated by protein kinase C epsilon (PKCepsilon). J. Biol. Chem. 287, 37570–37582 (2012).
10. Gartsbein, M. et al. The role of protein kinase C delta activation and STAT3 Ser727 phosphorylation in insulin-induced keratinocyte proliferation. J. Cell. Sci. 119, 470–481 (2006).
11. Glass, C. K. & Ogawa, S. Combinatorial roles of nuclear receptors in inflammation and immunity. Nat Rev Immunol 6, 44–55 (2006).
12. Levin, E. R. & Hammes, S. R. Nuclear receptors outside the nucleus: extranuclear signalling by steroid receptors. Nature Reviews Molecular Cell Biology 17, 783–797 (2016).
13. Chantranupong, L., Wolfson, R. L. & Sabatini, D. M. Nutrient-sensing mechanisms across evolution. Cell 161, 67–83 (2015).
14. Mossmann, D., Park, S. & Hall, M. N. mTOR signalling and cellular metabolism are mutual determinants in cancer. Nat Rev Cancer 18, 744–757 (2018).
15. Liu, G. Y. & Sabatini, D. M. mTOR at the nexus of nutrition, growth, ageing and disease. Nature Reviews Molecular Cell Biology 21, 183–203 (2020).
16. Sarkar, F. H., Li, Y., Wang, Z. & Kong, D. NF-kappaB signaling pathway and its therapeutic implications in human diseases. Int Rev Immunol 27, 293–319 (2008).
17. Kim, H. D. & O'Shea, E. K. A quantitative model of transcription factor-activated gene expression. Nature Structural & Molecular Biology 15, 1192–1198 (2008).
18. Lam, F. H., Steger, D. J. & O'Shea, E. K. Chromatin decouples promoter threshold from dynamic range. Nature 453, 246–250 (2008).
19. 19. Williams, M. S., Amaral, F. M., Simeoni, F. & Somervaille, T. C. A stress-responsive enhancer induces dynamic drug resistance in acute myeloid leukemia. J. Clin. Invest. 130, 1217–1232 (2020).
20. Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell 147, 577–589 (2011).
21. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461, 186–192 (2009).
22. Cairns, B. R. The logic of chromatin architecture and remodelling at promoters. Nature 461, 193–198 (2009).
23. Himanen, S. V. & Sistonen, L. New insights into transcriptional reprogramming during cellular stress. J. Cell. Sci. 132, jcs238402 (2019).
24. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. 10, 161–172 (2009).
25. Fyodorov, D. V., Zhou, B.-R., Skoultchi, A. I. & Bai, Y. Emerging roles of linker histones in regulating chromatin structure and function. Nature Reviews Molecular Cell Biology 19, 192–206 (2018).
26. Segal, E. et al. A genomic code for nucleosome positioning. Nature 442, 772–778 (2006).

27. Mavrich, T. N. et al. Nucleosome organization in the Drosophila genome. Nature 453, 358–362 (2008).
28. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nature Reviews Molecular Cell Biology 29, 2563 (2017).
29. Yudkovsky, N., Logie, C., Hahn, S. & Peterson, C. L. Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. Genes & Development 13, 2369–2374 (1999).
30. Varga-Weisz, P. ATP-dependent chromatin remodeling factors: nucleosome shufflers with many missions. Oncogene 20, 3076–3085 (2001).
31. Oruba, A., Saccani, S. & van Essen, D. Role of cell-type specific nucleosome positioning in inducible activation of mammalian promoters. Nature Communications 11, 1075–18 (2020).
32. Ramirez-Carrozzi, V. R. et al. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell 138, 114–128 (2009).
33. Johnson, T. A. et al. Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin in vivo. Nucleic Acids Res. 46, 203–214 (2018).
34. Gjidoda, A., Tagore, M., McAndrew, M. J., Woods, A. & Floer, M. Nucleosomes are stably evicted from enhancers but not promoters upon induction of certain pro-inflammatory genes in mouse macrophages. PLoS ONE 9, e93971 (2014).
35. Mieczkowski, J. et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nature Communications 7, 11485 (2016).
36. McPherson, C. E., Shim, E. Y., Friedman, D. S. & Zaret, K. S. An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. Cell 75, 387–398 (1993).
37. Li, Z., Schug, J., Tuteja, G., White, P. & Kaestner, K. H. The nucleosome map of the mammalian liver. Nature Structural & Molecular Biology 18, 742–746 (2011).
38. Ballaré, C. et al. Nucleosome-driven transcription factor binding and gene regulation. Mol. Cell 49, 67–79 (2013).
39. Vicent, G. P. et al. Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. Genes & Development 25, 845–862 (2011).
40. Vicent, G. P. et al. Two chromatin remodeling activities cooperate during activation of hormone responsive promoters. PLoS Genet. 5, e1000567 (2009).
41. Chávez, S. & Beato, M. Nucleosome-mediated synergism between transcription factors on the mouse mammary tumor virus promoter. Proceedings of the National Academy of Sciences 94, 2885–2890 (1997).
42. Comoglio, F. et al. Dissection of acute stimulus-inducible nucleosome remodeling in mammalian cells. Genes & Development 33, 1159–1174 (2019).
43. Jin, C. et al. H3.3/H2A.Z double variant–containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat. Genet. 41, 941–945 (2009).
44. Suto, R. K., Clarkson, M. J., Tremethick, D. J. & Luger, K. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol 7, 1121–1124 (2000).
45. Guillemette, B. et al. Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biol. 3, e384 (2005).
46. Gévry, N. et al. Histone H2A.Z is essential for estrogen receptor signaling. Genes & Development 23, 1522–1533 (2009).
47. Hu, G. et al. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell Stem Cell 12, 180–192 (2013).
48. Raisner, R. M. et al. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell 123, 233–248 (2005).
49. Wan, Y. et al. Role of the histone variant H2A.Z/Htz1p in TBP recruitment, chromatin dynamics, and regulated expression of oleate-responsive genes. Mol. Cell. Biol. 29, 2346–2358 (2009).
50. Li, B. et al. Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. Proceedings of the National Academy of Sciences 102, 18385–18390 (2005).
51. Adam, M., Robert, F., Larochelle, M. & Gaudreau, L. H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol. Cell. Biol. 21,

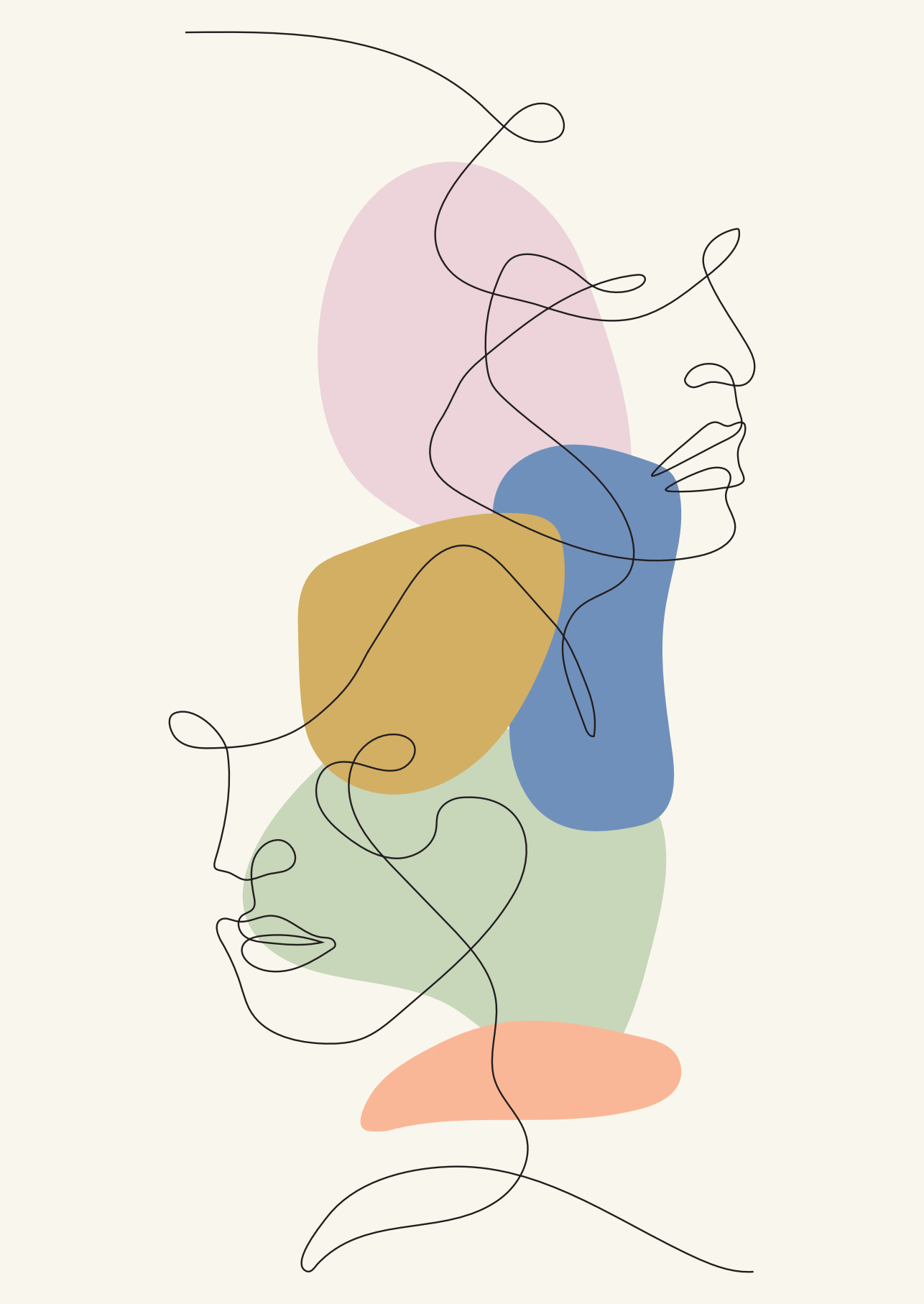
6270–6279 (2001).
52. Brunelle, M. et al. The histone variant H2A.Z is an important regulator of enhancer activity. Nucleic Acids Res. 43, 9742–9756 (2015).
53. Farris, S. D. et al. Transcription-induced chromatin remodeling at the c-myc gene involves the local exchange of histone H2A.Z. J. Biol. Chem. 280, 25298–25303 (2005).
54. Draker, R., Sarcinella, E. & Cheung, P. USP10 deubiquitylates the histone variant H2A.Z and both are required for androgen receptor-mediated gene activation. Nucleic Acids Res. 39, 3529–3542 (2011).
55. Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. Cell Res. 21, 421–434 (2011).
56. Jin, C. & Felsenfeld, G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & Development 21, 1519–1529 (2007).
57. Xu, M. et al. Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. Science 328, 94–98 (2010).
58. Hake, S. B. & Allis, C. D. Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proceedings of the National Academy of Sciences 103, 6428–6435 (2006).
59. Goldberg, A. D. et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140, 678–691 (2010).
60. Tagami, H., Ray-Gallet, D., Almouzni, G. & Nakatani, Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell 116, 51–61 (2004).
61. Ahmad, K. & Henikoff, S. Histone H3 variants specify modes of chromatin assembly. Proceedings of the National Academy of Sciences 99 Suppl 4, 16477–16484 (2002).
62. Schlesinger, S. et al. A hyperdynamic H3.3 nucleosome marks promoter regions in pluripotent embryonic stem cells. Nucleic Acids Res. 45, 12181–12194 (2017).
63. Mito, Y., Henikoff, J. G. & Henikoff, S. Histone replacement marks the boundaries of cis-regulatory domains. Science 315, 1408–1411 (2007).
64. Chen, P. et al. H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. Genes & Development 27, 2109–2124 (2013).
65. Schwartz, B. E. & Ahmad, K. Transcriptional activation triggers deposition and removal of the histone variant H3.3. Genes & Development 19, 804–814 (2005).
66. Nekrasov, M. et al. Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics. Nature Structural & Molecular Biology 19, 1076–1083 (2012).
67. Pradhan, S. K. et al. EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. Mol. Cell 61, 27–38 (2016).
68. Talbert, P. B. & Henikoff, S. Histone variants on the move: substrates for chromatin dynamics. Nature Reviews Molecular Cell Biology 18, 115–126 (2017).
69. Garcia-Ramirez, M., Rocchini, C. & Ausio, J. Modulation of chromatin folding by histone acetylation. J. Biol. Chem. 270, 17923–17928 (1995).
70. Tse, C., Sera, T., Wolffe, A. P. & Hansen, J. C. Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. Mol. Cell. Biol. 18, 4629–4638 (1998).
71. Zee, B. M. et al. In vivo residue-specific histone methylation dynamics. J. Biol. Chem. 285, 3341–3350 (2010).
72. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nature Structural & Molecular Biology 14, 1025–1040 (2007).
73. Bernstein, B. E. et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120, 169–181 (2005).
74. Kim, T. H. et al. A high-resolution map of active promoters in the human genome. Nature 436, 876–880 (2005).
75. Vermeulen, M. et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell 131, 58–69 (2007).
76. Foster, S. L., Hargreaves, D. C. & Medzhitov, R. Gene-specific control of inflammation by TLR-

induced chromatin modifications. Nature 447, 972–978 (2007).
77. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. 39, 311–318 (2007).
78. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283 (2011).
79. Local, A. et al. Identification of H3K4me1-associated proteins at mammalian enhancers. Nat. Genet. 50, 73–82 (2018).
80. Ostuni, R. et al. Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell 152, 157–171 (2013).
81. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences 107, 21931–21936 (2010).
82. Sanchez, R. & Zhou, M.-M. The role of human bromodomains in chromatin biology and gene transcription. Curr Opin Drug Discov Devel 12, 659–665 (2009).
83. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. FEBS Lett 586, 2692–2704 (2012).
84. Kulaeva, O. I. et al. Internucleosomal interactions mediated by histone tails allow distant communication in chromatin. J. Biol. Chem. 287, 20248–20257 (2012).
85. Borghini, L., Hibberd, M. & Davila, S. Changes in H3K27ac following lipopolysaccharide stimulation of nasopharyngeal epithelial cells. BMC Genomics 19, 969–12 (2018).
86. McKittrick, E., Gafken, P. R., Ahmad, K. & Henikoff, S. Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proceedings of the National Academy of Sciences 101, 1525–1530 (2004).
87. Colino-Sanguino, Y. et al. A Read/Write Mechanism Connects p300 Bromodomain Function to H2A.Z Acetylation. iScience 21, 773–788 (2019).
88. Valdés-Mora, F. et al. Acetylated histone variant H2A.Z is involved in the activation of neo-enhancers in prostate cancer. Nature Communications 8, 1346–17 (2017).
89. Valdés-Mora, F. et al. Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. Genome Research 22, 307–321 (2012).
90. Ishibashi, T. et al. Acetylation of vertebrate H2A.Z and its effect on the structure of the nucleosome. Biochemistry 48, 5007–5017 (2009).
91. Halley, J. E., Kaplan, T., Wang, A. Y., Kobor, M. S. & Rine, J. Roles for H2A.Z and its acetylation in GAL1 transcription and gene induction, but not GAL1-transcriptional memory. PLoS Biol. 8, e1000401 (2010).
92. Hargreaves, D. C., Horng, T. & Medzhitov, R. Control of inducible gene expression by signal-dependent transcriptional elongation. Cell 138, 129–145 (2009).
93. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIIH and Mediator. Nature 551, 204–209 (2017).
94. Sims, R. J., Belotserkovskaya, R. & Reinberg, D. Elongation by RNA polymerase II: the short and long of it. Genes & Development 18, 2437–2468 (2004).
95. Rasmussen, E. B. & Lis, J. T. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proceedings of the National Academy of Sciences 90, 7923–7927 (1993).
96. Brzozowski, A. M. et al. Molecular basis of agonism and antagonism in the oestrogen receptor. Nature 389, 753–758 (1997).
97. Sharma, D. & Fondell, J. D. Ordered recruitment of histone acetyltransferases and the TRAP/Mediator complex to thyroid hormone-responsive promoters in vivo. Proceedings of the National Academy of Sciences 99, 7934–7939 (2002).
98. Wang, G. et al. Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. Mol. Cell 17, 683–694 (2005).
99. Serrat, N. et al. The response of secondary genes to lipopolysaccharides in macrophages depends on histone deacetylase and phosphorylation of C/EBPβ. J Immunol 192, 418–426 (2014).
100. Eychenne, T. et al. Functional interplay between Mediator and TFIIB in preinitiation complex assembly in relation to promoter architecture. Genes & Development 30, 2119–2132 (2016).

101. Vihervaara, A. et al. Transcriptional response to stress is pre-wired by promoter and enhancer architecture. Nature Communications 8, 255–16 (2017).
102. Arner, E. et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science 347, 1010–1014 (2015).
103. Li, W. et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 498, 516–520 (2013).
104. Dukler, N. et al. Nascent RNA sequencing reveals a dynamic global transcriptional response at genes and enhancers to the natural medicinal compound celastrol. Genome Research 27, 1816–1829 (2017).
105. Adelman, K. et al. Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. Proc. Natl. Acad. Sci. U.S.A. 106, 18207–18212 (2009).
106. Strobl, L. J. & Eick, D. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. EMBO J. 11, 3307–3314 (1992).
107. Krumm, A., Meulia, T., Brunvand, M. & Groudine, M. The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. Genes & Development 6, 2201–2213 (1992).
108. Aida, M. et al. Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene. Mol. Cell. Biol. 26, 6094–6104 (2006).
109. Kininis, M., Isaacs, G. D., Core, L. J., Hah, N. & Kraus, W. L. Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol. Cell. Biol. 29, 1123–1133 (2009).
110. Bertucci, P. Y. et al. Progesterone receptor induces bcl-x expression through intragenic binding sites favoring RNA polymerase II elongation. Nucleic Acids Res. 41, 6072–6086 (2013).
111. Barboric, M., Nissen, R. M., Kanazawa, S., Jabrane-Ferrat, N. & Peterlin, B. M. NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. Mol. Cell 8, 327–337 (2001).
112. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. Cell 141, 432–445 (2010).
113. Takahashi, H. et al. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell 146, 92–104 (2011).
114. White, R. M. et al. DHODH modulates transcriptional elongation in the neural crest and melanoma. Nature 471, 518–522 (2011).
115. Tan, J. L. et al. Stress from Nucleotide Depletion Activates the Transcriptional Regulator HEXIM1 to Suppress Melanoma. Mol. Cell 62, 34–46 (2016).
116. Lee, J.-E. et al. Brd4 binds to active enhancers to control cell identity gene induction in adipogenesis and myogenesis. Nature Communications 8, 2217–12 (2017).
117. Winter, G. E. et al. BET Bromodomain Proteins Function as Master Transcription Elongation Factors Independent of CDK9 Recruitment. Mol. Cell 67, 5–18.e19 (2017).
118. Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. Genes & Development 32, 26–41 (2018).
119. Sartorelli, V. & Lauberth, S. M. Enhancer RNAs are an important regulatory layer of the epigenome. Nature Structural & Molecular Biology 27, 521–528 (2020).
120. Boettiger, A. N. & Levine, M. Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science 325, 471–473 (2009).
121. Proudfoot, N. J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science 352, aad9926–aad9926 (2016).
122. West, S., Gromak, N. & Proudfoot, N. J. Human 5‘ --> 3’ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432, 522–525 (2004).
123. Vilborg, A. et al. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc. Natl. Acad. Sci. U.S.A. 114, E8362–E8371 (2017).
124. Vilborg, A., Passarelli, M. C., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Widespread Inducible Transcription Downstream of Human Genes. Mol. Cell 59, 449–461 (2015).
125. Di Giammartino, D. C., Shi, Y. & Manley, J. L. PARP1 represses PAP and inhibits polyadenylation during heat shock. Mol. Cell 49, 7–17 (2013).
126. Shalgi, R., Hurt, J. A., Lindquist, S. & Burge, C. B. Widespread inhibition of posttranscriptional

splicing shapes the cellular transcriptome following heat shock. Cell Rep 7, 1362–1370 (2014).
127. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nature Reviews Molecular Cell Biology 18, 637–650 (2017).
128. Cardiello, J. F., Goodrich, J. A. & Kugel, J. F. Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. Mol. Cell. Biol. 38, 4229 (2018).
129. Warringer, J., Hult, M., Regot, S., Posas, F. & Sunnerhagen, P. The HOG pathway dictates the short-term translational response after hyperosmotic shock. Mol Biol Cell 21, 3080–3092 (2010).
130. Uesono, Y. & Toh-E, A. Transient inhibition of translation initiation by osmotic stress. J. Biol. Chem. 277, 13848–13855 (2002).
131. 131. Erickson, C. A. & Reedy, M. V. Neural crest development: the interplay between morphogenesis and cell differentiation. Curr. Top. Dev. Biol. 40, 177–209 (1998).
132. 132. Meredith, P. & Sarna, T. The physical and chemical properties of eumelanin. Pigment Cell Res 19, 572–594 (2006).
133. 133. Dupin, E., Glavieux, C., Vaigot, P. & Le Douarin, N. M. Endothelin 3 induces the reversion of melanocytes to glia through a neural crest-derived glial-melanocytic progenitor. Proceedings of the National Academy of Sciences 97, 7882–7887 (2000).
134. Dorsky, R. I., Moon, R. T. & Raible, D. W. Control of neural crest cell fate by the Wnt signalling pathway. Nature 396, 370–373 (1998).
135. Yasumoto, K.-I. et al. Microphthalmia-associated transcription factor interacts with LEF-1, a mediator of Wnt signaling. EMBO J. 21, 2703–2714 (2002).
136. Yasumoto, K., Yokoyama, K., Shibata, K., Tomita, Y. & Shibahara, S. Microphthalmia-associated transcription factor as a regulator for melanocyte-specific transcription of the human tyrosinase gene. Mol. Cell. Biol. 14, 8058–8070 (1994).
137. Aoki, H. & Moro, O. Involvement of microphthalmia-associated transcription factor (MITF) in expression of human melanocortin-1 receptor (MC1R). Life Sci 71, 2171–2179 (2002).
138. Du, J. & Fisher, D. E. Identification of Aim-1 as the underwhite mouse mutant and its transcriptional regulation by MITF. J. Biol. Chem. 277, 402–406 (2002).
139. Yasumoto, K., Yokoyama, K., Takahashi, K., Tomita, Y. & Shibahara, S. Functional analysis of microphthalmia-associated transcription factor in pigment cell-specific transcription of the human tyrosinase family genes. J. Biol. Chem. 272, 503–509 (1997).
140. Shakhova, O. Neural crest stem cells in melanoma development. Curr Opin Oncol 26, 215–221 (2014).
141. Shakhova, O. et al. Sox10 promotes the formation and maintenance of giant congenital naevi and melanoma. Nature Cell Biology 14, 882–890 (2012).
142. Mort, R. L., Jackson, I. J. & Patton, E. E. The melanocyte lineage in development and disease. Development 142, 620–632 (2015).
143. Zeng, Z., Richardson, J., Verduzco, D., Mitchell, D. L. & Patton, E. E. Zebrafish have a competent p53-dependent nucleotide excision repair pathway to resolve ultraviolet B-induced DNA damage in the skin. Zebrafish 6, 405–415 (2009).
144. Wada, T. et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes & Development 12, 343–356 (1998).
145. Orkin, S. H. & Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell 132, 631–644 (2008).
146. Samokhvalov, I. M., Samokhvalova, N. I. & Nishikawa, S.-I. Cell tracing shows the contribution of the yolk sac to adult haematopoiesis. Nature 446, 1056–1061 (2007).
147. North, T. E. et al. Runx1 Expression Marks Long-Term Repopulating Hematopoietic Stem Cells in the Midgestation Mouse Embryo. Immunity 16, 661–672 (2002).
148. Göthert, J. R. et al. In vivo fate-tracing studies using the Scl stem cell enhancer: embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. Blood 105, 2724–2732 (2005).
149. Sanchez, M. J., Holmes, A., Miles, C. & Dzierzak, E. Characterization of the first definitive hematopoietic stem cells in the AGM and liver of the mouse embryo. Immunity 5, 513–525

(1996).
150. Boisset, J.-C. et al. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. Nature 464, 116–120 (2010).
151. Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. Nature 464, 112–115 (2010).
152. Bertrand, J. Y. et al. Characterization of purified intraembryonic hematopoietic stem cells as a tool to define their site of origin. Proceedings of the National Academy of Sciences 102, 134–139 (2005).
153. Mendelson, A. & Frenette, P. S. Hematopoietic stem cell niche maintenance during homeostasis and regeneration. Nature Medicine 20, 833–846 (2014).
154. Rafii, S., Butler, J. M. & Ding, B.-S. Angiocrine functions of organ-specific endothelial cells. Nature 529, 316–325 (2016).
155. Kobayashi, H. et al. Angiocrine factors from Akt-activated endothelial cells balance self-renewal and differentiation of haematopoietic stem cells. Nature Cell Biology 12, 1046–1056 (2010).
156. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. Nature 481, 457–462 (2012).
157. Pinho, S. & Frenette, P. S. Haematopoietic stem cell activity and interactions with the niche. Nature Reviews Molecular Cell Biology 20, 303–320 (2019).
158. Crane, G. M., Jeffery, E. & Morrison, S. J. Adult haematopoietic stem cell niches. Nat Rev Immunol 17, 573–590 (2017).
159. Jenq, R. R. & van den Brink, M. R. M. Allogeneic haematopoietic stem cell transplantation: individualized stem cell and immune therapy of cancer. Nat Rev Cancer 10, 213–221 (2010).



2

RNA helicase DDX21 mediates nucleotide stress responses in neural crest and melanoma cells

Published in Nature Cell Biology, Vol 22, 372-379 (2020)

Chapter 2

RNA helicase DDX21 mediates nucleotide stress responses in neural crest and melanoma cells

Cristina Santoriello*, Audrey Sporrij*, Song Yang, Ryan A. Flynn, Telmo Henriques, Bilguujin Dorjsuren, Eugenia Custo Greig, Wyatt McCall, Meredith E. Stanhope, Maurizio Fazio, Michael Superdock, Asher Lichtig, Isaac Adatto, Brian J. Abraham, Marian Kalocsay, Michael Jurynec, Yi Zhou, Karen Adelman, Eliezer Calo, Leonard I. Zon#

* These authors contributed equally to this study
# Correspondence to: leonard.zon@enders.tch.harvard.edu

This work is published as the following research article:
Santoriello, C.*, Sporrij, A.*, Yang, S. et al. RNA helicase DDX21 mediates nucleotide stress responses in neural crest and melanoma cells. Nat Cell Biol 22, 372–379 (2020) https://doi.org/10.1038/s41556-020-0493-0

Abstract

The availability of nucleotides has a direct impact on transcription. The inhibition of dihydroorotate dehydrogenase (DHODH) with leflunomide impacts nucleotide pools by reducing pyrimidine levels. Leflunomide abrogates effective transcription elongation of genes required for neural crest development and melanoma growth *in vivo*[1]. To define the mechanism of action, we undertook an *in vivo* chemical suppressor screen for restoration of neural crest after leflunomide treatment. Surprisingly, we found that alterations in progesterone and progesterone receptor (Pgr) signaling strongly suppressed leflunomide-mediated neural crest defects in zebrafish. In addition, progesterone bypasses the transcriptional elongation block resulting from Paf complex deficiency, rescuing neural crest defects in *ctr9* morphant and *paf1(aln^{z24})* mutant embryos. Using proteomics, we found that Pgr binds the RNA helicase protein Ddx21. *ddx21*-deficient zebrafish show resistance to leflunomide-induced stress. At a molecular level, nucleotide depletion reduced the chromatin occupancy of DDX21 in human A375 melanoma cells. Nucleotide supplementation reversed the gene expression signature and DDX21 occupancy changes prompted by leflunomide. Together, our results show that DDX21 acts as a sensor and mediator of transcription during nucleotide stress.

Introduction

The maintenance of adequate nucleotide pools is necessary for many biological processes[2]. Changes in nucleotide metabolism can cause cellular transformation and tumorigenesis[3,4]. The inhibition of DHODH, an enzyme involved in *de novo* nucleotide biosynthesis, lowers pyrimidine levels and blocks transcription elongation in neural crest and melanoma cells[1]. Transcription elongation is highly regulated following the recruitment of RNA polymerase II (Pol II) to the transcriptional start site (TSS) and establishment of promoter-proximal pausing[5,6]. Positive transcription elongation factor b (p-TEFb), composed of cyclin-dependent kinase 9 (CDK9) and cyclin T1 (CCNT1), is able to relieve pausing and stimulate productive elongation[7,8]. HEXIM1, a component of the 7SK small nuclear ribonuclear protein (7SK snRNP), sequesters p-TEFb to enforce transcriptional pausing under nucleotide stress[9]. Despite these insights, the mechanisms through which nucleotide depletion is sensed are unknown.

Results

Progesterone confers resistance to nucleotide depletion in vivo

To elucidate how a block in pyrimidine biosynthesis affects transcription elongation, we performed a chemical suppressor screen for the DHODH inhibitor leflunomide in zebrafish *(Danio rerio).* We treated embryos with leflunomide and screened 2140 chemicals using whole mount in situ hybridization (WISH) for the neural crest marker *crestin*[10] (Figure 1A). We identified 30 compounds that were able to rescue the neural crest lineage following leflunomide treatment (Supplementary Table 1). To ensure specificity, we tested these chemicals in combination with A77 1726 (Terileflunomide), the active metabolite of leflunomide, and 2-[(3,5-difluoro-3'-methoxy-4-biphenylyl)amino]nicotinic acid, a DHODH inhibitor with a distinct chemical structure hereafter named iDH1 (Supplemental Figure 1A). We found that 13 chemicals were able to rescue *crestin* expression independent of the DHODH inhibitor used (Supplementary Table 1).

We performed metabolite profiling by mass spectrometry on A375 melanoma cells to evaluate nucleotide and precursor levels 24 hours post treatment. Analysis revealed that the 13 chemical hits fall into three distinct classes. The first class affects the biochemical activity of leflunomide, while the second class affects nucleotide pools. Chemicals of the third class bypass transcriptional defects through independent mechanisms (Figure 1B, 1C, Supplemental Figure 1B, 1C). Within this third group is the steroid hormone progesterone. Progesterone, an activator of Pgr, recued expression of the neural crest markers *crestin, sox10, pax3* and *foxd3* (Figure 1B, Supplemental Figure 2A). We then tested whether progesterone also rescues neural crest defects resulting from Paf complex deficiency. The Paf complex plays a crucial role in transcription elongation[11-13]. We found

that progesterone circumvents aberrant transcription in *ctr9*[12] morphants and *paf1(aln^{z24})*[14] mutants (Figure 1D). This establishes that progesterone rescues elongation defects consequent to leflunomide or Paf complex deficiency.

To understand the developmental aspects of rescue, we added progesterone at varying times after leflunomide treatment. We found that 12 to 15 somites is the earliest stage to rescue the neural crest (Supplemental Figure 2D). We performed RNA sequencing (RNA-Seq) on embryos treated with leflunomide, progesterone or both compounds at 8 and 15 somites. This captured transcriptional changes before onset and at emergence of the phenotype, respectively. Although DHODH inhibition is expected to lead to ubiquitous effects, different cell types may have varying sensitivities to inhibition depending on their ability to engage alternative mechanisms for nucleotide production, such as salvage pathways. Analysis of differentially expressed genes (DEGs) at 15 somites revealed that several lineages were likely affected by leflunomide, including muscle and skeletal systems (Supplementary Table 2). We additionally observed specific enrichment for neural crest-derived lineages. Mutations in DHODH cause Miller syndrome, a human craniofacial disorder[15]. This highlights the susceptibility of neural crest to DHODH deficiency during development. Together, the data illustrate the tissue specific nature of gene expression defects with the neural crest being prominently affected.

We observed little overlap (7%) in DEGs at 8 and 15 somites (Supplementary Table 2). We suspected neural crest effects to be obscured through bulk transcriptomic profiling of heterogeneous developing embryos. In order to resolve specific changes, we sorted *sox10^{+}* neural crest cells at 15 somites from transgenic embryos treated with leflunomide, progesterone or both compounds (Supplemental Figure 3A, 3B). Progesterone restores the gene signature imposed by leflunomide to near completion, rescuing over 83% of DEGs, including *foxd3* and *pax3* (Supplemental Figure 3C, Supplementary Table 3). We observed a similar (72%) transcriptional reversion in *ctr9* morphant *sox10^{+}* cells (Supplemental Figure 3D, Supplementary Table 4). Genes not rescued showed enrichment in developmental signaling pathways. We cannot exclude that these genes may require additional signals or time to restore their expression. Nonetheless, the data define progesterone as a strong chemical suppressor of leflunomide-mediated neural crest defects and support a role for nuclear hormone signaling in overcoming transcriptional defects.

We verified *pgr* expression during early embryonic development by RT-PCR (Supplemental Figure 4A). To test whether the effect of progesterone is mediated via its receptor, we abolished *pgr* expression using an ATG morpholino (MO). Embryos were treated with leflunomide or leflunomide plus progesterone, and *crestin* expression was assessed. We found that loss of *pgr* did not affect progesterone-mediated rescue. Rather, loss of *pgr* by itself was sufficient to rescue *crestin* following leflunomide treatment (Figure 2A). We ensured MO specificity through co-injection with wild-type (wt) or mismatched (7mm) *pgr* mRNA (Supplemental Figure 4B). In line with the previous observation, *crestin*

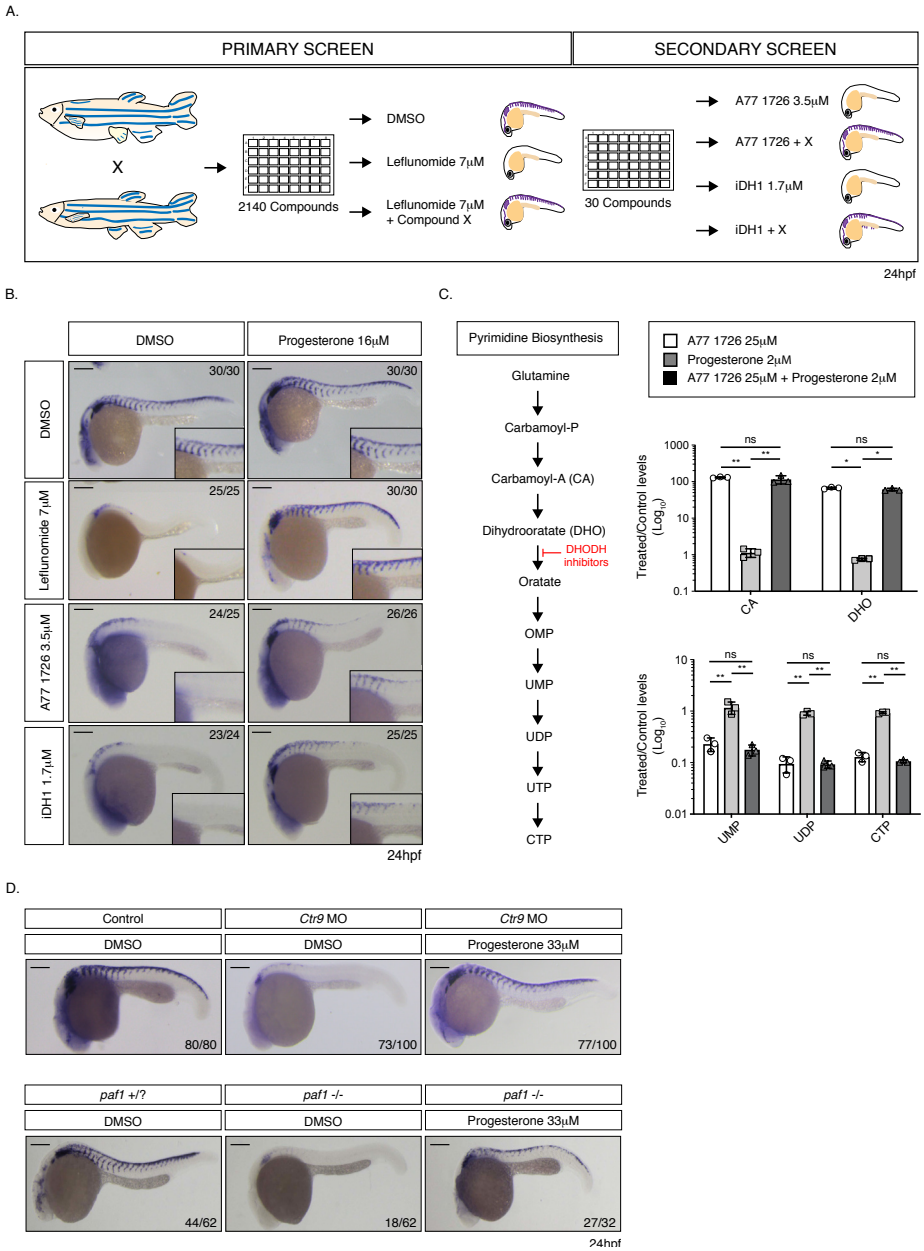


Figure 1. Progesterone confers resistance to nucleotide depletion *in vivo*. (A) Schematic overview of chemical suppressor screen performed in zebrafish embryos with the DHODH inhibitors leflunomide, A77 1726, and iDHODH1 (iDH1). *Crestin* expression was assessed at 24 hours post fertilization (hpf). (B) *In situ* hybridization for *crestin* at 24hpf in embryos treated with the indicated DHODH inhibitors plus or minus progesterone. Number of embryos displaying the represented phenotype is noted. (C) The pyrimidine biosynthesis pathway highlighting the enzymatic conversion inhibited by the leflunomide, A77 1726 and iDH1. Metabolite profiling of A375 melanoma cells exposed to indicated chemicals for 24 hours. (n = 3 biologically independent experiments, Mean ± SD, Two-Way ANOVA with Bonferroni Comparison, * = $p < 0.001$, ** = $p < 0.0001$, ns = $p > 0.05$). (D) *In situ* hybridization for *crestin* at 24hpf in *ctr9* morpholine (MO) injected or *paf* mutant zebrafish embryos exposed to DMSO or progesterone. Number of embryos displaying the presented phenotype is indicated. Scale bars represent 200μm.

rescue was reduced by co-injection of the *pgr* MO with mismatched *pgr* mRNA (Figure 2B). These data demonstrate that knockdown of *pgr* counteracts transcriptional defects arising from nucleotide depletion.

As *pgr* knockdown rescues neural crest, we hypothesized whether the progesterone dosage used in our studies decreased receptor expression through self-regulatory mechanisms. Effects of progesterone on PGR expression have been previously described in cancer[16,17]. We observed that progesterone reduced *pgr* levels by almost 2-fold in zebrafish embryos (Figure 2C). These findings may explain why both exposure to progesterone and loss of its receptor yield the same biological effect *in vivo.* We also found that leflunomide increased *pgr* expression, which prompted us to assess whether *pgr* overexpression can mimic the effects of leflunomide. Overexpression of *pgr* was not sufficient to abrogate *crestin* expression (Supplemental Figure 4C). This indicates that additional mechanisms are in play to establish loss of neural crest lineage through leflunomide. Together, the data support a function for Pgr during nucleotide stress.

The progesterone receptor interacts with RNA helicase DDX21

To clarify the role of Pgr, we conducted co-immunoprecipitation (Co-IP) followed by mass spectrometry to identify complex interaction partners of the receptor. We injected embryos with *pgr:flag:T2A:gfp* mRNA and exposed them to DMSO, progesterone, or leflunomide plus progesterone prior to Co-IP at the 15 somite stage. Amongst the top peptides identified was Hsp90, a known interactor of Pgr[18] (Figure 2D, Supplementary Table 5). We found associations between Pgr and proteins from two major biological processes. First, Pgr associates with nucleotide metabolism enzymes such as Atic and Impdh2. This highlights a function for Pgr in the regulation of nucleotide levels. We also identified proteins related to RNA metabolism as potential complex partners of Pgr, including Ddx21. Ddx21 is an RNA helicase with reported functions in transcription and ribosomal RNA (rRNA) metabolism[19-21]. We verified the association between DDX21 and PGR in human A375 melanoma cells with inducible PGR expression and in T47D cells, a human breast cancer cell line where both proteins are expressed naturally (Figure 2E, Supplemental Figure 5A, 5B). This revealed an association between PGR and DDX21 in several cellular contexts.

Due to its regulating function in Pol I and Pol II transcription[19-21], we studied the role of Ddx21 *in vivo.* We confirmed *ddx21* expression during early development by RT-PCR (Supplemental Figure 5C). We then knocked down *ddx21* using an ATG MO and exposed embryos to DMSO, leflunomide or iDH1. We observed that loss of *ddx21* rescued the neural crest following leflunomide or iDH1 treatment (Figure 3A, 3B). Co-injection of the MO with wild-type *ddx21* mRNA reduced rescue effects and confirmed MO specificity. Additionally, we assessed whether *crestin* rescue through loss of function recovers *in vivo* melanocyte differentiation potential. Pigmentation defects characterizing leflunomide-

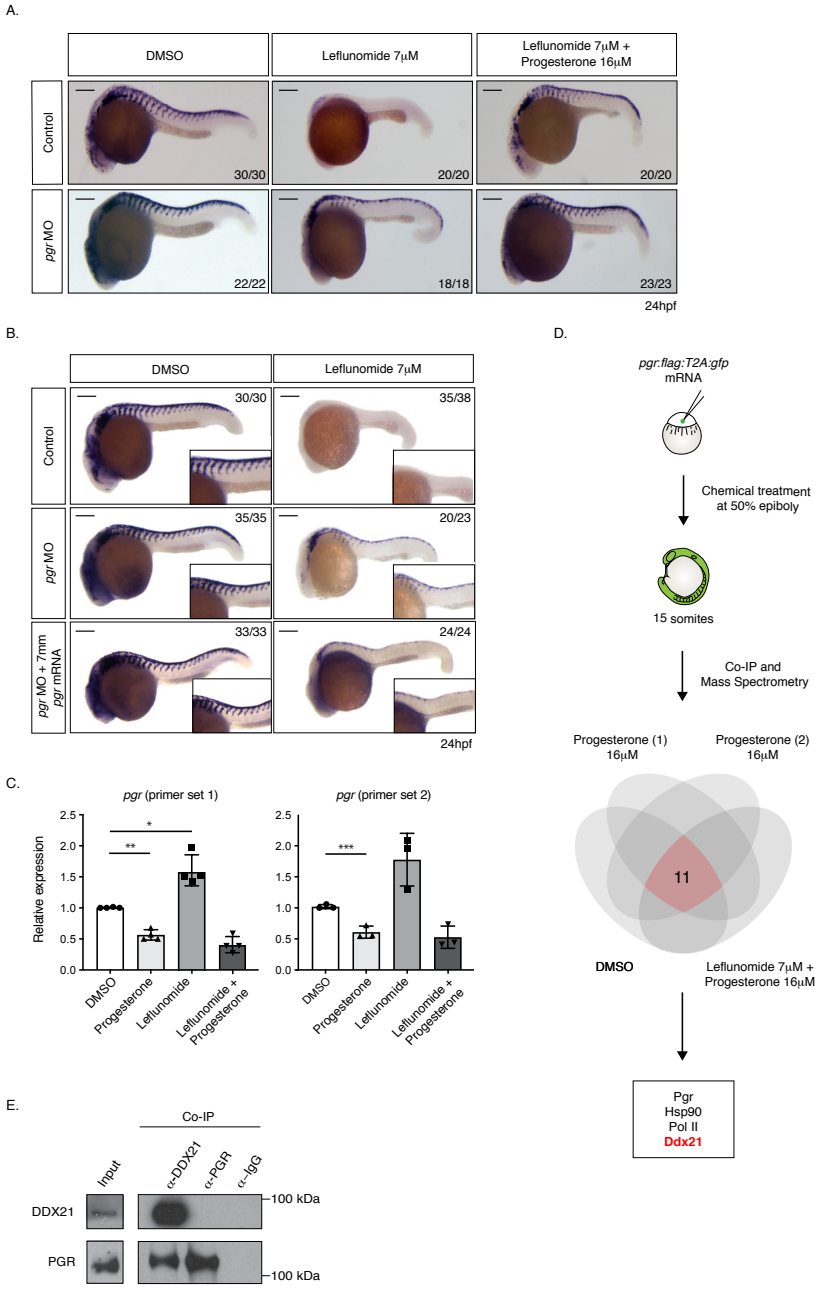


Figure 2. The progesterone receptor interacts with RNA helicase DDX21. (A, B) *In situ* hybridization at 24hpf for *crestin* in control, *pgr* morpholine (MO) and *pgr* mRNA injected embryos exposed to leflunomide in the presence and absence of progesterone. Number of embryos displaying the presented phenotype is indicated. Scale bars represent 200μm. (C) qPCR for *pgr* on zebrafish embryos at 24hpf using two different primer sets. (n ≥ 3 biologically independent experiments, Mean ± SD, Unpaired T-test with Welch Correction, * p = 0.017, ** p = 0.002, *** p = 0.021). (D) Co-Immunoprecipitation followed by mass-spectrometry of *pgr:flag:T2A:gfp* injected embryos identified 11 complex association partners of Pgr *in vivo*. (E) DDX21 associates with PGR in A375 melanoma cells (2 biologically independent experiments).

treated embryos were overcome by *ddx21* knockdown (Figure 3C). This established that knockdown of *ddx21* confers resistance to nucleotide depletion.

Nucleotide stress induces a shift in DDX21 binding from DNA to RNA

Ribosomal biosynthesis encompasses a significant demand for nucleotides. The majority of nucleotides reside in the rRNA of ribosomes[22]. Given the dual function of DDX21 in regulating mRNA transcription and rRNA transcription, processing and modifications[23], we hypothesized that DDX21 responds to changes in nucleotide levels. To determine whether nucleotide depletion affects DDX21 binding to rRNA, we performed infra-red crosslinking

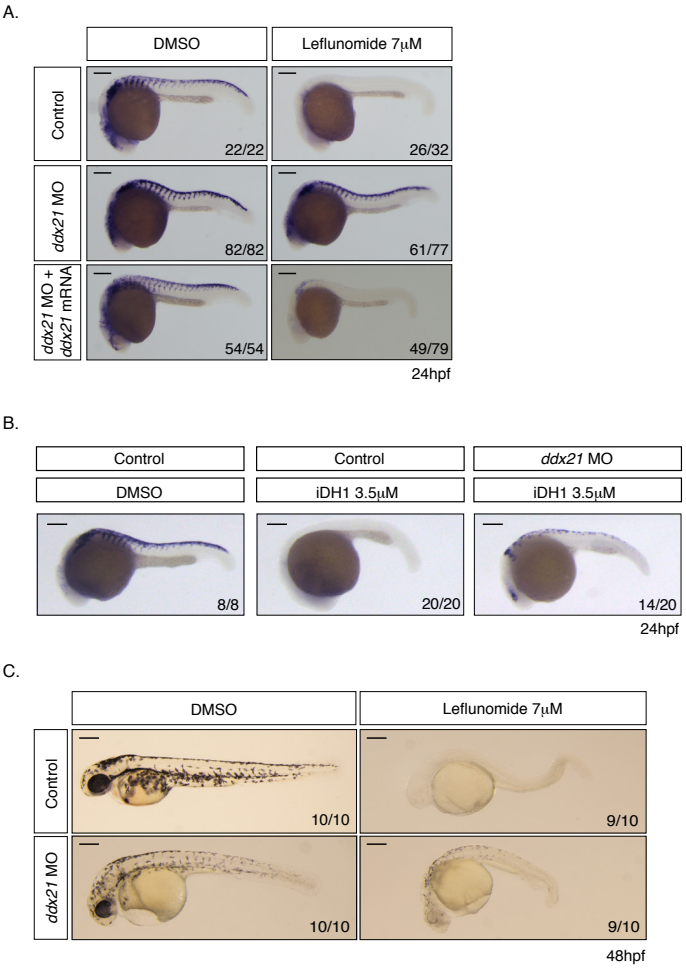


Figure 3. Loss of Ddx21 rescues neural crest defects under nucleotide depletion. (A) *In situ* hybridization at 24hpf for *crestin* in control, *ddx21* morpholine (MO), and MO plus *ddx21* mRNA injected embryos exposed to leflunomide. (B) *In situ* hybridization at 24hpf for *crestin* in control, *ddx21* morpholine (MO), and *ddx21* MO injected embryos exposed to iDH1. (C) Pigmentation of zebrafish embryos at 48hpf. Number of embryos displaying the presented phenotype is indicated in lower right corner. Scale bars represent 200μm (A, B) or 250μm (C).

and immunoprecipitation (irCLIP) for DDX21 in A375 cells after A77 1726 treatment. Remarkably, we observed that DDX21 binding to rRNA was reduced by nucleotide stress (Figure 4A). Simultaneously, there was a notable increase in binding of DDX21 to mRNA, including 469 new bound transcripts (Figure 4B, Supplementary Table 6). New DDX21-bound transcripts include mRNAs of genes involved in pyrimidine metabolism and the cell cycle such as CAD and CDKN1A, respectively (Figure 4C). The shift in DDX21 interaction from rRNA to mRNA is reversed by nucleotide supplementation, indicating that this effect is specific to a reduction of nucleotide pools. We hypothesized that binding of DDX21 to mRNA affects corresponding protein levels. We therefore performed proteomics analysis on A77 1726 treated A375 cells (Supplementary Table 7). We found no significant correlation between DDX21-bound mRNAs and their respective protein expression levels. The consequence of DDX21 binding to mRNAs is currently unknown. We do not exclude a role for DDX21 in mRNA processing during nucleotide stress. We conclude that nucleotide depletion leads to altered binding of DDX21 to mRNAs, but these changes are

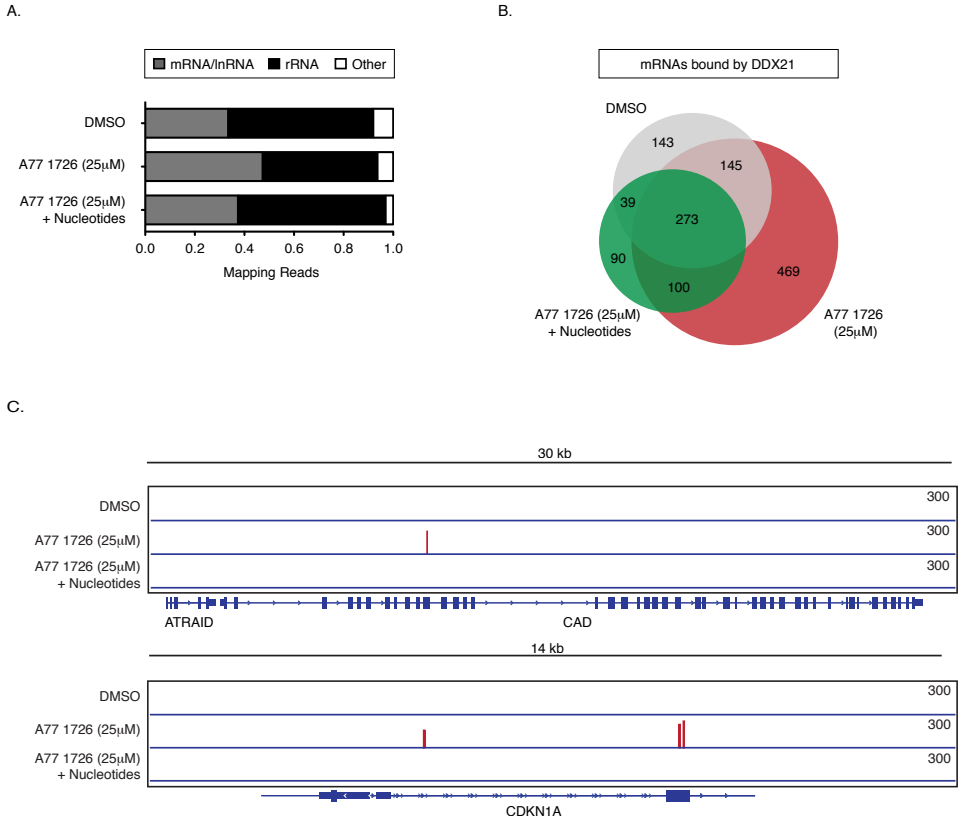


Figure 4. Nucleotide stress alters binding of DDX21 to RNA. (A) Annotation of DDX21-associated RNAs identified by irCLIP in A375 melanoma cells after DHODH inhibition with or without nucleotide supplementation. (B) Venn diagram comparing overlap in mRNA binding targets of DDX21 under different conditions. (C) Gene tracks showing irCLIP signal of DDX21 at the CAD and CDKN1A locus during nucleotide depletion.

not reflected at protein levels of targeted transcripts.
Under homeostatic conditions, DDX21 is localized to the nucleolus where it directly binds rRNAs and small nucleolarRNAs (snoRNAs)[19,24]. Following transcriptional stress, DDX21 relocalizes to the nucleoplasm[19,20]. Based on our observation that DDX21 engages mRNAs when nucleotide levels are low, we investigated whether DHODH inhibition affects DDX21 localization. We performed immunofluorescence (IF) staining for DDX21 in A375 cells 24 hours after A77 1726 treatment, the timepoint when nucleotide depletion is first observed[9]. We observed partial relocalization of DDX21 from the nucleolus to the nucleoplasm, which was reversed by nucleotide supplementation. No changes in total DDX21 protein levels were noted (Supplemental Figure 6A-6C). TCOF1, an unrelated nucleolar protein, did not display changes in localization upon nucleotide stress (Supplemental Figure 6D, 6E). These observations revealed that DDX21 can interpret nucleotide levels and relocalize to the nucleoplasm when nucleotide pools are limited.

DDX21 mediates transcriptional changes during nucleotide stress

Given that changes in DDX21 localization correlate with its chromatin occupancy[19], we investigated whether nucleotide stress affects DDX21 binding to chromatin. We performed chromatin immunoprecipitation sequencing (ChIP-Seq) for DDX21 and the CDK9 subunit of p-TEFb in A375 cells (Supplemental Figure 7A). Cells exposed to A77 1726 showed substantially reduced genomic occupancy of DDX21 and CDK9 compared to control cells. Remarkably, chromatin binding was restored following nucleotide addition (Figure 5A). Due to suggestions that DDX21 senses the transcriptional status of Pol I[21,25], we examined DDX21 occupancy in the 14kb rDNA repeat region. The data show that the genomic distribution of DDX21 along the rDNA genes for Pol I-directed transcripts is not affected by limited nucleotide pools (Supplemental Figure 7B). However, we previously observed reduced binding of DDX21 to rRNA during nucleotide stress (Figure 4A).
Taken together, these observations suggest that the bulk effect of DDX21 in the nucleolus is reduced, as well as the efficiency of ribosomal biogenesis. These data demonstrate that nucleotide depletion predominantly affects the activity of DDX21 on Pol II-mediated transcription and target genes.
To investigate whether reduced genomic occupancy of DDX21 correlates with gene expression changes, we performed RNA-Seq 24 and 48 hours post A77 1726 treatment in A375 cells. We found that majority of transcriptional changes occur 48 hours after DHODH inhibition. More specifically, 32% of the genome was at least 2-fold downregulated while 17% of genes showed significant upregulation (Figure 5B, Supplementary Table 8). The analysis of genes downregulated at 48 hours revealed high enrichment for genes involved in RNA metabolic processes (Supplemental Figure 8A). Nucleotide supplementation restores the gene signature imposed by A77 1726, indicating that the transcriptional changes originate from nucleotide stress (Figure 5C).

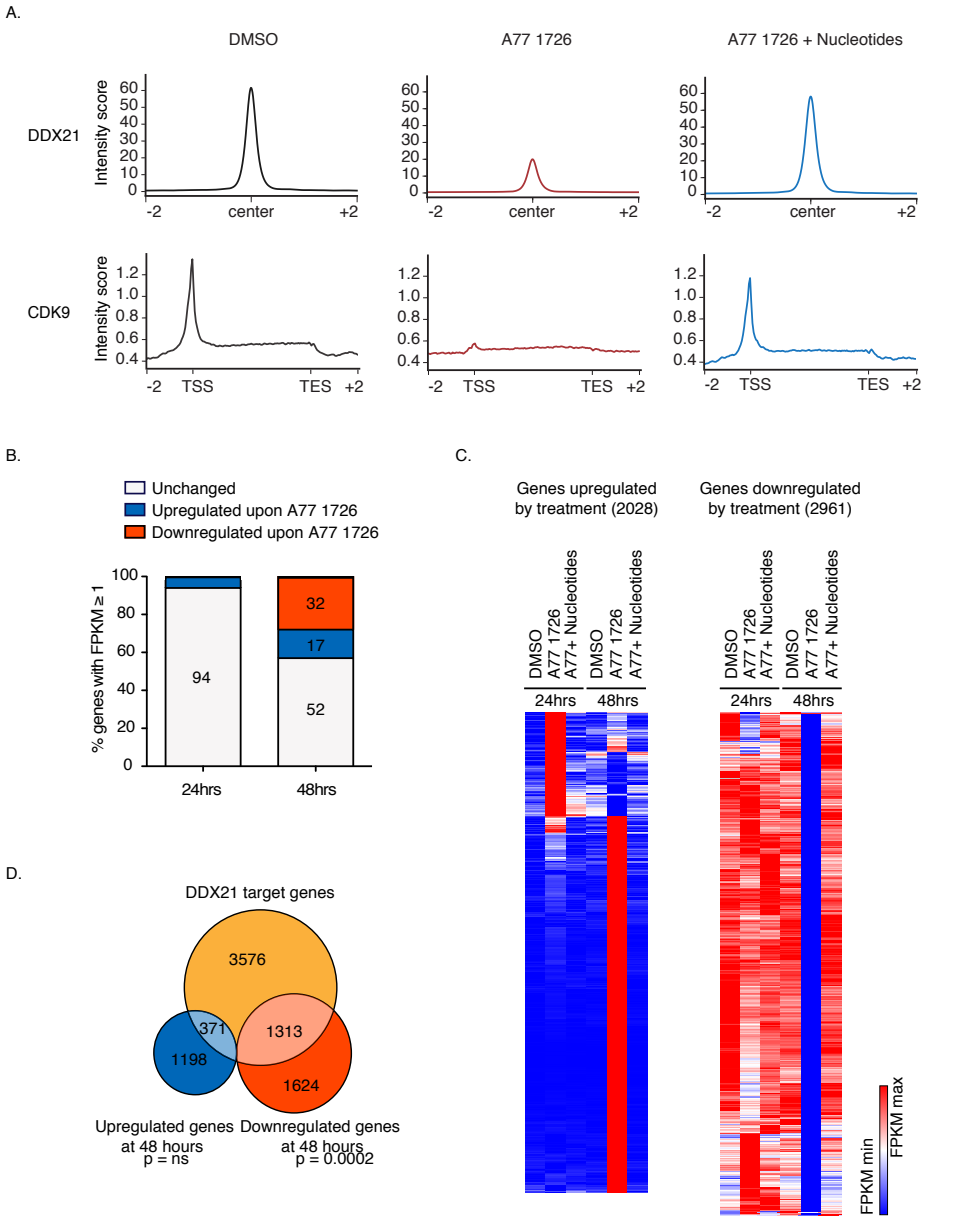


Figure 5. DDX21 mediates transcriptional changes during nucleotide stress. (A) Metagene analysis of DDX21 and CDK9 in A375 melanoma cells after 24 hours of treatment. Upper panel: Normalized peak intensities ±2kb from the DDX21 peak center are shown. Lower panel: Normalized CDK9 occupancy at transcription start site (TSS), along gene bodies and at transcription end site (TES). (B) The percentage of differentially expressed genes (DEGs) in A375 cells after 24 hours and 48 hours in the presence of A77 1726. DEG criteria: FPMK ≥1 after treatment, Log_2 fold change ≥1 or ≤ -1. (C) Hierarchical clustering heatmap of FPKM values from genes upregulated (2028) and downregulated (2961) at 24 hours and 48 hours post DHODH inhibition plus or minus nucleotide supplementation. (D) Venn diagram of DDX21 target genes with reduced DDX21 binding upon A77 1726 treatment as defined by ChIP-Seq (orange). Genes upregulated (blue) and downregulated (red) by RNA-Seq at 48 hours post treatment with A77 1726 in A375 melanoma cells (n = 3 biologically independent experiments, Hypergeometric Test, ns = not significant; under enrichment p-value = $2.7e^{-58}$).

We then compared DEGs at 48 hours to genes displaying reduced chromatin occupancy of DDX21 at 24 hours post A77 1726 treatment. We observed that 45% of the downregulated genes show loss of DDX21 binding during nucleotide depletion (Figure 5D). Gene set analysis of downregulated DDX21 targets uncovered enrichment in genes associated with RNA metabolism and the cell cycle (Supplemental Figure 8B). The data show that loss of DDX21 binding precedes transcriptional downregulation of genes such as CSTF2 (mRNA processing) and CCNB1 (cell cycle) under nucleotide stress. These observations could explain the transcriptional and proliferative defects seen in neural crest and melanoma cells after DHODH inhibition[1,9].

To test whether loss of DDX21 and CDK9 from chromatin impacts promoter-proximal pausing, we performed precision run-on sequencing (PRO-Seq) in A375 cells 24 hours post A77 1726 treatment[26]. Consistent with a broad increase in transcriptional pausing, we observed a genome-wide increase in signal at the promoter region. No changes within the gene body were found (Supplemental Figure 8C, 8D). The PRO-Seq data suggests that the increased Pol II pausing that occurs 24 hours after treatment underlies the transcriptional downregulation seen at 48 hours. Due to its strong interaction with CDK9[19], we hypothesize that DDX21 sequesters CDK9 from chromatin following nucleotide depletion. This could drive the observed genome-wide loss of CDK9 under nucleotide stress. Consequent increases in Pol II pausing lead to transcriptional defects driving cell cycle arrest. In summary, limited nucleotide pools reduce expression of DDX21-bound genes.

Discussion

Here, we employed an *in vivo* chemical suppressor screen in zebrafish to find chemicals that confer resistance to DHODH-mediated nucleotide stress and subsequent transcriptional defects. While we do not distinguish between cell autonomous and non-cell autonomous effects of leflunomide, we consider both mechanisms to play a role. We found that alterations in progesterone and Pgr signaling lead to neural crest resilience when nucleotide pools are limited. Progesterone can bypass the elongation block induced by DHODH inhibition or Paf complex deficiency *in vivo.* We suggest that the downregulation of Pgr reduces transcriptional activity of the receptor hereby ameliorating effects due to low abundance of nucleotides. We found that PGR interacts with DDX21 and propose that DDX21 acts as a mediator between nuclear hormone signaling and the nucleotide stress response. Competition between the two signals could occur when both nucleotide stress and progesterone are present. *In vitro,* DDX21, together with CDK9, falls off chromatin to prevent transcription elongation when nucleotide levels are limited. Following treatment with progesterone, DDX21 activity could be skewed towards its function in nuclear hormone signaling. We hypothesize that the presence of both progesterone and nucleotide depletion restores the interaction between DDX21

and CDK9 on chromatin thereby re-establishing efficient transcription and rescuing the neural crest lineage. In addition, our study reveals a function for DDX21 as a sensor of nucleotide pools and mediator of transcription during nucleotide stress. DDX21 occupancy is lost at Pol II, but not Pol I (rDNA) targets when nucleotide pools are limited. This shows that polymerase I and II hold intrinsic differences in their response to changes in nucleotide levels. While ribosomal genes are affected by nucleolar stress[20], cell cycle genes are affected by nucleotide stress through Pol II-mediated transcriptional effects. Patients undergoing chemotherapy regimens that target nucleotide metabolism often develop resistance[27]. This study suggests that modulating DDX21 levels could be an attractive therapeutic strategy to delay disease progression.

Methods

Zebrafish handling

Zebrafish were handled according to the vertebrate animal protocol as approved by the Harvard University Animal Care Committee. Strains used are AB (wild-type) and *paf1(aln^{z24})* mutants.

Chemical screen and drug treatment

Wild-type zebrafish embryos were treated with leflunomide from 50% epiboly until 24 hours post-fertilization in E3 embryo media and fixed with 4% Formaldehyde. Chemical libraries screened: LOPAC (n= 1280), ICCBL (n= 480) and NIH clinical collection (n=450). Other chemical used include leflunomide (Sigma L5025), Progesterone (Sigma, P0130), Esomeprazole (Enzo life sciences), Aphidicolin (Enzo life sciences, BML-CC101-0001) and A77 1726 (Enzo Life Sciences, ALX-430-096-M025). For nucleotide rescue experiments, a cocktail of 10µg/ml uridine 5′-monophosphate (Sigma, UMP) and 10µg/ml cytidine 5′-monophosphate (Sigma, CMP) was added to A77 1726. For *in situ* hybridizations, we followed the methods as described by Thisse C and Thisse, B[28,29].

Cell culture and generation of A375-iPRG (PRB) line

A375 cells (ATCC) and T47D cells (ATCC) were grown in DMEM or RPMI with 10% fetal bovine serum, 1X GlutaMAX and 1% Penicillin-Streptomycin. The full-length human PRB CDS was cloned into a pENTD (Life technologies) from a vector kindly provided by Brown Myles via Gateway recombination (Invitrogen). The primers used are F: caccATGACTGAGCTGAAGGCAAAGGG, R: CTTTTTATGAAAGAGAAGGGGTTTCAC (Supplementary Table 9). Subsequently PRB was subcloned to add a myc-tag using F: caccATGACTGAGCTGAAGGCAAAGGG, R: CTATAGTTCTAGAGGCTCGAGAG. The pENTD vector containing myc-tagged PRB was cloned into the pHage vector via Gateway recombination. Lentiviral particles were produced by co-transfection of 293T cells (ATCC) with sequence verified pHage-PRB and packaging plasmids pMD2.G (Addgene, #12259) and psPAX2 (Addgene #12260, gifts from Didier Trono), using FuGENE HD (Promega). Viral particles were harvested 48 hours and 72 hours post-transfection, concentrated by overnight PEG precipitation[30], resuspended in PBS, and stored at -80°C. A375 cells were overlaid with viral particles diluted in DMEM/1x Glutamax with 10% TET System Approved FBS (Clontech) and supplemented with 5µg/ml polybrene (Sigma) for 24 hours at 37°C. 48 hours post-transduction, infected cells were selected with 500µg/ml G418 (Gibco) for 7 days. Cell line is available upon request.

Targeted mass spectrometry and metabolite profile analyses

Targeted mass spectrometry and metabolite analysis were performed as previously described[9]. In short, samples were re-suspended using 20µL HPLC grade water.

5µL were injected and analysed using a hybrid 5500 QTRAP triple quadrupole mass spectrometer (AB/SCIEX) coupled to a Prominence UFLC HPLC system (Shimadzu) via selected reaction monitoring (SRM)[31]. Peak areas from the total ion current for each metabolite SRM transition were integrated using MultiQuant v2.0 software (AB/SCIEX). Spectral counts of metabolites were first normalized to the row average of the DMSO samples, followed by normalization to the median of the entire dataset.

cDNA cloning and in vitro transcription

Danio rerio Pgr cDNA was cloned into pENTR/D-TOPO (Thermo Fisher Scientific) using the following primers: F: CACCATGGACACGGTGAAC, R: CATCAGCGGTTTGACCATTCCTG. Final Destination vector cloning was performed using LR Clonase mix (Thermo Fisher Scientific) and 5'entry p5E-CMV/SP6 vector, 3'entry flag:bio:T2A:gfp vector and a pDestTol2CG2 backbone. The 7mm pgr construct was generated by NEB Q5 mutagenesis using pENTRD-pgr as a template and with the following primers: Q5_F:atacgtcccctgCTGATTCCACTGGGACGGTGACGGG, Q5_R: tgacggtatccatGGTGAAGGGGGCGGCCGC. hsDDX21 (wild-type) was subcloned using BamHI/XhoI into pCS2+ from the original plasmids provided by E. Calo[19]. Primers used: F:GTCAGGATCCATGCCGGGAAAACTCC, R:CGTACTCGAGTCATCATTGACCAAATGCTT. mRNA was generated using mMESSAGE mMACHINE SP6 Transcription Kit (Life technologies). 1nl of mRNA mix was injected into 1-cell stage embryos.

Protein extracts Co-IP and mass spectrometry of zebrafish embryos

500 embryos were injected with *pgr:flag:T2A:gfp* mRNA, treated with the respective chemicals and collected at 15 somites. Embryos were checked for GFP expression, dechorionated, deyolked[32] and resuspended in Pierce IP Buffer. Anti-Flag M2 Magnetic beads (Sigma, M8823) were used to pull down Pgr:Flag proteins. Protein complexes were eluted using 3xFlag peptides (Sigma, F4799). Proteins were concentrated by TCA precipitation and loaded on a 10% acrylamide gel. Gel was stained with Colloidal Blue and excised bands were cut into 1mm^3 pieces. Gel pieces were subjected to a modified in-gel trypsin digestion procedure[32]. Gel pieces were washed and dehydrated with acetonitrile for 10 minutes followed by removal of acetonitrile. Pieces were dried in a speed-vac. Rehydration of the gel pieces was done using 50mM ammonium bicarbonate solution containing 12.5ng/µl modified sequencing-grade trypsin (Promega) at 4ºC. After 45min, excess trypsin solution was removed and replaced with 50mM ammonium bicarbonate solution to cover the gel pieces. Samples were placed at 37ºC overnight. Peptides were extracted by removing the ammonium bicarbonate solution, followed by one wash with a solution containing 50% acetonitrile and 1% formic acid. The extracts were dried in a speed-vac and samples were stored at 4ºC. On the day of analysis, the samples were reconstituted in 5-10µl of HPLC solvent A (2.5% acetonitrile, 0.1% formic

acid). A nano-scale reverse-phase HPLC capillary column was created by packing 2.6µm C18 spherical silica beads into a fused silica capillary (100µm inner diameter x 30cm length) with a flame-drawn tip. After equilibrating the column each sample was loaded via a Famos auto sampler (LC Packings). A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). As peptides eluted, they were subjected to electrospray ionization and entered into an LTQ Orbitrap Velos Pro ion-trap mass spectrometer. Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences were determined by matching to protein databases using Sequest software (v.28 (rev. 12). All databases include a reversed version of the sequences and the data was filtered to a 1-2% percent false discovery rate.

Co-immunoprecipitation (Co-IP)

Cells were washed twice with 1X PBS. Nuclei were isolated with 0.05% Triton in PBS and lysed in nuclei lysis buffer (20mM Hepes-KOH pH7.9, 25% glycerol, 420mM NaCl, 1.5mM MgCl2, 0.2mM EDTA, 0.5mM DTT). 500ug of protein extracts was used and the salt concentration was diluted from 420mM to 150mM NaCl using 20mM Hepes-KOH pH7.9, 20% glycerol, 0.25mM EDTA, 0.05% NP-40. Cell lysates were pre-cleared with non-antibody bounds beads for 1 hour at 4°C. Antibody bounds protein G Dynabeads were added to pre-cleared lysate and samples incubated overnight at 4°C. Antibodies used: IgG (Cell Signalling #2729), Progesterone Receptor B (C1A2), Cell Signalling #3157) and DDX21 (Novus Biologicals NB100-1718). Protein-bead complexes were then washed 5 times with wash buffer (20mM Hepes-KOH pH7.9, 10% Glycerol, 150mM NaCl, 1.5 MgCl2, 0.2mM EDTA, 0.5mM DTT) and beads were boiled in 50µL Laemmli Buffer for 15min at 95°C to elute proteins.

Western blotting and immunofluorescence stainings

A375 cells were treated for 24 hours, washed in 1X PBS, and collected in RIPA buffer with protease and phosphatase inhibitors. Samples were run on acrylamide gel and transferred onto a nitrocellulose membrane. Membrane was blocked for one hour in 5% of milk or TBS-T and incubated overnight at 4°C with anti-DDX21 (Novus Biologicals NB100-1718, 1:3000), Progesterone Receptor B (C1A2), Cell Signalling #3157, 1:1000) or anti-actin (Abcam, ab14128, 1:1000). The next day, membranes were washed, incubated with HRP-conjugated secondary antibodies for 1 hour at room temperature, and developed with SuperSignal West Pico Plus Chemiluminescent substrate. DDX21 immunostaining was performed and quantified as described previously[19] using anti-DDX21 (NBP1-83310, 1:300) and anti-TCOF (NBP1-86909, 1:100). Images were acquired using an Andor Zyla VSC-04746 camera, calibrated at 1.63µm/px, binning of 1x1 and an exposure of 70ms.

RNA extraction, RT-PCR and qRT-PCR

RNA isolation was performed using Trizol (Life Technologies, 15596-026) and the RNAeasy Kit (Qiagen). cDNA was synthesized with the Superscript Vilo cDNA synthesis kit (Life Technologies, 11754-050).

RT-PCR primers used:

Actin	Fwd: ACCTCATGAAGATCCTGACC
	Rev: TGCTAATCCACATCTGCTGG
Pgr1	Fwd: AAAGCTGCTACGACTCCACC
	Rev: AGACGACATGCGGGACAATT
Pgr2	Fwd: CAGACAGCATACACCGCATT
	Rev: GCTGTTGAGGAGGAGGTGAG
ddx21	Fwd: ATCCAGCATGCCGTCAAAGA
	Rev: TCAAACAGGTACGAGACGCC

qPCR primers used:

bactin	Fwd: CGAGCAGGAGATGGGAACC
	Rev: CAACGGAAACGCTCATTGC
sox10	Fwd: ATATCCGCACCTGCACAA
	Rev: CGTTCAGCAGTCTCCACAG
crestin	Fwd: AGTGCCTGCCAATGTTCAC
	Rev: CTGAAAAAGGCCGATGAGTT
foxd3	Fwd: CATGCAAAACAAGCCCAAG
	Rev: ATGAGGGCGATGTACGAGTAG
mitfa	Fwd: GGCGGTTTAATATCAATGACAGA
	Rev: GGTGCCTTTATTCCACCTCA
snail2	Fwd: TTATAGTGAACTGGAGAGTCCAACA
	Rev: TCCATACTGTTATGGGATTGTACG
neurog1	Fwd: CGTGCCATTATCTTCAACACA
	Rev: CGATCTCCATTGTTGATAACCTT
myf5	Fwd: GCTACAACTTTGACGCACAAAA
	Rev: CACGATGCTGGACAAACACT
pgr (1)	Fwd: TTCTCGCTGGGTTGGAGAAC
	Rev: GCATAGCCAAGCAAAGGTCG
pgr (2)	Fwd: CAGCATACACCGCATTCTCC
	Rev: ACCCTCAACAGCTGTCTTGA

Cell sorting and RNA-sequencing

300 embryos were treated with 7uM leflunomide, 16uM Progesterone or both drugs from 50% epiboly until 15 somites. Embryos were dechorionated with pronase and treated

with Liberase. Sox10+ cells were sorted on a BD FACS aria II. 16.5ng of RNA from Sox10+ sorted cells was used for library preparation. For RNA-sequencing of zebrafish and human cells, isolated RNA was subjected to rRNA depletion (Zebrafish: RiboGone, Takara #634847, Human: Epicentre, RZG1224, MRZ11124C). For zebrafish: NGS libraries were generated using the Smarter Low Input and Low input library prep kits (Takara). For human: libraries were prepared using the NEBnext Ultra RNA library prep kit (NEB, E7530S). Libraries were sequenced on Illumina Hi-Seq2000. Quality control of RNA-Seq datasets was performed by FastQC and Cutadapt to remove adaptor sequences and low-quality regions. The high-quality reads were aligned to UCSC build danRer7 for zebrafish or hg19 for human using Tophat 2.0.11 without novel splicing form calls. Transcript abundance and differential expression were calculated with Cufflinks 2.2.1. FPKM values were used to normalize and quantify each transcript; the resulting list of differential expressed genes are filtered by $\log^2$ fold change ≥ 1.

Morpholino injection and pgr overexpression

Morpholinos were obtained from Gene Tools, LLC, and injected into the yolk of 1-cell stage embryos with 2ng MO in 1nl volume. The ctr9 MO has been described by Bai et al., 2013[33]. Morpholino sequences are:

ctr9 (NM_001083583): GATTTCAATGGATCCCCGAGACATG

pgr (NM_001166335.1): GGAGAAGTGTTCACCGTGTCCATTC

ddx21 (ENSDART00000093149.5): ATTCTGGGAGACTCTTTGACGGCAT

100pg of wild-type pgr:flag:T2A:gfp mRNA (wild-type pgr mRNA) or mismatch pgr:flag:T2A:gfp mRNA containing 7 silent mutations (7mm pgr mRNA) was co-injected for MO experiments. For overexpression, 100pg of wild-type pgr:flag:T2A:gfp mRNA was injected.

ChIP-sequencing and analysis

ChIP-Seq was performed as previously described[9,34]. A375 cells were treated with 25μM DMSO, 25μM A771726 or 25μM A771726 plus nucleotides. After 24 hours, cells were fixed in 1% formaldehyde and subjected to Co-IP using DDX21 (Novus Biologicals, NB100-1718) and Cdk9 (Santa Cruz, C-20) antibodies. Libraries were sequenced on Illumina Hi-Seq2000. All datasets were aligned to UCSC build version hg19 of the human genome using Bowtie2 (version 2.2.1) with the following parameters: --end-to-end, -N0, -L20. We used the MACS2 version 2.1.0[35] peak finding algorithm to identify regions of ChIP-Seq peaks, with a q-value threshold of enrichment of 0.05 for all datasets. The genome-wide occupancy profile figures were generated by deeptools[36] using the reference-point mode and the scale-regions mode. The genomic distribution of DDX21 ChIP was plotted using the ChIPSeeker R package, annotatePeak to assign peaks to a genomic annotation, which includes whether a peak is in the TSS, Exon, 5' UTR, 3' UTR, Intronic or Intergenic. The genome annotation is from the R-bioconductor annotation

packages. To map DDX21 ChIP-seq peak to the rDNA data, we followed published work[19]. Briefly, we obtained the DNA consensus sequence of the 43kb ribosomal locus NCBI (GeneBank ID: U13369.1). Using the unique 43kb region, we used Bowtie to map ChIP-seq reads with standard mapping parameters to the hg19 human genome build. Data was visualized with Integrative Genomics Viewer.

Infrared Crosslinking and Immunoprecipitation (irCLIP)

irCLIP was performed as described[37]. DDX21 (Novus Biologicals, NB100-1718) bound to Protein G Dynabeads (Thermo Fisher Scientific) was used to IP for 8 hours at 4°C. irCLIP data was processed using the irCLIP pipeline (https://github.com/ChangLab/irCLIP/tree/lite). PCR duplicates were removed using unique molecular identifiers (UMI) in the RT primer region. Adapter and barcode sequences were trimmed, and reads were mapped step-wise to repetitive and then non-repetitive (GRCh38) genomes. Specific parameters used for the irCLIP pipeline are as follows: -f 18 (trims 17nt from the 5' end of the read), -l 16 (includes all reads longer than 16nt), –bm 29 (minimum MAPQ score from bowtie2 of 29 is required for mapping; unique mapping only), and –tr 2,3 (repetitive genome) and –tn 2,3 (non-repetitive genome) RT stop intersection (n,m; where n = replicate number and m = number of unique RT stops required per n replicates). Using the –tr/tn 2,3 parameters, a minimum of 6 RT stops are required to support any single nucleotide identified as crosslinking site.

Whole cell proteomics

Cells were lysed by homogenization in lysis buffer (2% SDS, 150mM NaCl, 50mM Tris pH 7.4). Lysates were reduced with 5mM DTT, alkylated with 15mM iodoacetamide for 30 minutes in the dark, quenched with 50mM fresh DTT and proteins precipitated by methanol/chloroform precipitation. Digests were carried out in 200mM EPPS, pH 8.5 in presence of 2% acetonitrile (v/v) with LysC (Wako, 2mg/ml at 1:75) for 3 hours at RT and subsequently trypsinated (Promega, #V5111, stock 1:75) overnight at 37°C. Missed cleavage rate was assayed from a small aliquot by mass spectrometry. For whole proteome analysis, digests containing 60µg of peptide material were directly labelled with TMT reagents. After quenching of TMT labelling reactions with hydroxylamine, TMT labelling reactions were mixed, solvent evaporated to near completion and TMT labelled peptides purified and desalted by acidic reversed phase C_{18} chromatography. Peptides were then fractionated by alkaline reversed phase chromatography into 96 fractions and combined into 12 samples. Before mass spectrometric analysis, peptides were desalted over Stage Tips[38]. Data were collected by a MultiNotch MS3 TMT method[39] using an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) coupled to an EasynLC1200 HPL system (Thermo Fisher Scientific). The 100µm inner diameter capillary column used was packed with C_{18} resin (SepPax Technologies Inc., 1.8µm). Peptides of each fraction were separated over 3 hours acidic acetonitrile gradients by LC prior to

mass spectrometry (MS) injection. The first scan of the sequence was an MS1 spectrum (Orbitrap analysis; resolution 120,000; mass range 400–1400 Th). MS2 analysis followed collision-induced dissociation (CID, CE=35) with a maximum ion injection time of 150 ms and an isolation window of 0.7 Da. In order to obtain quantitative information, MS3 precursors were fragmented by high-energy collision-induced dissociation (HCD) and analyzed in the Orbitrap at a resolution of 50,000 at 200Th. Further details on LC and MS parameters and settings used were described recently[40]. Peptides were searched with a SEQUEST (v.28, rev. 12) based software against the human proteome (Uniprot 02/2014) with added common contaminant proteins. For this, spectra were first converted to mzXML. Searches were performed using a mass tolerance of 50 ppm for precursors, fragment ion tolerance 0.9Da to maximize sensitivity in conjunction with SEQUEST searches and linear discriminant analysis. For the searches maximally 2 missed cleavages per peptide were allowed. We searched dynamically for oxidized methionine residues (+15.9949 Da) and applied a target decoy database strategy and a false discovery rate (FDR) of 1% set for peptide-spectrum matches following filtering by linear discriminant analysis (LDA) and a final collapsed protein-level FDR of 1%. Quantitative information on peptides was derived from MS3 scans. Quant tables were generated requiring an MS2 isolation specificity of >65% for each peptide and a sum of TMT s/n of >150 over all channels for any given peptide and exported to Excel and further processed therein. Details of the TMT intensity quantification method and further search parameters applied were described recently[40].

PRO-Seq analysis

Cells were collected, washed with PBS and washed with wash buffer (10mM Tris-Cl, pH 8.0, 10mM KCl, 250mM sucrose, 5mM MgCl2, 0.5mM DTT, 10% Glycerol). Cells were permeabilized with 0.1% Igepal for 2 minutes at RT, spun and resuspended in freezing buffer (50mM Tris-CL pH 8.3, 40% glycerol, 5mM MgCl2, 0.5mM DTT). Nuclei were counted, and flash frozen. PRO-Seq libraries were generated using 1 million nuclei per sample. Samples were spiked with 40,000 Drosophila S2 cells as normalization control. Samples were sequenced on the NextSeq using a High Output 75-cycle kit to an average depth of 10 million mappable reads per sample. To remove adapter sequence and low quality 3' ends, we used cutadapt 1.14 to discard reads shorter than 20nt (-m 20 -q 10), and removing a single nucleotide from the 3' end of all trimmed reads to allow successful alignment with Bowtie 1.2.2. Remaining read pairs were aligned to the Drosophila Dm3 genome index to determine spike-normalization ratios based on uniquely mapped reads. Counts of pairs mapping uniquely to spike-in RNAs were determined for each sample. Depth normalization was used for each bedGraph. Reads mapped to dm3 were excluded from further analysis, and unmapped pairs were aligned to the human hg19 genome. Identical parameters were utilized in each alignment described above (-m1, -v2, -X1000, --un). Pairs mapping uniquely to hg19, representing biotin-labeled RNAs were

separated, and strand-specific counts of the 3'-end mapping positions determined at single nucleotide resolution, genome-wide, and expressed in bedGraph format with "plus" and "minus" strand labels, as appropriate. Combined bedGraphs were generated after deduplication by summing counts per nucleotide of all 3 replicates for each condition. Read counts were calculated per gene (from transcription start site to transcription end site), in a strand-specific manner, based on default annotations (Ensembl hg19), using feature Counts. Differentially expressed genes were identified using DESeq2 v1.18.1 under R 3.3.1. PRO-Seq size factors were determined based on spike normalization (for Control: 1.401, 1.365, 1.422; drug-treated: 0.973, 1.158, 0.832; rescued: 1.225, 1.805, 1.089). At an adjusted p-value threshold of $p<0.05$, 3135 affected genes were identified in A77 1726 treated cells and 2 affected genes in A77 1726 plus nucleotides treated cells, as differentially expressed compared to A375 control cells. UCSC Genome Browser tracks were generated from the combined replicates per condition, normalized as in the differential expression analysis.

Statistics and reproducibility

All data represents 3 or more independent experiments, unless otherwise indicated here. In Supplemental Figure 3B, qPCR data of sorted sox10^{+}-GFP cells represents one experiment of three technical replicates due to limited material. In Supplemental Figure 4A, RT-PCR data of *pgr* expression during early development is from one independent experiment. Results were in line with previous reports of describing *pgr* expression during early zebrafish development as determined by *in situ* hybridization and single-cell RNA-Seq in zebrafish embryos by Bertrand *et al.,* 2007[41] and Wagner *et al.,* 2018[42], respectively. All statistical analysis was performed using GraphPad Prism 5 Software.

Data availability

Deep-sequencing (ChIP-Seq, RNA–Seq, PRO-Seq, irCLIP) data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE128086. Mass spectrometry data have been deposited in ProteomeXchange with the primary accession code PXD014433. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Acknowledgments

We thank the following for their support and contribution: E. Fast, A. Choudhuri, and J.M. Ordovas-Montanes for critical reading of our manuscript; M. Rossman, E. Patton, J. Johansson, R. A. Young, and J. Wysocka for discussions about the project; M. Brown, L. Krug, D. Grunwald, S. Spengler, M. Yuan, J. Asara, L. Rubin, A. Avanites, and T. Schlaeger for help with reagents and experiments. This work was supported by NIH grant P50-GM107618 (M.K.), the Hope Funds for Cancer Research (B.J.A.) and the following grants from L.I.Z: Cancer Biology R01 CA103846, NIH Melanoma PPG, P01CA63222, Melanoma Research Alliance, Starr Cancer Consortium grant.

Author contributions

C.S., A.S, B.D., W.M., E.C.G., M.E.S., M.F., performed experiments. A.L., M.S., B.J.A., S.Y., provided formal data analysis. I.A. assisted with the zebrafish logistics. T.H., R.A.F., M.K., E.C., performed experiments and analysed data. M.J, K.A., T.H., and Y.Z., provided insights on data interpretation. C.S., A.S., and L.I.Z. wrote the manuscript. L.I.Z. conceived and managed the study.

Conflicts of interest

B.J.A. is a shareholder of Syros Pharmaceuticals. L.I.Z. is founder and stockholder of Fate, Inc., Scholar Rock, Camp4 therapeutics and a scientific advisor for Stemgent. The other authors declare no competing interests.

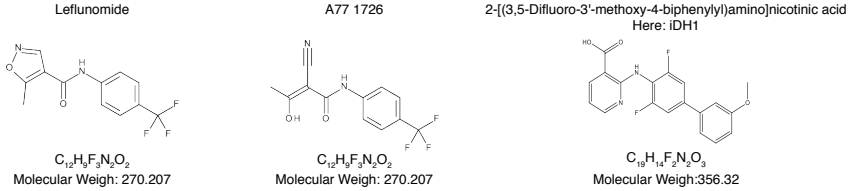
References

1. White, R. M. et al. DHODH modulates transcriptional elongation in the neural crest and melanoma. Nature 471, 518–522 (2011).
2. Lane, A. N. & Fan, T. W.-M. Regulation of mammalian nucleotide metabolism and biosynthesis. Nucleic Acids Res. 43, 2466–2485 (2015).
3. Chabosseau, P. et al. Pyrimidine pool imbalance induced by BLM helicase deficiency contributes to genetic instability in Bloom syndrome. Nature Communications 2, 368–6 (2011).
4. Liu, W. et al. Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. Proc. Natl. Acad. Sci. U.S.A. 109, 8983–8988 (2012).
5. Guo, J. & Price, D. H. RNA polymerase II transcription elongation control. Chemical Reviews 113, 8583–8603 (2013).
6. Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. eLife 6, R106 (2017).
7. Peng, J., Liu, M., Marion, J., Zhu, Y. & Price, D. H. RNA polymerase II elongation control. Cold Spring Harb. Symp. Quant. Biol. 63, 365–370 (1998).
8. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. 13, 720–731 (2012).
9. Tan, J. L. et al. Stress from Nucleotide Depletion Activates the Transcriptional Regulator HEXIM1 to Suppress Melanoma. Mol. Cell 62, 34–46 (2016).
10. 10. Luo, R., An, M., Arduini, B. L. & Henion, P. D. Specific pan-neural crest expression of zebrafish Crestin throughout embryonic development. Dev. Dyn. 220, 169–174 (2001).
11. Jurynec, M. J. et al. The Paf1 Complex and P-TEFb have reciprocal and antagonist roles in maintaining multipotent neural crest progenitors. Development dev.180133 (2019). doi:10.1242/dev.180133
12. Akanuma, T., Koshida, S., Kawamura, A., Kishimoto, Y. & Takada, S. Paf1 complex homologues are required for Notch-regulated transcription during somite segmentation. EMBO Rep. 8, 858–863 (2007).
13. Yoo, H.-S., Seo, J.-H. & Yoo, J.-Y. CTR9, a component of PAF complex, controls elongation block at the c-Fos locus via signal-dependent regulation of chromatin-bound NELF dissociation. PLoS ONE 8, e61055 (2013).
14. Cretekos, C. J. & Grunwald, D. J. alyron, an insertional mutation affecting early neural crest development in zebrafish. Developmental Biology 210, 322–338 (1999).
15. Rainger, J. et al. Miller (Genee-Wiedemann) syndrome represents a clinically and biochemically distinct subgroup of postaxial acrofacial dysostosis associated with partial deficiency of DHODH. Hum. Mol. Genet. 21, 3969–3983 (2012).
16. Read, L. D., Snider, C. E., Miller, J. S., Greene, G. L. & Katzenellenbogen, B. S. Ligand-modulated regulation of progesterone receptor messenger ribonucleic acid and protein in human breast cancer cell lines. Molecular Endocrinology 2, 263–271 (1988).
17. Diep, C. H., Ahrendt, H. & Lange, C. A. Progesterone induces progesterone receptor gene (PGR) expression via rapid activation of protein kinase pathways required for cooperative estrogen receptor alpha (ER) and progesterone receptor (PR) genomic action at ER/PR target genes. Steroids 114, 48–58 (2016).
18. Botos, J., Xian, W., Smith, D. F. & Smith, C. L. Progesterone receptor deficient in chromatin binding has an altered cellular state. J. Biol. Chem. 279, 15231–15239 (2004).
19. Calo, E. et al. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature 518, 249–253 (2015).
20. Calo, E. et al. Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. Nature 554, 112–117 (2018).
21. Xing, Y.-H. et al. SLERT Regulates DDX21 Rings Associated with Pol I Transcription. Cell 169, 664–678.e16 (2017).
22. Valvezan, A. J. et al. mTORC1 Couples Nucleotide Synthesis to Nucleotide Demand Resulting in a Targetable Metabolic Vulnerability. Cancer Cell 32, 624–638.e5 (2017).
23. Sarkar, M. & Ghosh, M. K. DEAD box RNA helicases: crucial regulators of gene expression and

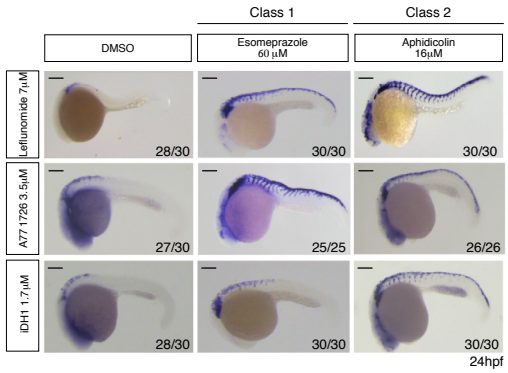
oncogenesis. Front Biosci (Landmark Ed) 21, 225–250 (2016).
24. Sloan, K. E. et al. The association of late-acting snoRNPs with human pre-ribosomal complexes requires the RNA helicase DDX21. Nucleic Acids Res. 43, 553–564 (2015).
25. Song, C., Hotz-Wagenblatt, A., Voit, R. & Grummt, I. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. Genes & Development 31, 1370–1381 (2017).
26. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science 339, 950–953 (2013).
27. Luengo, A., Gui, D. Y. & Vander Heiden, M. G. Targeting Metabolism for Cancer Therapy. Cell Chem Biol 24, 1161–1180 (2017).
28. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc 3, 59–69 (2008).
29. Thisse, B. & Thisse, C. In situ hybridization on whole-mount zebrafish embryos and young larvae. Methods Mol. Biol. 1211, 53–67 (2014).
30. Kutner, R. H., Zhang, X.-Y. & Reiser, J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. Nat Protoc 4, 495–505 (2009).
31. Yuan, M., Breitkopf, S. B., Yang, X. & Asara, J. M. A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. Nat Protoc 7, 872–881 (2012).
32. Link, V., Shevchenko, A. & Heisenberg, C.-P. Proteomics of early zebrafish embryos. BMC Dev. Biol. 6, 1–9 (2006).
33. Bai, X. et al. TiF1-gamma plays an essential role in murine hematopoiesis and regulates transcriptional elongation of erythroid genes. Developmental Biology 373, 422–430 (2013).
34. Batsché, E., Yaniv, M. & Muchardt, C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nature Structural & Molecular Biology 13, 22–29 (2006).
35. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137–9 (2008).
36. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–5 (2016).
37. Zarnegar, B. J. et al. irCLIP platform for efficient characterization of protein-RNA interactions. Nature Methods 13, 489–492 (2016).
38. Rappsilber, J., Friesen, W. J., Paushkin, S., Dreyfuss, G. & Mann, M. Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode. Anal. Chem. 75, 3107–3114 (2003).
39. McAlister, G. C. et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal. Chem. 86, 7150–7158 (2014).
40. Paulo, J. A. et al. Quantitative mass spectrometry-based multiplexing compares the abundance of 5000 S. cerevisiae proteins across 10 carbon sources. J Proteomics 148, 85–93 (2016).
41. Bertrand, S. et al. Unexpected novel relational links uncovered by extensive developmental profiling of nuclear receptor expression. PLoS Genet. 3, e188 (2007).
42. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science 360, 981-987 (2018).

Supplemental Figures

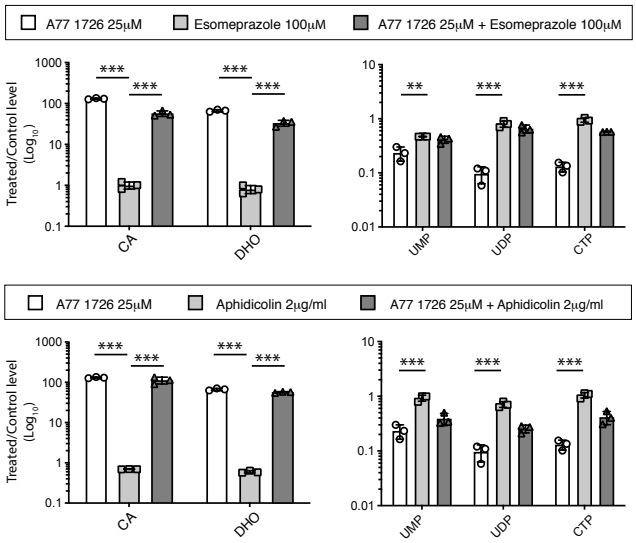
A.



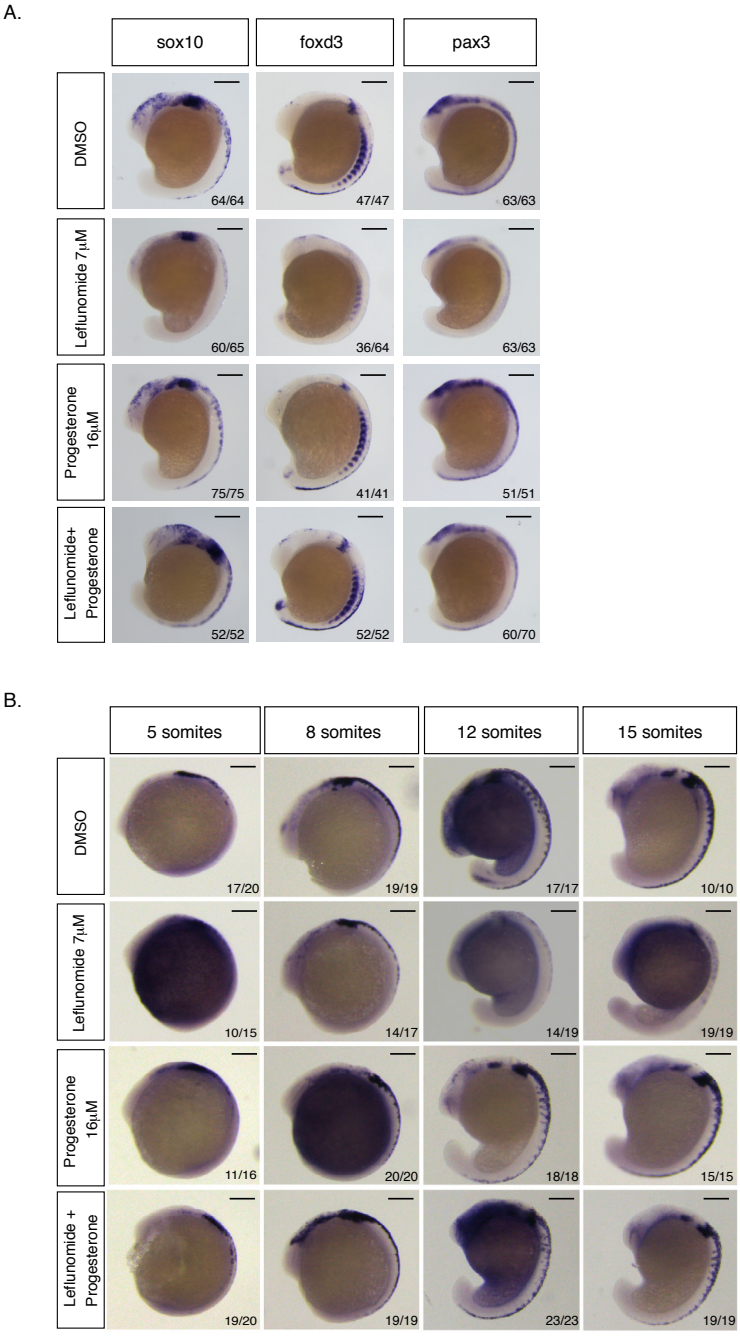
B.



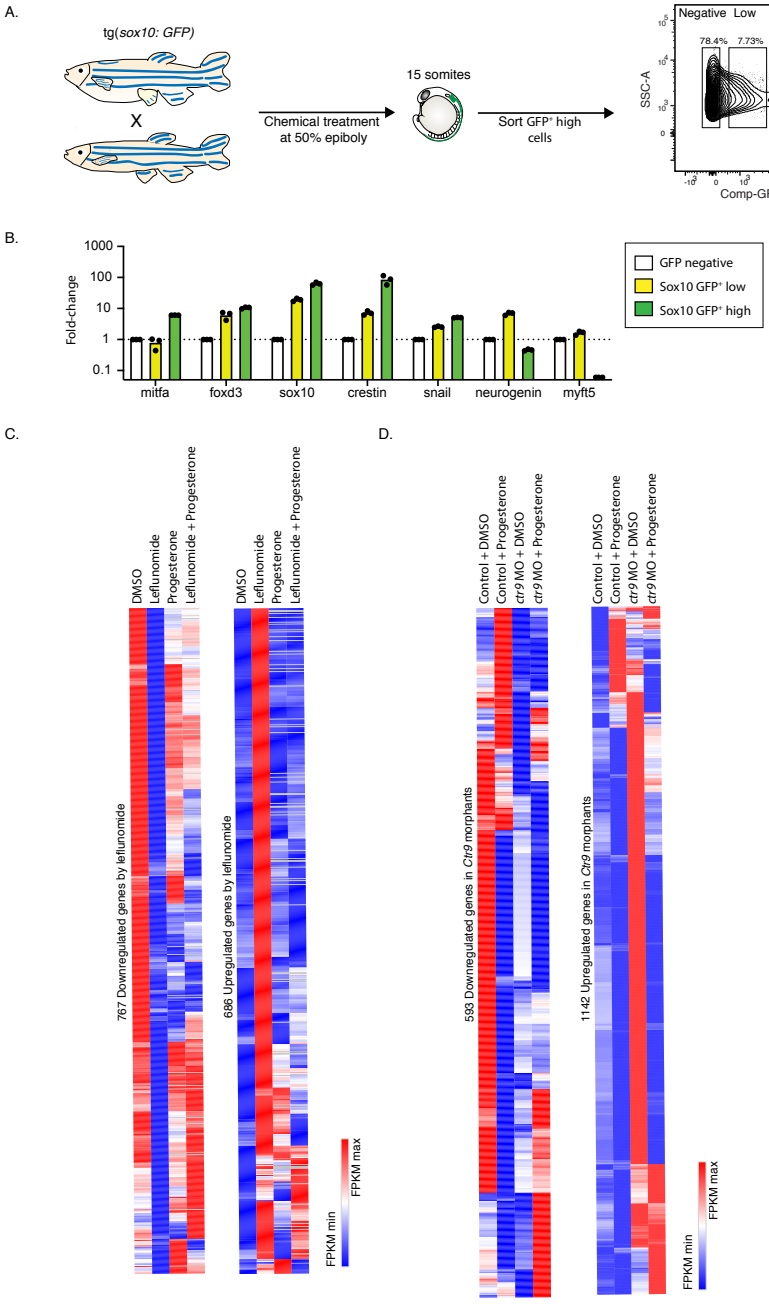
C.



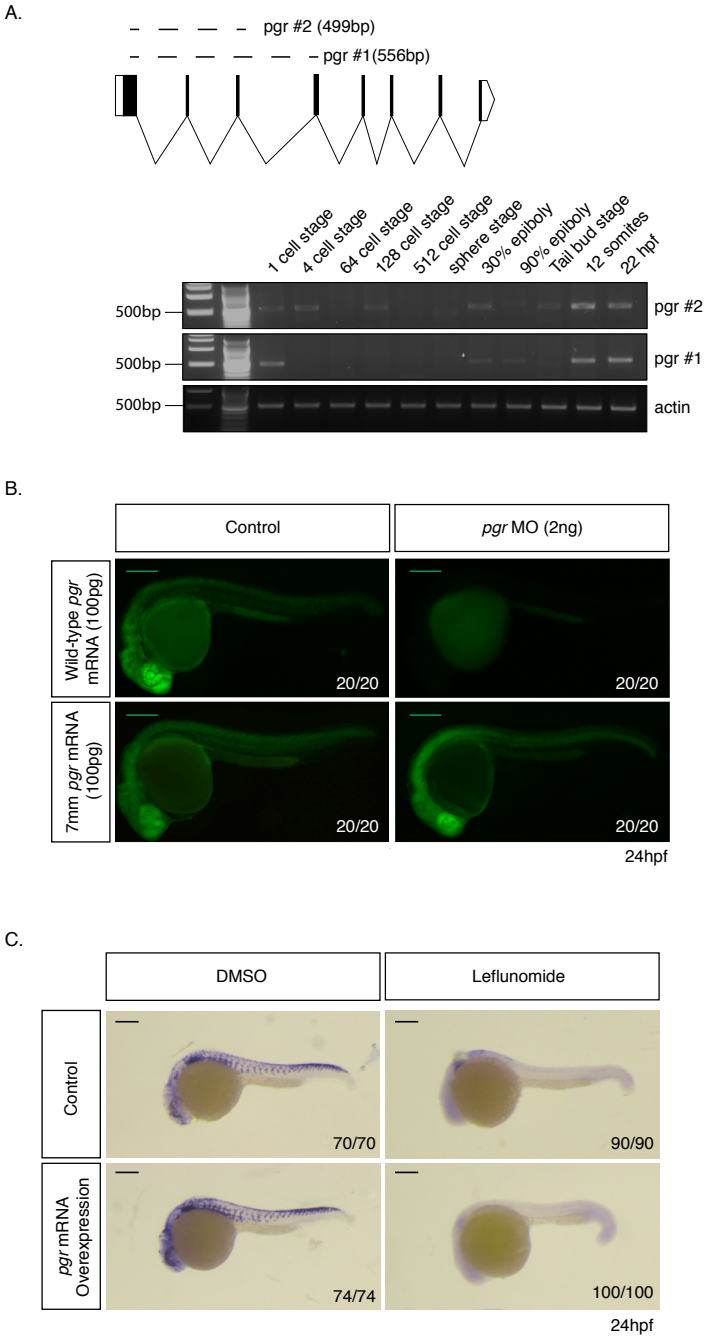
Supplemental Figure 1. Chemical suppressor screen for Leflunomide. (A) Chemical structure of leflunomide, the active metabolite of leflunomide known as A77 1726 and the independent DHODH inhibitor here named iDHODH1 (iDH1). (B) Lateral view of embryos at 24hpf treated with chemicals as indicated and subjected to *in situ* hybridization for *crestin*. Number of embryos displaying the indicated phenotype is indicated in the lower right corner. Scale bars represent 200µm. (C) Metabolite profiling in A375 melanoma cells. Upper panel: A375 cells exposed to A77 1726, Esomeprazol or both compounds for 24 hours. Lower panel: cells exposed to A77 1726, Aphidicolin or both chemicals. (n = 3 independent experiments, Mean ± SD, Two-Way ANOVA with Bonferroni Comparison, ** = $p < 0.01$, *** = $p < 0.0001$).



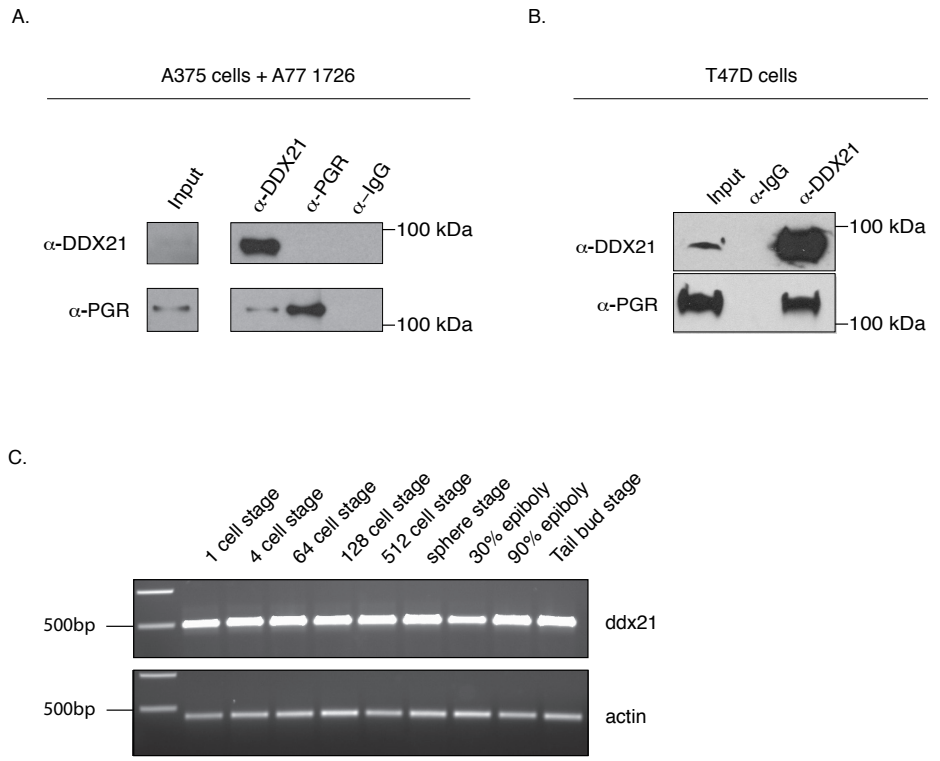
Supplemental Figure 2. Effects of Leflunomide and Progesterone on neural crest during embryonic development. (A) *In situ* hybridization for the neural crest cell markers *sox10, foxd3 and pax3* in zebrafish embryos at the 15-somite stage following treatment with indicated chemicals. (B) Lateral view of embryos subjected to *in situ* hybridization for *crestin* at 5, 8, 12 and 15 somites. Number of embryos displaying the presented phenotype is indicated in lower right corner. Scale bars represent 200μm.



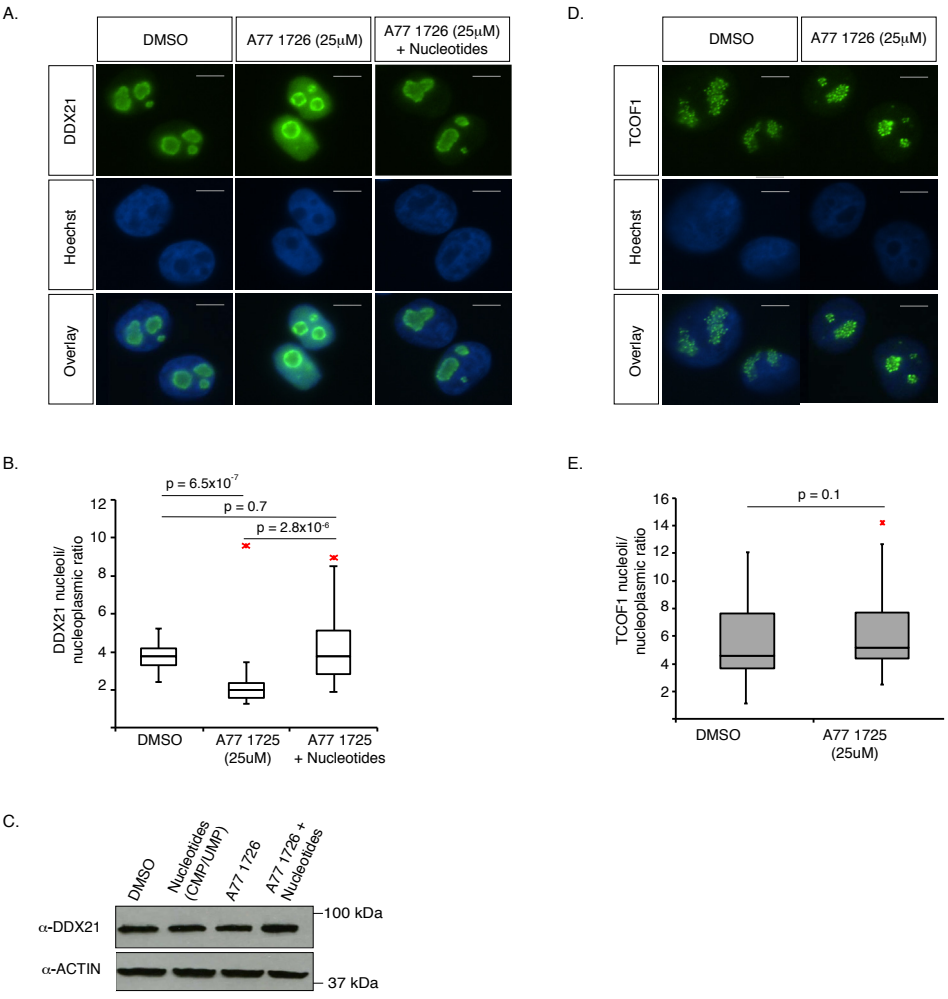
Supplemental Figure 3. Progesterone restores transcriptional changes caused by DHODH inhibition. (A) Schematic representation workflow to sort neural crest cells, here defined as *sox10:*GFP positive cells. (B) qPCR on whole embryos, sorted GFP low and sorted GFP high neural crest cells for neural crest (*mitfa, foxd3, sox10, crestin, snail2*) and non-neural crest (*neurogenin and myf5*) genes (Mean ± SD). (C, D) Hierarchical clustering heatmap of genes downregulated or upregulated in *sox10:*GFP high cells sorted from leflunomide-treated zebrafish (C) or *ctr9* morphants (D). Differentially expressed genes criteria: $\log_2$ fold change ≥ 1.5 or ≤ 1.5. (n = 3 biologically independent experiments).



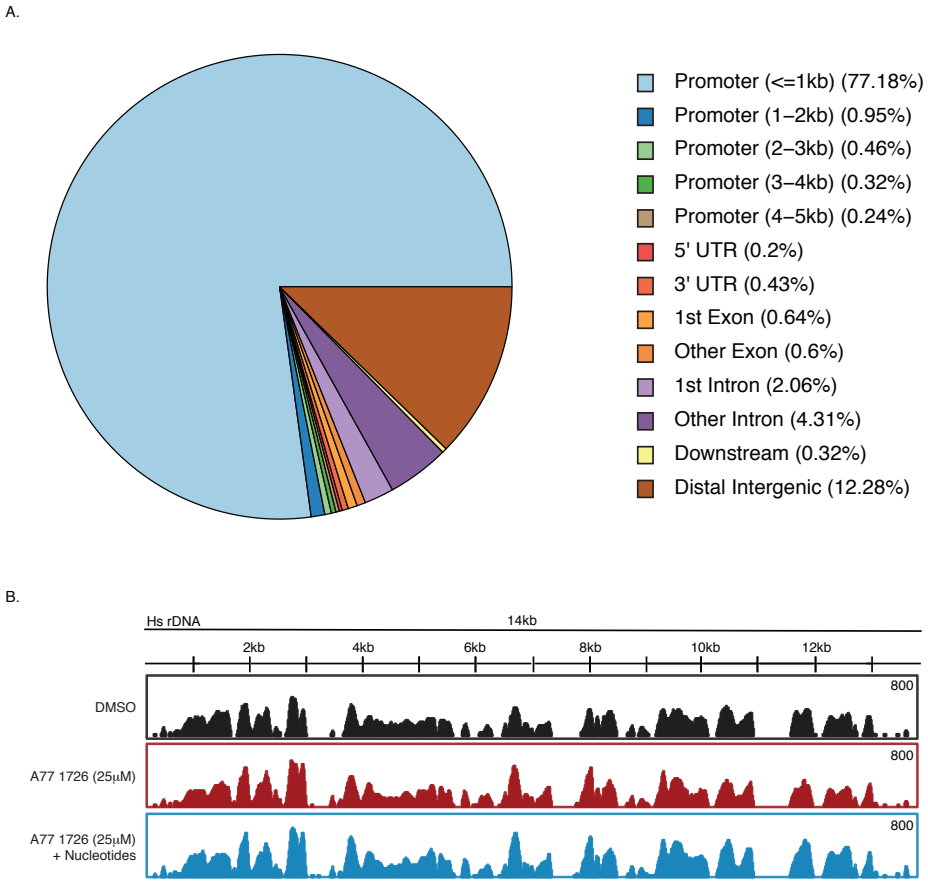
Supplemental Figure 4. Progesterone receptor expression and perturbation effects. (A) RT-PCR for *pgr* using two primer sets to reveal receptor expression during early development. (B) GFP positive embryos injected with *pgr:flag:T2A:gfp* mRNA or mismatched *pgr* mRNA with a *pgr* morpholine (MO) reveals MO specificity. (C) *In situ* hybridization for *crestin* at 24hpf in control and *pgr* mRNA injected embryos with or without leflunomide treatment. Number of embryos displaying the phenotype represented is indicated. Scale bars represent 200µm.



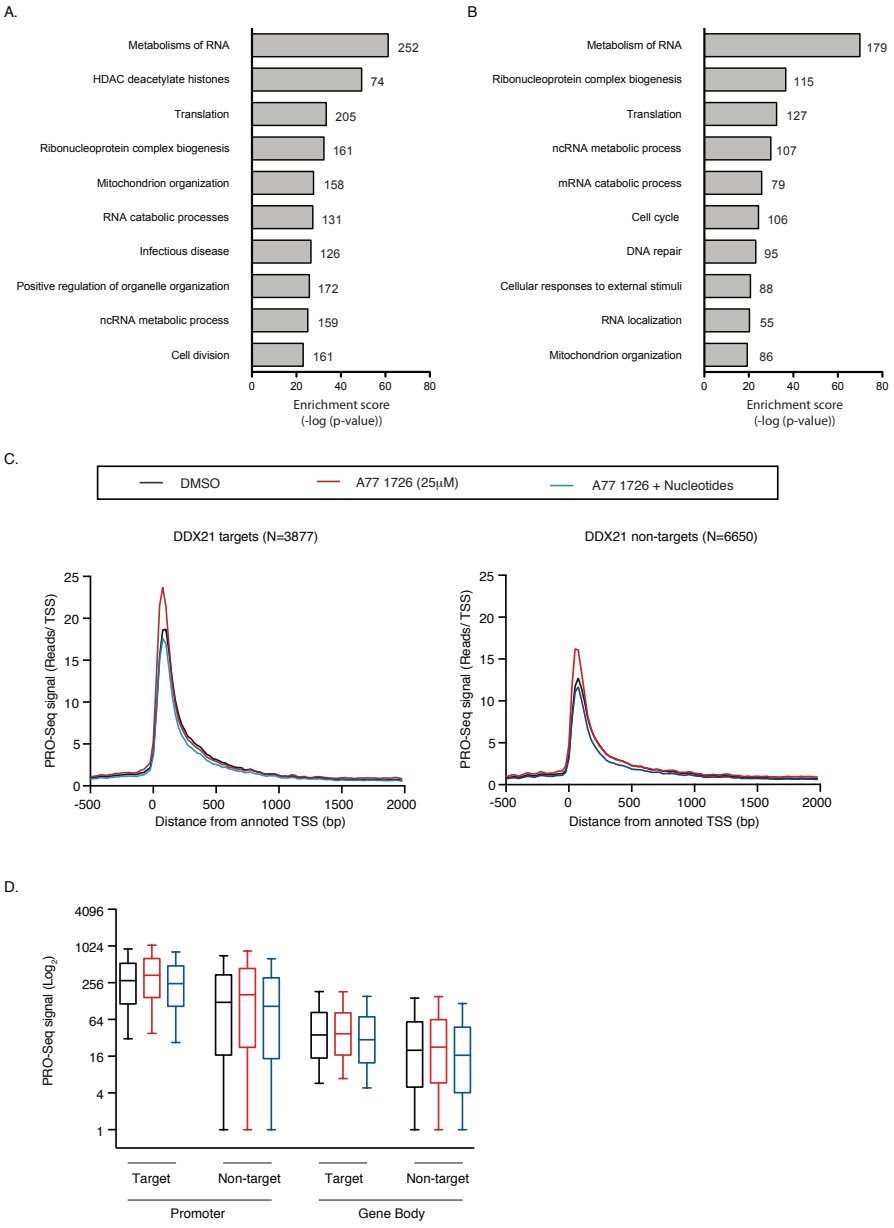
Supplemental Figure 5. DDX21 interacts with PGR and loss of *pgr* function rescues *crestin* expression *in vivo*. (A, B) DDX21 associates with PGR in A375 melanoma cells containing a doxycycline inducible PGR expression and in T47D breast cancer cells (n = 2 independent biological experiments per cell line). (C) RT-PCR for *ddx21* to reveal receptor expression during early development (n = 3 independent experiments).



Supplemental Figure 6. DDX21 relocalizes from the nucleolus to the nucleoplasm upon nucleotide depletion. (A, D) DDX21 and TCOF1 immunofluorescence staining in A375 melanoma cells (n = 4 independent experiments). Scale bars represent 100μm. (B, E) Quantification of nucleoli to nucleoplasm ratio (n = 5 sections per condition, Two-sided Wilcoxon-Mann-Whitney test). Box plots represent median value and 25th and 75th percentiles. Whiskers are minima and maxima. Red asterisks indicate outliers. (C) Western blot analysis for DDX21 in A375 cells treated for 24 hours with DMSO, A77 1726 or A77 1726 plus nucleotides. ACTIN was used as loading control. Immunoblot are representative of at least 2 independent experiments.



Supplemental Figure 7. Genome wide annotation for DDX21 bound regions in A375 cells. (A) Pie chart indicating the location of DDX21 peaks relative to gene locations in A375 melanoma cells treated with DMSO for 24hrs as determined by ChIP-Seq. (B). Gene track of DDX21 binding at 14kb rDNA region in A375 melanoma cells treated for 24 hours with DMSO, A77 1726 or and A77 1726 plus nucleotides.



Supplemental Figure 8. Gene Ontology analysis and PRO-Seq analysis in A375 cells. (A, B) Gene Ontology (GO) term enrichment analysis of genes downregulated (A) and downregulated DDX21 target genes (B) in A375 melanoma cells 48 hours post treatment with A77 1726 (n = 3 independent biological experiments, hypergeometric test and Benjamini-Hochberg correction). The number of genes associated to each GO term is shown at the end of each bar within the graph. (C) PRO-seq in A375 cells. Nascent transcription at the transcription start site (TSS) and at the gene body of DDX21 target and non-DDX21 target genes in cells treated for 24hrs with DMSO, A77 1726 or A77 1726 plus nucleotides (n = 3 biologically independent experiments). (D) Box plot of PRO-Seq signal shows no difference of nascent transcription in DDX21-bound and non-bound genes, as no changes are observed in promoter to gene body ratio between DDX21 targets and non-targets. Box plots represent median value and 25th and 75th percentiles. Whiskers are from 10th and 90th percentile (n = 3 biologically independent experiments).

Supplementary Tables

Supplementary Table 1:
Chemical screen in zebrafish embryos

Supplementary Table 2:
RNA-Seq of whole zebrafish embryos

Supplementary Table 3:
RNA-Seq of Sox10-GFP sorted zebrafish cells

Supplementary Table 4:
RNA-Seq of ***ctr9*** MO injected, Sox10-GFP sorted zebrafish cells

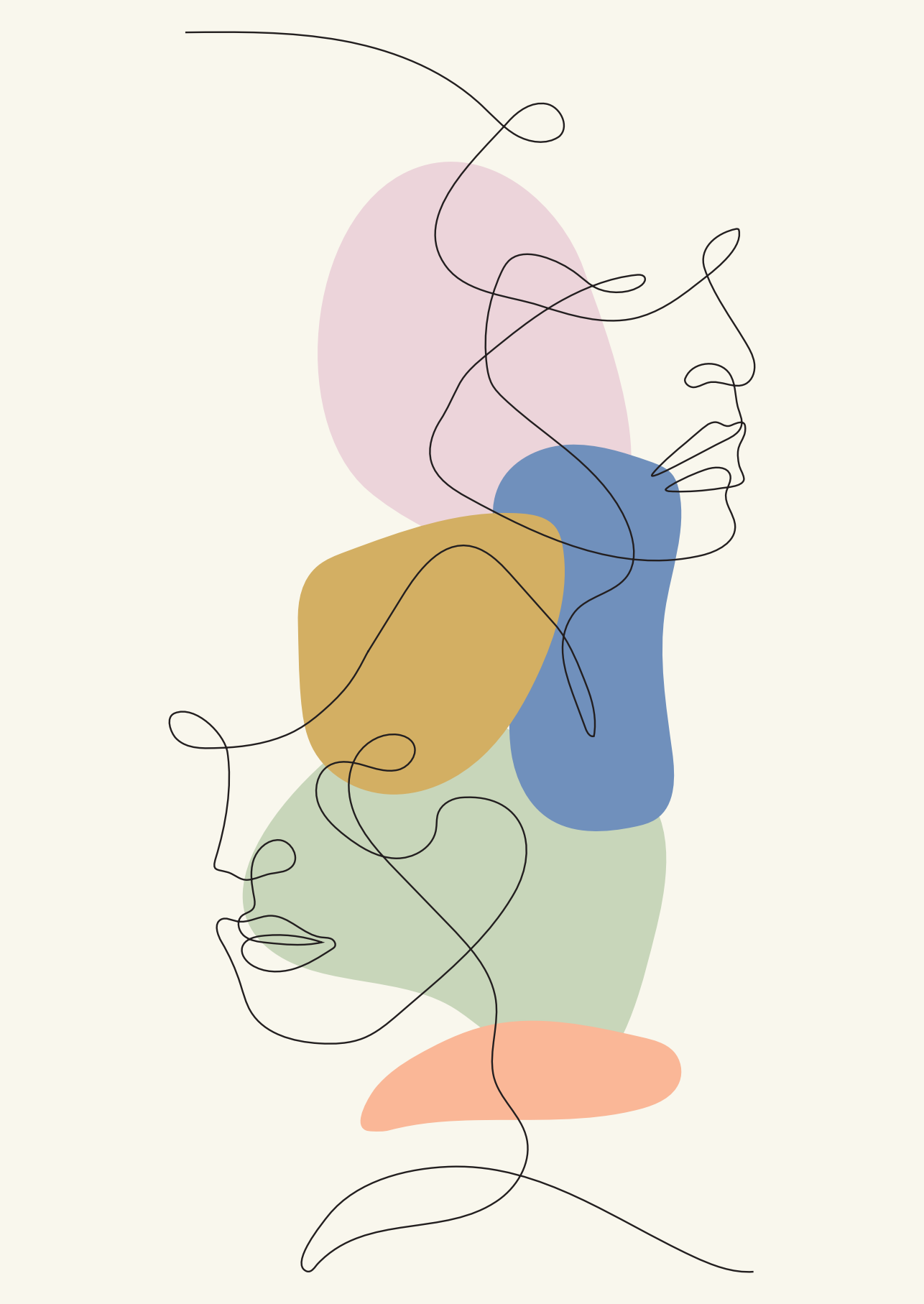
Supplementary Table 5:
Mass Spectrometry for Pgr in zebrafish embryos

Supplementary Table 6:
irCLIP in A375 human melanoma cells

Supplementary Table 7:
Whole Proteomic in A375 human melanoma cells

Supplementary Table 8:
RNA-Seq and ChIP-Seq in A375 human melanoma cells

The supplementary tables are available at:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7185069/



3

Multiomics dissection of the acute molecular stress response to dmPGE$_2$ in human hematopoietic stem and progenitor cells

Chapter 3

A multiomics approach to dissect the acute molecular stress response to dmPGE$_2$ in human hematopoietic stem and progenitor cells

Audrey Sporrij, Eva M. Fast, Margot E. Manning, Marian Kalocsay, Song Yang, Leonard I. Zon#

Correspondence to: leonard.zon@enders.tch.harvard.edu

Abstract

The lipid mediator Prostaglandin E2 (PGE$_2$) and its stable derivative 16,16-dimethyl-PGE$_2$ (dmPGE$_2$) act as important regulators of hematopoietic stem and progenitor cell (HSPC) fate during development, homeostasis and regeneration. Brief *ex vivo* pulse exposure of HSPCs to dmPGE$_2$ increases engraftment following transplantation, highlighting its clinical potential. A detailed understanding of the cellular changes that underlie altered HSPC function after dmPGE$_2$ treatment is currently lacking. Using a multiomics approach, we identified that dmPGE$_2$-induced changes in HSPC homing and proliferation have a transcriptional origin, while enhanced survival is controlled on a post-translational level. The integrated approach used in this study provides the first comprehensive dissection of the transcriptomic, proteomic and metabolomic changes observed in HSPCs in response to the inflammatory stressor dmPGE$_2$.

Introduction

Hematopoietic stem cells (HSCs) give rise to all mature blood lineages and play pivotal roles during development, homeostasis and stress. At steady state, HSCs remain quiescent to minimize exposure to potential harmful conditions. During infection and injury, HSCs respond to environmental signals and reestablish a balanced hematopoietic system through self-renewal and differentiation[1].

Their ability to repopulate hematopoiesis makes HSC therapeutically valuable. Transplantation of hematopoietic stem can progenitor cells (HSPCs) restores functional hematopoiesis in patients suffering from a variety of blood and immune disorders. Despite scientific advances, HSPC transplantation still faces major challenges such as low donor cell numbers or cell quality. These issues compromise long-term engraftment in recipients and thus directly impacts patient outcomes[2]. Many sought to address clinical limitations through *ex vivo* expansion of HSPCs prior to transplantation. Given that stress signals are potent regulators of HSPCs, much interest has been given to modifying HSPC fate by targeting stress pathways[3].

Prostaglandins are physiologically active lipids produced in response to mechanical, chemical or immunological stimuli. They sustain a variety of homeostatic and pathogenic functions. This includes carrying out roles in the inflammatory response to injury or infection[4]. Prostaglandin E$_2$ (PGE$_2$) is one of the most abundant prostaglandins produced in the body[5]. PGE$_2$ and its stable derivative 16,16-dimethyl-PGE$_2$ (dmPGE$_2$) were first identified as regulators of HSPC development and homeostasis in a zebrafish chemical screen[6]. The effects of dmPGE$_2$ were later shown to be conserved across vertebrate species[7,8]. *Ex vivo* pulse exposure of HSPCs to dmPGE$_2$ facilitates *in vivo* engraftment in mice through enhancement of stem cell homing, proliferation and survival. Subsequent non-human primate studies and a phase I clinical study suggests benefits to HSC transplantation outcomes in humans[2].

In HSPCs, dmPGE$_2$ predominantly exerts its effects by binding to PGE$_2$ receptor (EP) subtypes EP2 and EP4[7]. Upon ligand binding, these G-coupled protein receptors enhance intracellular cycling-AMP (cAMP) levels and initiates protein kinase A (PKA) signaling. This cascade mediates activity of numerous downstream factors, including Wnt and β-Catenin, to control cell migration and proliferation[9]. Improved engraftment by dmPGE$_2$ results, in part, from transcriptional regulation of genes implicated in HSPC fate. However, a comprehensive understanding of the cellular changes that underlie altered HSPC function after dmPGE$_2$ treatment is currently lacking. As external stress stimuli are actively studied as modulators of HSPC fate, in-depth knowledge of the cellular response to inductive signals can further advance our efforts to improve transplantation outcomes.

Here, we employ transcriptomics, proteomics, phosphoproteomics and metabolomics to characterize the response to dmPGE$_2$ in human HSPCs. We show that dmPGE$_2$ acts directly on HPSCs by concomitant regulation of multiple cellular processes. Enhanced

HSPC engraftment originates from transcriptional induction of genes implicated in HSPC migration, rapid translation of cell proliferation regulators, and post-translational phosphorylation of anti-apoptosis factors. In conclusion, our comprehensive multiomics approach identifies the complex cellular network that underlies $dmPGE_2$-mediated effects on HSPC fate.

Results

To gain a comprehensive understanding of mechanisms activated by $dmPGE_2$ in human blood cells, we exposed mobilized peripheral blood $CD34^+$ HSPCs or U937 myeloid leukemia cells to 10μM $dmPGE_2$ for 2 hours and performed extensive transcriptome, proteome, phospho-proteome and metabolome profiling (Figure 1).

$dmPGE_2$ induces an acute transcriptional response that modulates HSPC fate

We first aimed to assess the acute transcriptional response to $dmPGE_2$ by performing RNA-sequencing (RNA-Seq) on $CD34^+$ HSPCs after 2 hours of stimulation. We identified a total of 687 consistent differentially expressed genes (DEGs) after $dmPGE_2$ treatment in comparison to vehicle treated (DMSO) control cells (Figure 2A). The effect of $dmPGE_2$ on gene expression was overwhelmingly stimulatory. Specifically, 535 genes were at least 1.5-fold upregulated while 152 genes were greater than 1.5-fold downregulated in expression. Hierarchical clustering and heatmap analysis of the 3 independent biological replicates illustrates the high reproducibility between replicates, while also highlighting expression changes between the treatments (Figure 2B). Among the upregulated set were genes representative of cAMP/PKA signaling including PDE4B and PTGS2[10], cell migration genes such as CXCL2 and CXCL8[11], and genes know to be involved in HSPCs proliferation for instance CCND2, NR4A1 and JUNB[12,13]. Amongst the repressed genes, we found genes associated with self-renewal and cell division, such as HOXB4 and CCNF[14] (Figure 2B, Supplemental Figure 1A). To validate the gene expression changes identified by RNA-Seq, we performed RT-qPCR in HSPCs for representative genes and

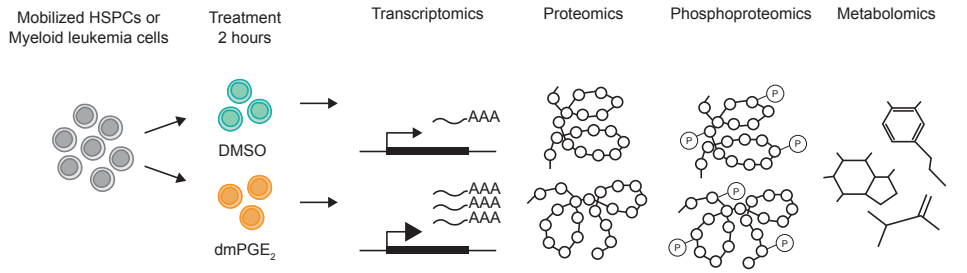


Figure 1. A multiomics approach to dissect to the response to $dmPGE_2$. Schematic representation of the experimental approach used in this study. Cells are stimulated for 2 hours with $dmPGE_2$ or vehicle control (DMSO), unless otherwise indicated.

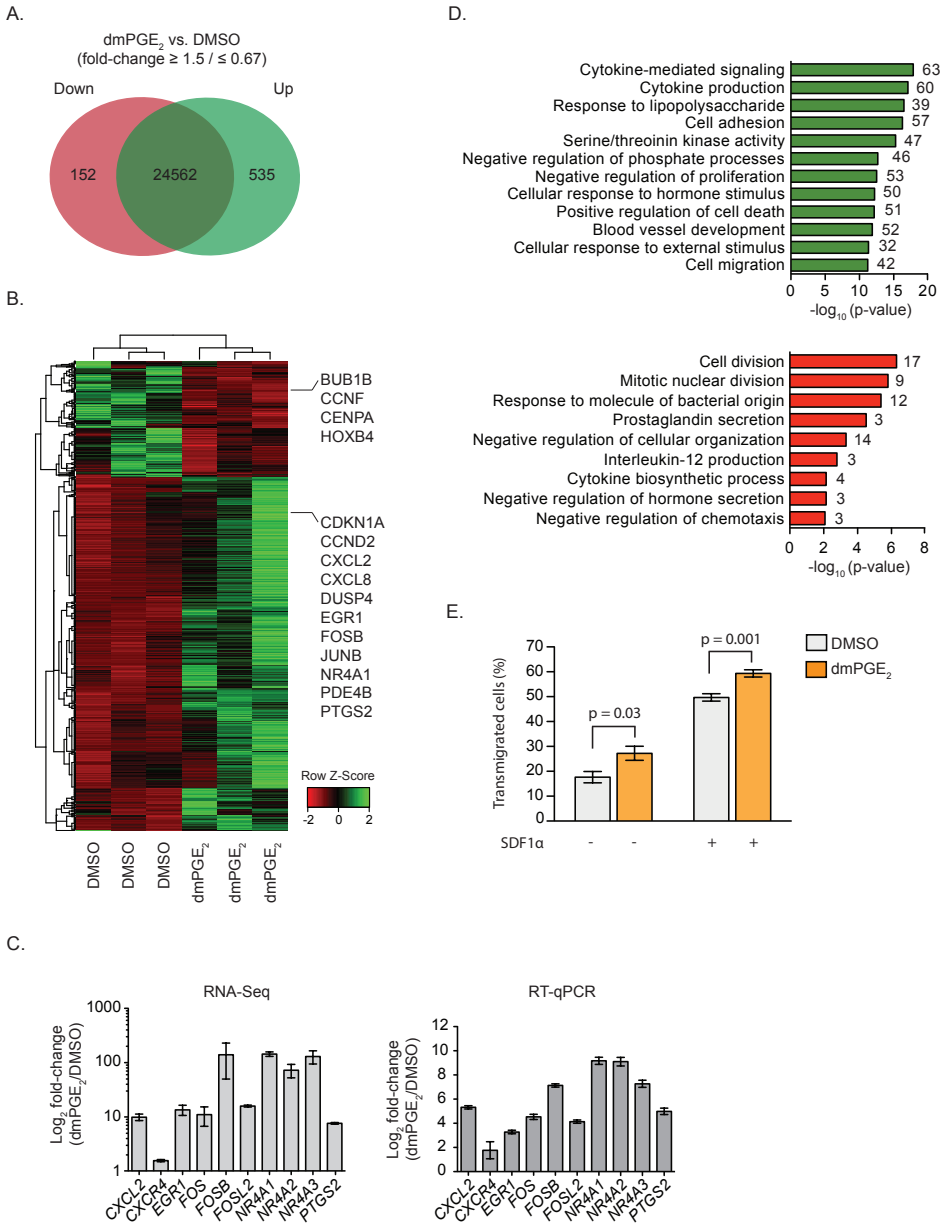


Figure 2. dmPGE$_2$ induces acute transcriptional changes that alter HSPC fate. (A) Number of genes upregulated (535) and downregulated (152) in CD34$^+$ HSPCs after 2-hour treatments, as determined by RNA-Seq. Differentially expressed gene (DEG) criteria: FPKM ≥ 1 after treatment; fold change ≥ 1.5 or ≤ 0.67 (n = 3 biologically independent experiments). (B) Hierarchical clustering heatmap of FPKM values from DEGs (687). (C) RT-qPCR validation in CD34$^+$ HSPCs of DEGs (n = 3 biologically independent experiments; mean values ± SEM). (D) GO term enrichment analysis of genes upregulated (upper panel) and down-regulated (lower panel) in CD34$^+$ HSPCs 2 hours post dmPGE$_2$ treatment. The number of genes associated with each GO term are shown at the end of each bar within the graph. P-values were calculates using hypergeometric test and Benjamini-Hochberg correction. (E) Transwell migration of CD34$^+$ HSPCs exposed to dmPGE$_2$ or DMSO for 2 hours prior to cell seeding. Transwell migration was assessed after 24 hours (n = 3 biologically independent experiments; mean values ± SEM).

found the data to be concurrent (Figure 2C).

Next, we performed gene ontology (GO) term enrichment analysis on upregulated and downregulated genes to understand the biological profile of our differential gene signature (Figure 2D). Enriched biological GO terms associated with upregulated genes predominantly relate to inflammatory stress signaling and cell migration whereas cell division and production of prostaglandin was overrepresented among downregulated genes. These observations align with previously described *in vivo* phenotypes of enhanced HSPC homing and proliferation after $dmPGE_2$ pulse exposure[8]. To assess functional conservation of GO term enriched biological processes and *in vivo* phenotypes, we assessed the migratory response after $dmPGE_2$ treatment using an *in vitro* transwell migration assay. We observed that a 2-hour pulse exposure to $dmPGE_2$ results in greater HSPC migration, when assessed 24 hours post-stimulation (Figure 2E). These data suggest that $dmPGE_2$ enhanced engraftment *in vivo* is, at least in part, driven by transcriptional induction of migration genes.

Given that short-term exposure to $dmPGE_2$ induces long-term effects *in vitro* (Figure 2E) and *in vivo*[2,6,7], we assessed whether $dmPGE_2$-induced differential expression persists in HSPCs. Using RNA-Seq, we investigated whether transcriptional changes induced by 2-hour pulse (T2) of $dmPGE_2$ are retained 24 hours (T24) after exposure (Figure 3A, Supplemental Figure 1B-1D). Assessment of gene expression changes at 24 hours revealed that few T2 DEGs retain their differential expression (Figure 3B). Hierarchical clustering of T2 DEGs at both timepoints further illustrates this observation. T24 $dmPGE_2$ treated and control HSPCs most closely associate with T2 control HSPCs (Figure 3C). More precisely, we found that only 22 of T2 upregulated and 7 of T2 downregulated genes remain increased or decreased in expression at T24, respectively (Figure 3D-3F). One of the genes that remains induced is the chemokine CXCL8. CXCL8 expression peaks between 12h and 24h after induction in response to LPS and hypoxia[15,16]. Similar kinetics may be in play in response to $dmPGE_2$ (Figure 3F). The chemokine was recently identified as a positive regulator of HSPC engraftment in non-cell autonomous manners[17]. Induction of CXCL8 potentially plays an important part in the mechanism by which $dmPGE_2$ promotes HSPC engraftment *in vivo*. The NF-κB subunit RELB also is upregulated 24h post-stimulation with $dmPGE_2$ (Figure 3F). The subunit RelB is known to show slower activation in response to stress since NF-κB controls a delayed response to stress that persists longer after stress sensing[18-21]. RelB has a key role in silencing expression of the pro-inflammatory cytokines TNF-α and IL-1β. The transcription factor KLF5 associates with NF-κB to regulate genes involved in inflammation and is known as a second phase gene[22-24]. This could explain the similarity in transcriptional kinetics of KLF5 and RELB. BMP6, a members of the transforming growth factor β (TGF-β) superfamily, that is further upregulated 24h post $dmPGE_2$ stimulation has been described to suppress inflammation *in vivo*[25-27] (Figure 3F). Induction of anti-inflammatory mediators at 24 hours post-stimulation presents a negative feedback mechanism to suppress

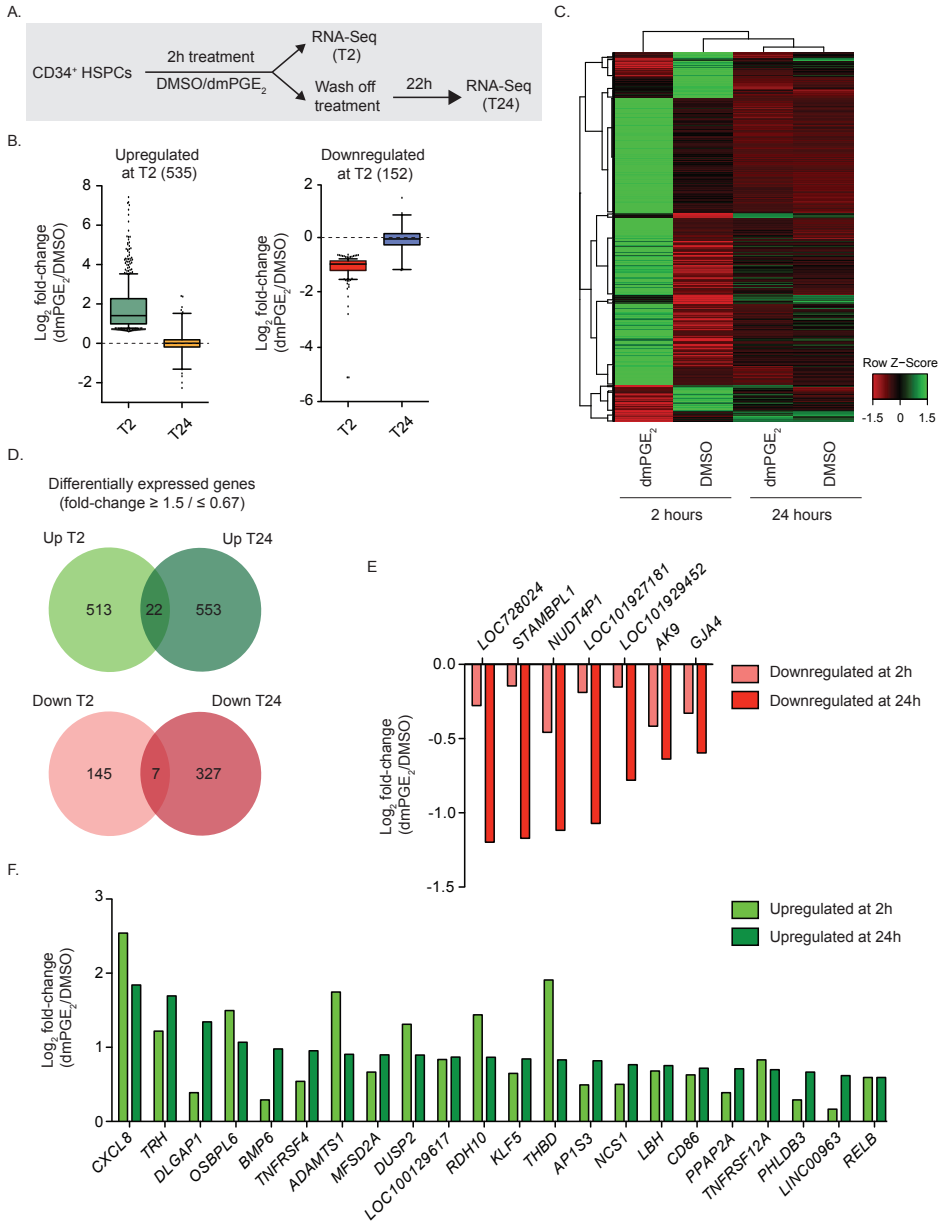


Figure 3. Low retention of transcriptional changes 24 hours after dmPGE$_2$ stimulation in HSPCs. (A) Schematic representation of the experimental approach used to study gene expression 24 hours after dmPGE$_2$ treatment. CD34$^+$ HSPCs were exposed to dmPGE$_2$ or DMSO for 2h after which stimuli were washed out. Cells were then harvested for RNA-Seq analysis (T2) or placed back in culture media for an additional 22 hours, after which they were harvest for RNA-Seq analysis (T24). (B) Comparison of gene expression changes of T2 DEGs (687) at 2 hours and 24 hours. Box plots shows median, 25th and 75th percentiles. Whiskers are from 10th to 90th percentile. (C) Hierarchical clustering heatmap of FPKM values of T2 DEGs (687) at 2 hours and 24 hours post-stimulation. (D) Venn diagram showing overlap between the genes upregulated (upper panel, 22) and downregulated (lower panel, 7) at both timepoints. DEG criteria: FPKM ≥ 1 after treatment; fold change ≥ 1.5 or ≤ 0.67. (E, F) Expression level changes of genes that remain downregulated (E) and upregulated (F) at 2 hours and 24 hours after dmPGE$_2$ treatment, as determined by RNA-Seq.

persistent inflammatory signaling. Together, these data indicate that dmPGE$_2$ induces an acute transcriptional response that supports cell migration. While gene expression changes are predominantly transient, we observe sustained induction of genes involved in re-establishing homeostasis after inflammatory stress.

Acute transcriptional changes are reflected in the cellular proteome

To determine if gene expression changes are reflected on a protein level, we performed proteomics on dmPGE$_2$ and DMSO treated U937 myeloid leukemia cells. U937 cells show a stress response to dmPGE$_2$ that is highly similar to HSPCs on a molecular and functional level (Supplemental Figure 2). We used the iTRAQ method to determine differential peptide levels by mass spectroscopy. We identified a total of 8883 common proteins in duplicate experiments. 194 (2%) proteins showed a statistically significant (p-value ≤ 0.05) change in expression, of which 113 proteins (58%) were more abundant and 81 proteins (42%) were less abundant in dmPGE$_2$ exposed cells (Figure 4A). Only 13 out of the 133 increased proteins and 3 out of the 81 decreased proteins show a change in protein ratio of ≥ 1.5 or ≤ 0.67 after 2 hours of dmPGE$_2$, (Figure 4B, Supplemental Figure 3A).

We found profound changes in proteins encoded by well-known immediate early genes (IEGs)[28]. IEGs are rapidly transcriptionally activated and translated in response to a variety of cellular stimuli. Indeed, proteins ≥ 1.5-fold induced show gene expression changes in the same direction (Supplemental Figure 3B). Proteins encoded by IEG and upregulated by dmPGE$_2$ include the transcription factors NR4A1, FOS, FOSB, and JUNB, which are important regulators of HSPC proliferation[12,29,30]. Besides transcription factors, IEGs encode other functional proteins found to be induced by dmPGE$_2$. We observed protein enrichment of the growth factor OSM[31], signal transduction regulator RGS2[32], and the cell cycle regulators CDKN1A (p21$^{cip1/waf1}$)[33,34] and BTG1[35]. This data indicates that increased proliferation observed in response to dmPGE$_2$ is mediated through acute transcriptional and translational regulation of cell cycle mediators.

KLF6, GPD1 and ALG3 were significantly less abundant in HSPCs after pulse exposure to dmPGE$_2$. *KLF6* is transcriptionally induced by inflammatory stress and the transcript is efficiently translated into protein. However, maximal expression of *KLF6* mRNA and protein levels are reached 1h after stress stimulation, after which protein expression goes down[23]. Similar kinetics are likely in place in response to dmPGE$_2$, where KLF6 protein levels are decreased 2h post-stimulation. GPD1 is a rate-limiting enzyme in glycerol formation. It catalyzes the conversion of dihydroxyacetone phosphate (DHAP) and nicotine adenine dinucleotide (NADH) to glycerol-3-phosphate (G3P) and NAD+. Decreased GPD1 protein expression limits growth in anaerobic environments[36,37]. Lower GPD1 protein levels suggest metabolic adaptation to altered environmental conditions. ALG3 is a mannosyltransferase involved in N-linked glycosylation. Downregulation of

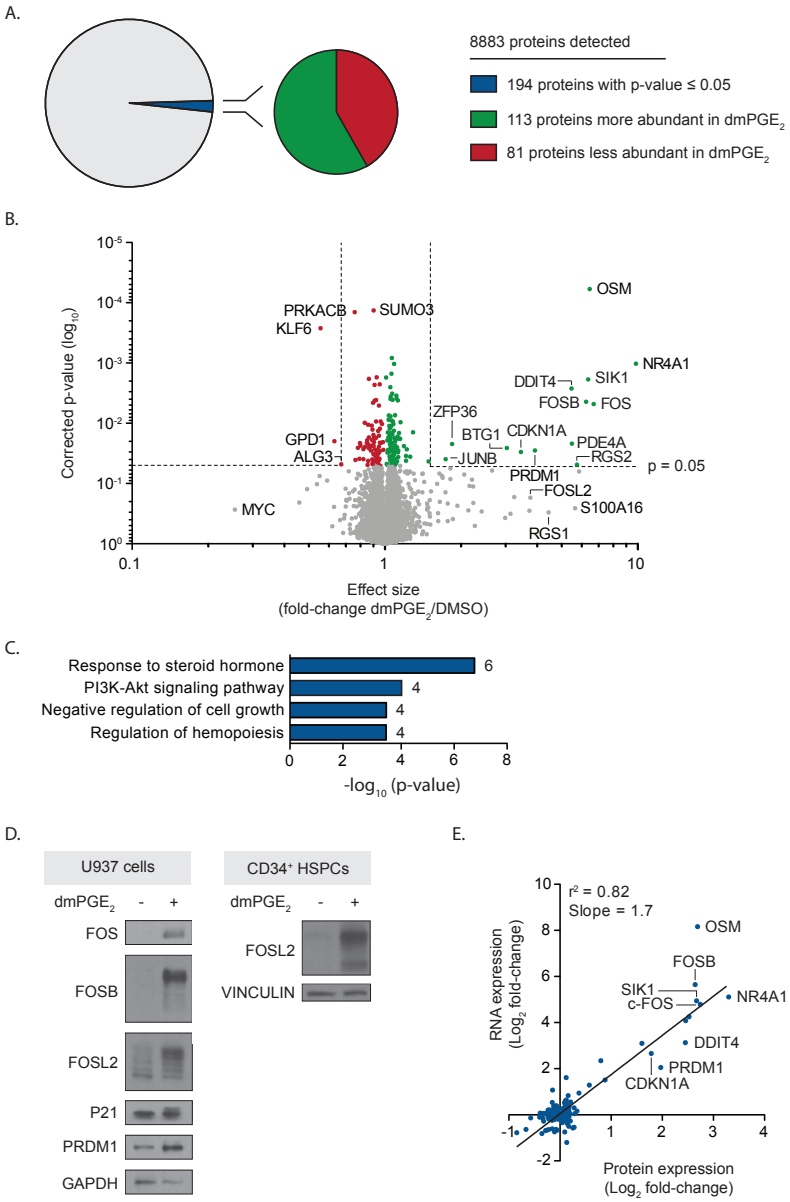


Figure 4. Acute transcriptional changes are reflected in the cellular proteome. (A) Mass-spectrometry based cellular proteome analysis of U937 cells identified 194 (2%, in blue) out of 8883 detected proteins with a significant differential abundance after 2 hours of dmPGE$_2$ treatment. 113 (58%, in green) presented with higher protein levels whereas 81 (42%) presented lower proteins levels (n = 2 biologically independent experiments). (B) Volcano plot of TMT mass-spectrometry results showing changes in protein enrichment. Proteins with a p-value ≤ 0.05 (194) are marked in red (downregulated) and green (upregulated). Higher effect size represents higher enrichment in dmPGE$_2$. P-value was determined by t-test and Benjamini-Hochberg corrected. (C) GO term enrichment analysis of significantly differential expressed proteins (194). The number of genes associated with each GO term are shown at the end of each bar within the graph. P-values were calculates using hypergeometric test and Benjamini-Hochberg correction. (D) Western blot analysis in U937 cells and CD34$^+$ HSPCs stimulated with vehicle control or dmPGE$_2$ for 2 hours. GAPDH and VINCULIN are used as loading controls. (E) Linear regression analysis of protein and mRNA fold changes in U937 cells. Values were plotted of significant differentially abundant proteins (p-value ≤ 0.05; 194).

ALG3 inhibits cell proliferation and migration *in vivo*[38]. Decreased protein expression of GPD1 and ALG3 may contribute to reestablishing homeostasis upon dmPGE$_2$ treatment in HSPCs.

GO term analysis of differential expressed proteins highlights biological processes related to cell signaling and cell growth (Figure 4C). This reveals that activation of cellular adaptation pathways in response to dmPGE$_2$ is reflected on a protein level.

We validated altered protein abundance of several peptides by western blot and found induction to be conserved in CD34$^+$ HSPCs (Figure 4D). Detectable enrichment of FOSL2 levels indicates that also non-significantly ($p > 0.05$) induced proteins can potentially contribute to biological response to dmPGE$_2$.

The observation that proteins with ≥ 1.5-fold or ≤ 0.67-fold abundance showed altered gene expression (Supplemental Figure 3B), prompted us to compare whole proteome to transcriptome changes. We identified a strong, positive ($R^2 = 0.82$, Slope = 1,7) correlation between significant proteins (n = 194) and expression changes of the corresponding genes 2 hours after dmPGE$_2$ treatment (Figure 4E). The positive relation between DEGs (n = 687) and corresponding protein levels was less profound (Supplemental Figure 3C, 3D). This is expected due to the temporal delay between transcription and translation that becomes especially eminent during dynamic transcriptional states[39].

Altogether, our results demonstrate that dmPGE$_2$ affects the cellular transcriptome and proteome. Acute changes are more pronounced at the mRNA level rather than in protein abundance. The immediate changes that are observed in the cellular proteome upon dmPGE$_2$ are driven by modulation of transcription of the corresponding genes.

dmPGE$_2$-mediated apoptotic effects are post-translationally regulated

Previous studies revealed that dmPGE$_2$ exerts its effects on HSPCs through cAMP signaling and activation of PKA, which regulates intracellular phosphorylation. To elucidate phosphorylation events following dmPGE$_2$ on a proteome-wide scale, we performed duplicate mass-spectrometry based phosphoproteomic profiling in U937 cells stimulated with dmPGE$_2$.

We quantified a total of 2609 phosphopeptides, of which 167 proteins (6%) were significantly (p-value ≤ 0.05) altered in abundance 2 hours post-dmPGE$_2$ stimulation. Specifically, 45 phosphoproteins (27%) were more abundant whereas 122 phosphopeptides (73%) were less abundant after dmPGE$_2$ treatment (Figure 5A). 37 out of the 45 increased phosphoproteins and 66 out of the 122 decreased phosphoproteins have a fold change in abundance of ≥ 1.5 or ≤ 0.67 (Figure 5B, Supplemental Figure 3E).

Ontology analysis on phosphoproteins with a differential presence of ≥ 1.5-fold or ≤ 0.67-fold revealed enrichment for several biological processes (Figure 5C). In line with transcriptional and translational effects, we found enrichment for phosphoproteins

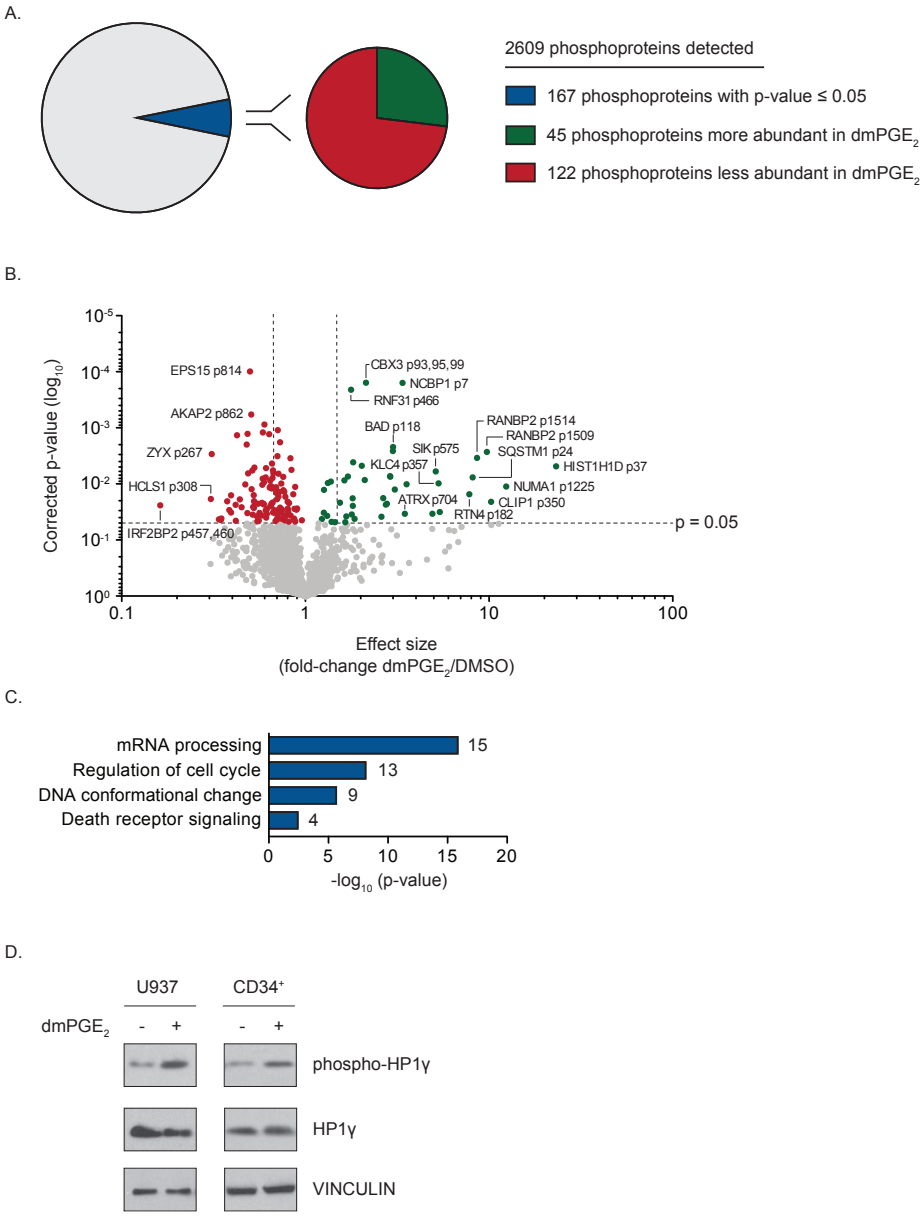


Figure 5. Rapid phosphorylation of chromatin regulators in response to dmPGE$_2$. (A) Mass-spectrometry based cellular phosphoproteome analysis of U937 cells identified 167 (6%, in blue) out of 2609 detected proteins with a significant differential abundance after 2 hours of dmPGE$_2$ treatment. 45 (27%, in green) presented with higher protein levels whereas 122 (73%) presented lower proteins levels (n = 2 biologically independent experiments). (B) Volcano plot of TMT mass-spectrometry results showing changes in phosphoprotein enrichment. Phosphoproteins with a p-value ≤ 0.05 (194) are marked in red (downregulated) and green (upregulated). Higher effect size represents higher enrichment in dmPGE$_2$. P-value was determined by t-test and Benjamini-Hochberg corrected. (C) GO term enrichment analysis of significantly differential expressed proteins (167). The number of genes associated with each GO term are shown at the end of each bar within the graph. P-values were calculates using hypergeometric test and Benjamini-Hochberg correction. (D) Western blot analysis in U937 cells and CD34$^+$ HSPCs. VINCULIN is used as a loading control.

associated with mRNA processing and cell cycle. Phosphoproteome analysis further identified novel roles for $dmPGE_2$ in biological processes not identified by transcriptome of proteome studies.

We found phosphoproteins related to DNA conformation such as SETD1A, CBX3, ATRX, and HIST1H1D. SETD1A is a methyltransferase that catalyzes methylation of histone 3 lysine 4, a post-translational modification associated with transcriptional regulation[40-42]. While the precise effects of post-translational modifications of SETD1A are unknown, phosphorylation is suggested to affect its function[43]. CBX3, also known as heterochromatin protein 1 gamma (HP1γ), is a transcriptional repressor that binds methylated lysine 9 resides of histone H3. Phosphorylation of HP1 by protein kinase A reduces its affinity for H3K9me leading to the transient displacement of HP1[44,45]. We confirmed increased levels of phospho-HP1 by western blot and furthermore showed conservation of protein induction in $CD34^+$ HSPCs (Figure 5D). ATRX belongs to the SWI/SNF family of chromatin remodelers and is involved in deposition and remodeling of H3.3-containing nucleosomes. ATRX phosphorylation correlates with chromatin association and is concomitant with its binding to HP1. A role in gene regulation during cell cycle progression has been attributed to phosphorylated ATRX[46]. Linker histone H1.3 (HIST1H1D) interacts with linker DNA between nucleosomes and functions in the compaction of chromatin into higher order structures. Linker histone H1 phosphorylation impairs chromatin aggregation and leads to local chromatin relaxation[47]. Our phosphoproteome data indicates that epigenetic regulation of chromatin factors through post-translational modifications may regulate $dmPGE_2$-induced transcriptional changes.

We also found increased phosphoprotein levels of cell death regulators. This includes BAD. BAD mediates apoptosis by binding to pro-apoptotic regulators and is controlled by multiple kinases and phosphatases[48-50]. Phosphorylation of BAD inhibits dimerization with Bcl-2 and Bcl-xL hereby preventing Bax/Bak-mediated apoptosis. BAD phosphorylation is thus anti-apoptotic. Phosphorylation of BAD at serine 118, as is seen upon $dmPGE_2$, is a main target of PKA and inhibits apoptosis[51]. Assessment of apoptosis by flow cytometry revealed that indeed $dmPGE_2$ decreases apoptosis in U937 cells by nearly 40% (Figure 6D, 6E). The reduction in cell death conserved in $CD34^+$ HSPCs (Figure 6F). Blockage of transcription using actinomycin D (AD) or inhibition of protein translation using cycloheximide (CHX) during $dmPGE_2$ exposure does not impact the anti-apoptotic effects observed after stimulation (Figure 6A-6E). This work reveals that increased cell survival of human blood cells after $dmPGE_2$ treatment is regulated on a post-translational level.

Metabolome profiling indicates epigenetic regulation of the $dmPGE_2$ response

To identify metabolic pathways critical for early adaptation to $dmPGE_2$, we assessed over 300 metabolites by mass-spectrometry (Figure 7A). Analysis with a minimum cutoff of 1.2-fold change and p-value of ≤ 0.05 showed 15 metabolites (5%) with significant differential

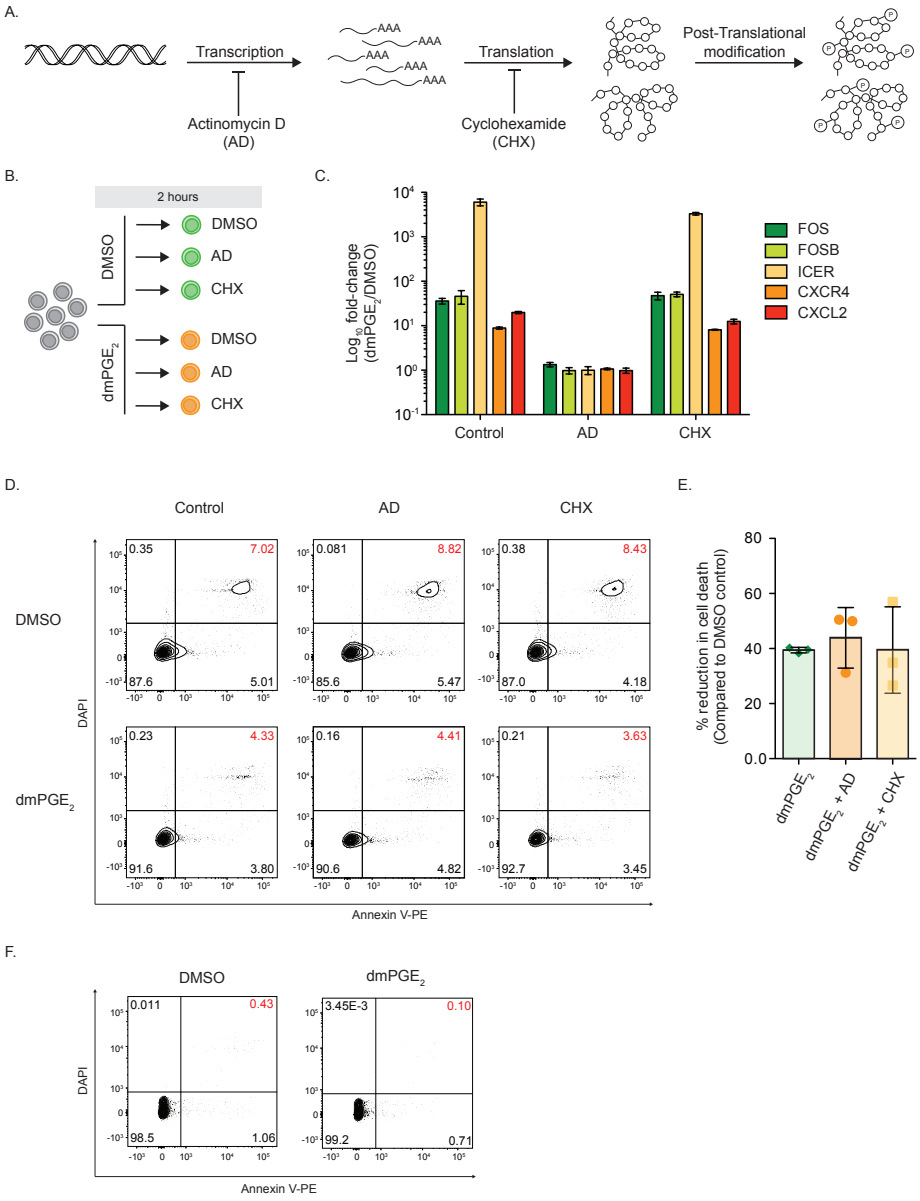


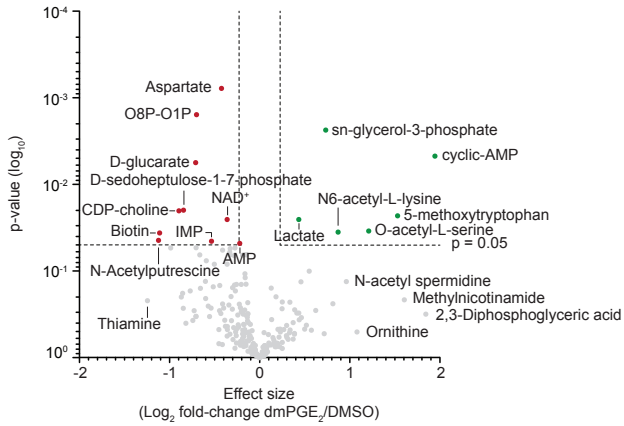
Figure 6. Anti-apoptotic effects of dmPGE$_2$ are post-translationally regulated. (A) Illustration of the mechanisms of action of actinomycin D (AD) and Cyclohexamide (CHX), inhibiting transcription and translation respectively. Post-translational phosphorylations remain unaffected by AD and CHX. (B) Schematic representation of experimental set up in U937 cells. Cells were treated for 2 hours with dmPGE$_2$ or DMSO in the presence or absence or AD or CHX. (C) RT-qPCR analysis for selected dmPGE$_2$ response genes shows efficient inhibition of transcriptional induction by AD, whereas gene expression changes remain unaffected by CHX treatment (n = 3 biologically independent experiments). The data are shown as the mean ± SD. (D) Representative FACS plots of cell apoptosis analysis in U937 cells by Annexin V/DAPI flow cytometry after 2 hours of treatment with DMSO or dmPGE$_2$ in the presence or absence of AD or CHX. (E) Quantification of cell death reduction relative to the corresponding DMSO control, as determined by Annexin-V/DAPI FACS analysis (n = 3 biologically independent experiments). The data are shown as the mean ± SD. (F) Representative FACS plots of cell apoptosis analysis in CD34$^+$ HSPCs by Annexin V/DAPI flow cytometry after 2 hours of treatment with DMSO or dmPGE$_2$.

presence in $dmPGE_2$ treated HSPCs, compared to control HSPCs (Figure 7B).
A noticeable observation was a near 4-fold increase in cyclic-AMP (cAMP) in response to $dmPGE_2$. This is in line with previous observations that $dmPGE_2$ mediates cellular effects in part via cAMP signaling in HSPCs. 5-methoxytryptophan (5-MTP) is a known suppressor of prostaglandin synthase 2/cyclooxygenase 2 (PTGS2/COX2) expression[52]. PTGS2 is an enzyme responsible for the majority of prostaglandin synthesis during inflammation. Upregulation of 5-MTP in HSPCs after $dmPGE_2$ treatment therefore presents an endogenous negative feedback mechanism to attenuate stress signaling after exposure to inflammatory stimuli. Negative regulation of PTGS2/COX2 by 5-MTP is further supported by GO term enrichment for prostaglandin secretion in the T2 downregulated gene set (Figure 2D). Changes in metabolism often result in altered epigenetic regulation[53]. We were particularly intrigued by the increased presence of N6-acetyl-L-lysine (acetyl-lysine) after $dmPGE_2$. Acetyl-lysine is actively associated with transcription in eukaryotes. Post-translational modification by lysine acetylation of transcription factors (TFs) and chromatin factors affects gene expression levels. This data further strengthens our indication of $dmPGE_2$-mediated epigenetic regulation of transcription.
Having determined the differential levels of metabolites during stress, we next subjected the data to metabolic pathway impact analysis to identify the relevant pathways that were perturbed. We considered metabolic pathways with pathway impact values of ≥ 0.1 and p-value of ≤ 0.05 be significantly perturbed[54,55] (Figure 7C). The most significantly altered pathways were nicotinate and nicotinamide metabolism, and glycerophospholipid metabolism. Nicotinate and nicotinamide metabolism can inhibit the release of inflammatory factors so as to achieve the effect of anti-inflammation, through regulation of NF-κB activity[56,57]. *Ex vivo* culture of HSPCs in the presence of nicotinamide facilitates their expansion and increases engraftment efficiency[58,59]. Activation of the nicotinate and nicotinamide metabolism by $dmPGE_2$ can contribute to the enhanced HSPC function observed after transplantation. Altered glycosphingolipid and glycerophospholipid metabolism, can drive inflammatory cascades and exert protective effects in relation to tissue damage[60-62]. These data provide a link between the transcriptional stress response to $dmPGE_2$, epigenetic regulation and metabolism in HSPCs. $dmPGE_2$ induces cellular adaptation at a transcriptional, translational, post-translational and metabolic level.

Discussion

Hematopoietic repopulation after transplantation is a multistep process directly affected by the ability of HSPCs to engraft and self-renewal in the bone marrow niche. $dmPGE_2$ was previously shown to increase homing, survival and proliferation of HSPCs. Short-term *ex vivo* pulse exposure to $dmPGE_2$ demonstrated direct and stable beneficial effects on HSPC function after transplantation *in vivo*. Here, we combined transcriptome, proteome, phosphoproteome and metabolome analysis in human $CD34^+$ HSPCs and U937 myeloid

A.



B.

Metabolite Name	DMSO	dmPGE$_2$	FC
cyclic-AMP	7800	30110	3.9
5-methoxytryptophan	4458	12898	2.9
O-acetyl-L-serine	10166	23853	2.3
N6-acetyl-L-lysine	41973	76871	1.8
sn-glycerol-3-phosphate	291829	485502	1.7
Lactate	12522198	16938007	1.4
NAD+	961049	748373	0.8
Aspartate	7143370	5333886	0.7
IMP	102542	70833	0.7
O8P-O1P	73811	45494	0.6
D-glucarate	14559	8915	0.6
D-sedoheptulose-1-7-phosphate	293635	163560	0.6
CDP-choline	30359	15327	0.5
Biotin	42197	19569	0.5
N-Acetylputrescine	49478	22759	0.5

C.

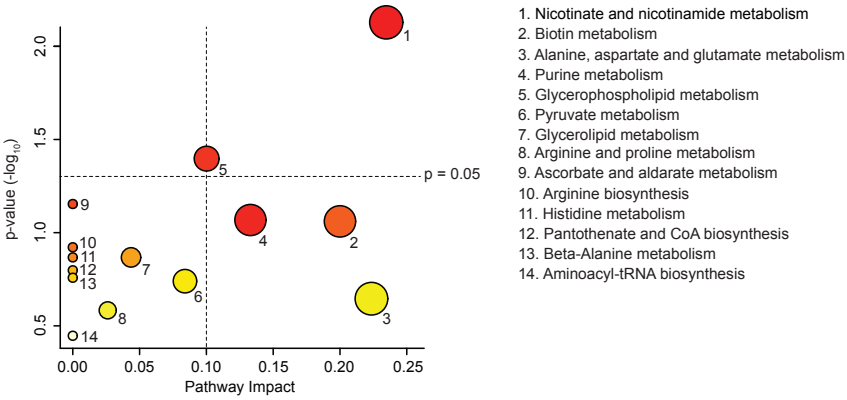


Figure 7. Metabolome profiling indicates epigenetic regulation of the dmPGE$_2$ response. (A) Volcano plot of metabolomics analysis showing changes in metabolite levels. Metabolomic analysis of CD34$^+$ HSPCs identified 15 (5%) out of 303 detected metabolites with a significant differential abundance after 2 hours of dmPGE$_2$ treatment. Metabolites with an effect size ≥ 1.2 or ≤ 0.83 and a p-value ≤ 0.05 are marked in red (downregulated, 9) and green (upregulated, 6). Higher effect size represents higher enrichment in dmPGE$_2$. P-value was determined by t-test (n = 3 biologically independent experiments). (B) A table summarizing the 15 identified metabolites that were significantly altered in dmPGE$_2$, their median corrected abundance and effect size. Fold change (FC) is presented as median average of dmPGE$_2$ values divided by DMSO control values. (C) Metabolome view of the pathway analysis considering significantly altered metabolites with quantity fold change ≥ 1.2 or ≤ 0.83. The most significant pathways are represented. Larger circles or circle in darker colors (red > orange > yellow) indicated higher impact of the metabolic pathways.

leukemia cells, and revealed the cellular responses that underlie the effects of $dmPGE_2$ on HSPCs.

Acute transcriptional reprogramming in response to stress is one of the main mechanisms of cellular adaptation[63]. We determined the specific transcriptional program that is rapidly provoked in response to $dmPGE_2$ in HSPCs. We find that enhanced homing is driven by changes in expression of genes involved cell migration. Using RNA-Seq analysis, we identified significant upregulation of *CXCL1*, *CXCL2*, and *CXCL8*, amongst others, in response to $dmPGE_2$. HSPC mobilization upon inflammatory stress has been observed in the bone marrow. Stress-activated HSPCs migrate from the osteoblastic niche to the closest sinusoidal, vascular niche to produce new blood cells that directly enter circulation[64,65]. This migratory phenotype evoked by $dmPGE_2$ may represent the natural response of HSPCs to inflammatory stresses as is seen when residing in their natural bone marrow niches.

While the majority of acute transcriptome changes in migratory genes after $dmPGE_2$ treatment are transient, we find *CXCL8* to show sustained upregulation 24h post-stimulation. CXCL8 signaling acts as a positive non-autonomous regulator of HSPC colonization during development and during engraftment upon transplantation *in vivo*[17]. This is established through binding of CXCL8 to its receptors CXCR1 and CXCR2, which are strongly expressed on endothelial cells. CXCR1 and CXCR2 receptor activation induces functional and structural changes in the stem cell niche that are favorable to HSPC colonization. Sustained induction of *CXCL8* by $dmPGE_2$ may enhance engraftment by priming of the bone marrow niche. We also found sustained upregulation of BMP by $dmPGE_2$. HSC-derived BMP-6 induces differentiation of mesenchymal cell to osteoblasts, which play a central role in the stem cell niche[66].These findings suggest that $dmPGE_2$-mediated long-term gene expression changes direct the formation and/or remodeling of the bone marrow niche to support HSPC colonization[67-69].

Gene expression changes also underlie cell cycle effects seen after $dmPGE_2$ stimulation. We found transcriptional changes in regulators of stem cell proliferation, including *NR4A1/2/3*, *EGR1*, *FOS*, *FOSB*, and *CDKN1A*. These gene are well described immediate early genes and the induction of their corresponding protein levels as identified by proteomics analysis is therefore expected. By controlling cycle entry and progression, these factors affect the balance between HSPC activation and quiescence[12,29,34,70,71]. Many inflammatory stress signals, including $dmPGE_2$, induce transient HSPC proliferation to increase the output of progenitors and subsequent new blood cells[8,72]. We also found upregulation of the anti-proliferative proteins CDKN1A ($p21^{cip1/waf1}$) and BTG[35,73]. We suggest that upregulation of these factors allows for activated HSCs to return back to a quiescent state once homeostasis is re-established. HSPCs reversibly switch between quiescence and proliferation during stress[74]. Maintenance of HSPC quiescence is crucial to prevent premature stem cell exhaustion and ensure long-term hematopoietic repopulation. The ability of $dmPGE_2$ to transcriptionally regulate HSPC proliferation may

contribute to enhanced, long-term engraftment seen after transplantation.

Stress-induced transcriptional responses are the result of the dynamic regulation of chromatin structure. Using phosphoproteomics and metabolomics, we found that dmPGE$_2$ likely mediates transcriptional responses through regulation of chromatin organization. Analysis of significantly altered phosphoproteins revealed enrichment in factors associated with chromatin conformation, such as SETD1A, CBX3, ATRX, and HIST1H1D. Post-translational phosphorylation of these factors impairs chromatin condensation, which is concomitant with transcriptional activation. Increased levels of acetylated lysine are further indicative of dmPGE$_2$-mediated epigenetic regulation in HSPCs. Lysine acetylation has emerged as a major post-translational modification for transcription factors and histones. The activity of numerous stress-related TFs is impacted by lysine acetylation, which promotes nuclear localization, TF dimerization, and DNA binding affinity[75]. Additionally, acetylation of lysine residues on histone tails reduces DNA affinity and alters nucleosome conformations[76]. The resulting increase in chromatin accessibility is associated with enhanced transcriptional activity in eukaryotic cells[77], and is in line with observed genes expression changes in HSPCs exposed to dmPGE$_2$.

Apoptosis is an important regulator of HSPCs and directly impacts transplantation outcomes[78]. Transient inhibition of apoptosis decreases the number of donor cells that are lost and improves engraftment[79]. Surprisingly, we found that the anti-apoptotic phenotype observed in HSPCs is regulated through post-translational mechanisms[80]. Loss of transcriptional or translational activity does not impede anti-apoptotic effects of dmPGE$_2$. Environmental signals activate kinases that inhibit components of the apoptotic pathway through post-translational phosphorylation. Phosphorylation of BAD suppresses its pro-apoptotic function, representing a transcription-independent manner to inhibit cell death[81,82]. Stress-induced promotion of cell survival has been observed in hematopoietic cells, where inflammatory signaling by the cytokine IL-3 induces BAD phosphorylation through protein kinase B (PKB, also known as AKT)[83,84]. We propose that the dynamics regulation of BAD by dmPGE$_2$ present a mechanism to rapidly affect HSPC viability. dmPGE$_2$ enhanced HSPC survival improves long-term engraftment and repopulation after transplantation.

In conclusion, the work presented here indicates that dmPGE$_2$-induced changes in homing and proliferation originate from transcriptional alterations. The acute, transient induction of migration and cell cycle genes, combined with sustained upregulation of factors that prime the bone marrow niche for HSPCs colonization, promote stem cell homing and engraftment upon transplantation. Our work furthermore reveals that enhanced survival of HSPCs after pulse exposure to dmPGE$_2$ is controlled on a post-translational level. The integrated multiomics approach used in this study provides the first comprehensive dissection of the transcriptome-, proteome- and metabolome-wide changes seen in HSPCs in response to the inflammatory stressor dmPGE$_2$.

Methods

Expansion of CD34⁺ cells

Human CD34⁺ hematopoietic stem and progenitor cells (HSPCs), isolated from peripheral blood of granulocyte colony-stimulating factor (G-CSF) mobilized healthy volunteers, were purchased from the Fred Hutchinson Cancer Research Center. The cells were maintained in suspension culture as previously described by Trompouki et al., 2011. Briefly, cells were expanded in StemSpan medium (Stem Cell Technologies Inc.) supplemented with StemSpan CC100 cytokine mix (Stem Cell Technologies Inc.) and 2% Penicillin-Streptomycin for a total of 6 days.

Cell culture

U937 cells were maintained in suspension culture in RPMI-1640 supplemented with 10% (v/v) heat-inactivated fetal bovine serum, 1X Glutamax and 1% Penicillin-Streptomycin at 37°C in a humidified atmosphere of 5% CO_2.

$dmPGE_2$ treatment

16,16-dimethyl prostaglandin E_2 ($dmPGE_2$) was purchased reconstituted in DMSO from Cayman Chemicals (cat. #14750), aliquoted and stored in -80°C until use. Cells were counted, collected and resuspended in StemSpan medium with 2% PenStrep (CD34+ Cells) or RPMI with 1% Penicillin-Streptomycin, but in the absence of additional cytokines or growth factors. Cells were treated with $dmPGE_2$ (Cayman chemicals) or DMSO (vehicle control) for 2 hours. In indicated experiments, Actinomycin D and Cyclohexamide was added to cells at the same time as $dmPGE_2$ or DMSO. Treatment with Actinomycin D or Cyclohexamide was performed at final concentration of 5µg/ml and 10µg/ml, respectively.

Transwell migration assay

In vitro cell migration was determined using a 2-chamber Costar Transwell (6.5-mm diameter, 5-µm pore; Cambridge, MA). $dmPGE_2$ and DMSO treated CD34⁺ HSPCs were cultured and treated as described above. After 2 hours of treatment, cells were spun down and placed in fresh media. Cells were then added to the top chamber of the transwell system, with or without recombinant human SDF1α in the bottom chamber, and incubated at 37°C for 24 hours. Cells transmigrated to the bottom chamber were enumerated by CellTiter-Glo luminescent Cell Viability Assay (Progema). Percentage migration was calculated by dividing cells migrated into the lower well by the total cell input.

qPCR analysis

RNA was extracted from using the RNeasy plus mini kit (Qiagen). cDNA synthesis was performed using the Superscript VILO (Invitrogen) and using equal amounts of starting

RNA. The cDNA was analyzed with the Light Cycler 480 II SYBR green master mix (Applied Biosystems), and the QuantStudio 12K Flex (Applied Biosystems). All samples were prepared in triplicate. The PCR cycle conditions used are: (a) 95°C for 5 min, (b) [95°C for 10 sec, 54°C for 10 sec, 72°C for 15 sec] X 40 cycles. The analysis of Ct values was performed using 2^-ΔΔT method. The PCR primer-pairs used are:

CXCL2	Fwd: AAACCGAAGTCATAGCCACACTC
	Rev: AGCCACCAATAAGCTTCCTCCTTC
CXCR4	Fwd: CCTATGCAAGGCAGTCCATGT
	Rev: GGTAGCGGTCCAGACTGATGA
EGR1	Fwd: GCGAGCAGCCCTACGAGCAC
	Rev: TGCAGGCTCCAGGGAAAAGC
FOS	Fwd: TGCCTCTCCTCAATGACCCTGA
	Rev: ATAGGTCCATGTCTGGCACGGA
FOSB	Fwd: TTCTGACTGTCCCTGCCAAT
	Rev: CGGGGTCAGATGCAAAATAC
FOSL2	Fwd: GCAGTTGGGTTTCTGGCTTGAG
	Rev: TCCTGCTACTCCTGGCTCATTC
GAPDH	Fwd: GAAGGTCGGAGTCAACGGATTT
	Rev: GAATTTGCCATGGGTGGAAT
ICER	Fwd: CACCATGGCTGTAACTGGAGATGAC
	Rev: AGGTCCAAGTCAAAGACAGTTACTC
NR4A1	Fwd: GAGTGCACAGAAGAACT
	Rev: CACAGGAGGAGGAAGA
NR4A2	Fwd: CACAGGTTGCAATGCGTTCG
	Rev: TCAATTATTGCTGGCGGTGG
NR4A3	Fwd: CGTCGAAACCGATGTCAGTA
	Rev: GACGACCTCTCCTCCCTTTC
PTGS2	Fwd: GAATCATTCACCAGGCAAATTG
	Rev: TCTGTACTGCGGGTGGAACA

Western blotting

Cells were treated for 2 hours, washed in 1X PBS, and collected in RIPA buffer with protease and phosphatase inhibitors. Samples were run on acrylamide gel and transferred onto a nitrocellulose membrane. Membrane was blocked for one hour in 5% of milk or TBS-T and incubated overnight at 4ºC with primary antibody. Antibodies used were: VINCULIN (Abcam 73412), GAPDH (Abcam 9482), PRDM1 (Cell signaling Technologies 9115), CDKN1A (Cell Signaling Technologies 2947), FOS (Cell Signaling Technologies 2250) FOSB (Cell Signaling Technologies 2251), FOSL2 (Abcam 19967), CBX3 (Abcam 213167), phospho-CBX3 (Abcam 45270). The next day, membranes

were washed, incubated with HRP-conjugated secondary antibodies for 1 hour at room temperature (RT), and developed with SuperSignal West Pico Plus Chemiluminescent substrate.

RNA-Seq

RNA from one million cells was isolated using the RNeasy plus mini kit (Qiagen #74134). 5µg of RNA was subjected to ribosomal and mitochondrial RNA depletion using the RiboZero Gold kit (Human/Mouse/Rat, Epicentre #MRZG12324) according to manufacturer›s instructions. The ribo-zero treated RNA was used to create multiplexed RNA-Seq libraries using the NEBNext Ultra RNA Library Prep Kit (Illumina E7530) according to the manufacturer's instructions. Briefly 500pg of ribozero treated RNA was fragmented and used to produce cDNA libraries using the NEBnext Ultra RNA library prep kit (NEB, E7530S) according to the manufacturer's protocol. Purified double-stranded cDNA underwent end-repair and dA-tailing reactions following manufacturer's reagents and reaction conditions. The obtained DNAs was adaptor ligated using adaptors and enzymes provided in the NEBNext Multiplex Oligos kit for Illumina (NEB#E7335) and following recommended reaction conditions. Eluted DNA was enriched with PCR reaction using Fusion High-Fidelity PCR Master Mix kit (NEB, M0531S) and specific index primers supplied in NEBNext Multiplex Oligo Kit for Illumina (Index Primer Set 1, NEB, E7335L). Conditions for PCR used are as follows: 98°C, 30 sec; [98°C, 10 sec; 65°C, 30 sec; 72°C, 30 sec] X 15 cycles; 72°C, 5 min; hold at 4°C. PCR reaction mix was purified using Agencourt AMPure XP beads (1X of reaction volume). Libraries were eluted in 20µl elution buffer. All the libraries went through quality control analysis using an Agilent Bioanalyzer before subjection to next-generation sequencing using the Illumina HiSeq 2500 platform. Quality control of RNA-Seq datasets was performed by FastQC and Cutadapt to remove adaptor sequences and low-quality regions. The high-quality reads were aligned to UCSC hg19 for human using Tophat 2.0.11 without novel splicing form calls. Transcript abundance and differential expression were calculated with Cufflinks 2.2.1. FPKM values were used to normalize and quantify each transcript.

Proteomics and phosphoproteomics

Cells were lysed by homogenization (QIAshredder) in lysis buffer (2% SDS, 150mM NaCl, 50mM Tris pH 7.4). Lysates were reduced with 5mM DTT, alkylated with 15mM iodoacetamide for 30 minutes in the dark, quenched with 50mM fresh DTT and proteins precipitated by methanol/chloroform precipitation. Digests were carried out in 200mM EPPS, pH 8.5 in presence of 2% acetonitrile (v/v) with LysC (Wako, 2mg/ml at 1:75) for 3 hours at room temperature and subsequently trypsinated (Promega, #V5111, stock 1:75) overnight at 37°C. Missed cleavage rate was assayed from a small aliquot by mass spectrometry. For whole proteome analysis, digests containing approximately 60µg of peptide material were directly labeled with TMT reagents (Thermo Fisher

Scientific). Labeling efficiency and TMT ratios were assayed by mass spectrometry, while labeling reactions were stored at -80°C. After quenching of TMT labeling reactions with hydroxylamine, TMT labeling reactions were mixed, solvent evaporated to near completion and TMT labeled peptides purified and desalted by acidic reversed phase C_{18} chromatography. Peptides were then fractionated by alkaline reversed phase chromatography into 96 fractions and combined into 12 samples. Before mass spectrometric analysis, peptides were desalted over StageTips.

Phosphopeptides were enriched from digested material containing approximately 4mg of peptide material per sample by Fe-NTA chromatography (Thermo Fisher Scientific #A32992). After TMT labeling, phosphopeptides were subjected to alkaline reverse fractionation by a linear acetonitrile gradient into 96 fractions and combined into 24 samples. TMT labeled, fractionated phosphopeptides were desalted over StageTips prior to mass spectrometry analysis.

Data were collected by a MultiNotch MS3 TMT method using an Orbitrap Lumos mass spectrometer (Thermo Fisher Scientific) coupled to a Proxeon EASY-nLC 1200 liquid chromatography (LC) system (Thermo Fisher Scientific). The 100µm inner diameter capillary column used was packed with C_{18} resin (SepPax Technologies Inc., 1.8µm). Peptides of each fraction were separated over 3 hours acidic acetonitrile gradients by LC prior to mass spectrometry (MS) injection. The first scan of the sequence was an MS1 spectrum (Orbitrap analysis; resolution 120,000; mass range 400–1400Th). MS2 analysis followed collision-induced dissociation (CID, CE=35) with a maximum ion injection time of 150ms and an isolation window of 0.7Da. In order to obtain quantitative information, MS3 precursors were fragmented by high-energy collision-induced dissociation (HCD) and analyzed in the Orbitrap at a resolution of 50,000 at 200Th. Further details on LC and MS parameters and settings used were described by Paulo et al., 2016. Peptides were searched with a SEQUEST (v.28 (rev. 12)) based software against a size-sorted forward and reverse database of the human proteome (Uniprot 02/2014) with added common contaminant proteins. For this, spectra were first converted to mzXML. Searches were performed using a mass tolerance of 50ppm for precursors, fragment ion tolerance 0.9Da. This wide mass tolerance maximizes sensitivity in conjunction with SEQUEST searches and linear discriminant analysis. For the searches maximally 2 missed cleavages per peptide were allowed. We searched dynamically for oxidized methionine residues (+15.9949Da) and applied a target decoy database strategy and a false discovery rate (FDR) of 1% set for peptide-spectrum matches following filtering by linear discriminant analysis (LDA) and a final collapsed protein-level FDR of 1%. Quantitative information on peptides was derived from MS3 scans. Quant tables were generated requiring an MS2 isolation specificity of >65% for each peptide and a sum of TMT s/n of >150 over all channels for any given peptide and exported to Excel and further processed therein. Details of the TMT intensity quantification method and further search parameters applied are described by Paulo et al., 2016.

Metabolomics

Half of the metabolite layer was resuspended in 20µL LC/MS grade water, 5µL were injected over a 15min gradient using a 5500 QTRAP triple quadrupole mass spectrometer (AB/SCIEX) coupled to a Prominence UFLC HPLC system (Shimadzu) via SRM of a total of 287 SRM transitions using positive and negative polarity switching corresponding to 259 unique endogenous water-soluble metabolites. Samples were separated using a Amide XBridge HPLC hydrophilic interaction liquid chromatographic (HILIC) column (3.5µm; 4.6mm inner diameter × 100mm length; Waters) at 300µL/min. Gradients were run starting from 85% buffer B (LC/MS grade acetonitrile) to 40% B from 0–5 min; 40% B to 0% B from 5–16 min; 0% B was held from 16–24 min; 0% B to 85% B from 24–25 min; 85% B was held for 7 min to re-equilibrate the column. Buffer A was comprised of 20mM ammonium hydroxide/20 mM ammonium acetate, pH 9.0 in 95:5 water/acetonitrile. Peak areas from the total ion current for each metabolite SRM transition were integrated using MultiQuant version 2.1.1 software (AB/SCIEX). The other half of the metabolite layer was re-suspended in 20µL LC/MS grade water, 5µL were analyzed by positive/negative polarity switching mode using a hybrid QExactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific) via a data-dependent analysis (DDA Top8). Metabolites were delivered and separated using an EASY-nLC nanoflow HPLC (Thermo Fisher Scientific) at 225nL/min using self-packed 15cm length × 75µm inner diameter. C_{18} fritted microcapillary columns. Solvent gradient conditions were 25 minutes from 3% B buffer to 38% B (B buffer: 100% acetonitrile; A buffer: 0.1% formic acid/99.9% water). The data were analyzed using Elements (Proteome Software) with the NIST MS/MS spectral database (http://chemdata.nist.gov/mass-spc/msms-search/) and HMDB metabolite database followed by statistical analysis and pathway analysis with MetaboAnalyst 3.0 software.

Acknowledgments

A.S. was supported by a Boehringer Ingelheim PhD Fellowship. This work was supported by the following grants from L.I.Z.: R01 HL04880, P015PO1HL32262-32, 5P30 DK49216, 5R01 DK53298, 5U01 HL10001-05, R24 DK092760, 1R24OD017870-01.

Author contributions

A.S., E.M.F., and M.E.M., performed experiments. M.K. performed mass-spectrometry experiments and proteomics data analysis. S.Y., provided RNA-sequencing data analysis. A.S. and L.I.Z. wrote the manuscript.

Conflict of interests

L.I.Z. is founder and stockholder of Fate, Inc., Scholar Rock, Camp4 therapeutics and a scientific advisor for Stemgent. The other authors declare no competing interests.

References

1. Wilkinson, A. C., Igarashi, K. J. & Nakauchi, H. Haematopoietic stem cell self-renewal in vivo and ex vivo. Nat. Rev. Genet. 21, 541–554 (2020).
2. Cutler, C. et al. Prostaglandin-modulated umbilical cord blood hematopoietic stem cell transplantation. Blood 122, 3074–3081 (2013).
3. King, K. Y. & Goodell, M. A. Inflammatory modulation of HSCs: viewing the HSC as a foundation for the immune response. Nat Rev Immunol 11, 685–692 (2011).
4. Ricciotti, E. & FitzGerald, G. A. Prostaglandins and inflammation. Arterioscler. Thromb. Vasc. Biol. 31, 986–1000 (2011).
5. Miller, S. B. Prostaglandins in health and disease: an overview. Semin Arthritis Rheum 36, 37–49 (2006).
6. North, T. E. et al. Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. Nature 447, 1007–1011 (2007).
7. Goessling, W. et al. Prostaglandin E2 enhances human cord blood stem cell xenotransplants and shows long-term safety in preclinical nonhuman primate transplant models. Cell Stem Cell 8, 445–458 (2011).
8. Hoggatt, J., Singh, P., Sampath, J. & Pelus, L. M. Prostaglandin E2 enhances hematopoietic stem cell homing, survival, and proliferation. Blood 113, 5444–5455 (2009).
9. Goessling, W. et al. Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. Cell 136, 1136–1147 (2009).
10. Klein, T., Shephard, P., Kleinert, H. & Kömhoff, M. Regulation of cyclooxygenase-2 expression by cyclic AMP. Biochim Biophys Acta 1773, 1605–1618 (2007).
11. Sinclair, A. et al. CXCR2 and CXCL4 regulate survival and self-renewal of hematopoietic stem/progenitor cells. Blood 128, 371–383 (2016).
12. Freire, P. R. & Conneely, O. M. NR4A1 and NR4A3 restrict HSC proliferation via reciprocal regulation of C/EBPα and inflammatory signaling. Blood 131, 1081–1093 (2018).
13. Santaguida, M. et al. JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. Cancer Cell 15, 341–352 (2009).
14. Antonchuk, J., Sauvageau, G. & Humphries, R. K. HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. Cell 109, 39–45 (2002).
15. Desbaillets, I., Diserens, A. C., Tribolet, N., Hamou, M. F. & Van Meir, E. G. Upregulation of interleukin 8 by oxygen-deprived cells in glioblastoma suggests a role in leukocyte activation, chemotaxis, and angiogenesis. J. Exp. Med. 186, 1201–1212 (1997).
16. Lingam, I. et al. Serial blood cytokine and chemokine mRNA and microRNA over 48 h are insult specific in a piglet model of inflammation-sensitized hypoxia-ischaemia. Pediatr Res 74, 50–12 (2020).
17. Blaser, B. W. et al. CXCR1 remodels the vascular niche to promote hematopoietic stem and progenitor cell engraftment. J. Exp. Med. 214, 1011–1027 (2017).
18. Masat, E. et al. RelB activation in anti-inflammatory decidual endothelial cells: a master plan to avoid pregnancy failure? Sci Rep 5, 14847–7 (2015).
19. Saccani, S., Pantano, S. & Natoli, G. Two waves of nuclear factor kappaB recruitment to target promoters. J. Exp. Med. 193, 1351–1359 (2001).
20. Jung, K. H. et al. RNA sequencing reveals distinct mechanisms underlying BET inhibitor JQ1-mediated modulation of the LPS-induced activation of BV-2 microglial cells. J Neuroinflammation 12, 36–18 (2015).
21. Elkon, R., Linhart, C., Halperin, Y., Shiloh, Y. & Shamir, R. Functional genomic delineation of TLR-induced transcriptional networks. BMC Genomics 8, 394–17 (2007).
22. Chanchevalap, S. et al. Kruppel-like factor 5 is an important mediator for lipopolysaccharide-induced proinflammatory response in intestinal epithelial cells. Nucleic Acids Res. 34, 1216–1223 (2006).
23. Cullingford, T. E. et al. Differential regulation of Krüppel-like factor family transcription factor expression in neonatal rat cardiac myocytes: effects of endothelin-1, oxidative stress and cytokines. Biochim Biophys Acta 1783, 1229–1236 (2008).

24. Sur, I., Undén, A. B. & Toftgård, R. Human Krüppel-like factor5/KLF5: synergy with NF-kappaB/Rel factors and expression in human skin and hair follicles. Eur J Cell Biol 81, 323–334 (2002).
25. Varas, A. et al. Blockade of bone morphogenetic protein signaling potentiates the pro-inflammatory phenotype induced by interleukin-17 and tumor necrosis factor-α combination in rheumatoid synoviocytes. Arthritis Res Ther 17, 192–10 (2015).
26. Eddowes, L. A. et al. Antiviral activity of bone morphogenetic proteins and activins. Nat Microbiol 4, 339–351 (2019).
27. Dendooven, A. et al. Loss of endogenous bone morphogenetic protein-6 aggravates renal fibrosis. Am. J. Pathol. 178, 1069–1079 (2011).
28. Halterman, M. W. et al. The endoplasmic reticulum stress response factor CHOP-10 protects against hypoxia-induced neuronal death. J. Biol. Chem. 285, 21329–21340 (2010).
29. Hanna, R. N. et al. The transcription factor NR4A1 (Nur77) controls bone marrow differentiation and the survival of Ly6C- monocytes. Nat Immunol 12, 778–785 (2011).
30. Land, R. H. et al. The Orphan Nuclear Receptor NR4A1 Specifies a Distinct Subpopulation of Quiescent Myeloid-Biased Long-Term HSCs. STEM CELLS 33, 278–288 (2014).
31. Mahony, C. B., Pasche, C. & Bertrand, J. Y. Oncostatin M and Kit-Ligand Control Hematopoietic Stem Cell Fate during Zebrafish Embryogenesis. Stem Cell Reports 10, 1920–1934 (2018).
32. Park, I. K. et al. Molecular cloning and characterization of a novel regulator of G-protein signaling from mouse hematopoietic stem cells. J. Biol. Chem. 276, 915–923 (2001).
33. Yu, H., Yuan, Y., Shen, H. & Cheng, T. Hematopoietic stem cell exhaustion impacted by p18 INK4C and p21 Cip1/Waf1 in opposite manners. Blood 107, 1200–1206 (2006).
34. Cheng, T. et al. Hematopoietic stem cell quiescence maintained by p21cip1/waf1. Science 287, 1804–1808 (2000).
35. Doidge, R., Mittal, S., Aslam, A. & Winkler, G. S. The anti-proliferative activity of BTG/TOB proteins is mediated via the Caf1a (CNOT7) and Caf1b (CNOT8) deadenylase subunits of the Ccr4-not complex. PLoS ONE 7, e51331 (2012).
36. Hubmann, G., Guillouet, S. & Nevoigt, E. Gpd1 and Gpd2 fine-tuning for sustainable reduction of glycerol formation in Saccharomyces cerevisiae. Appl Environ Microbiol 77, 5857–5867 (2011).
37. Xie, J. et al. GPD1 Enhances the Anticancer Effects of Metformin by Synergistically Increasing Total Cellular Glycerol-3-Phosphate. Cancer Res. 80, 2150–2162 (2020).
38. Yang, Y. et al. ALG3 Is Activated by Heat Shock Factor 2 and Promotes Breast Cancer Growth. Med Sci Monit 24, 3479–3487 (2018).
39. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165, 535–550 (2016).
40. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837 (2007).
41. Briggs, S. D. et al. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes & Development 15, 3286–3295 (2001).
42. Shilatifard, A. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr Opin Cell Biol 20, 341–348 (2008).
43. Karra, A. S. et al. ERK2 phosphorylates the epigenetic regulator CXXC-finger protein 1 (CFP1). bioRxiv 562173 (2019).
44. Ayoub, N., Jeyasekharan, A. D., Bernal, J. A. & Venkitaraman, A. R. HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. Nature 453, 682–686 (2008).
45. Lomberk, G., Bensi, D., Fernandez-Zapico, M. E. & Urrutia, R. Evidence for the existence of an HP1-mediated subcode within the histone code. Nature Cell Biology 8, 407–415 (2006).
46. Bérubé, N. G., Smeenk, C. A. & Picketts, D. J. Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. Hum. Mol. Genet. 9, 539–547 (2000).
47. Lopez, R. et al. Linker histone partial phosphorylation: effects on secondary structure and chromatin condensation. Nucleic Acids Res. 43, 4463–4476 (2015).
48. Tan, Y., Demeter, M. R., Ruan, H. & Comb, M. J. BAD Ser-155 phosphorylation regulates BAD/

Bcl-XL interaction and cell survival. J. Biol. Chem. 275, 25865–25869 (2000).
49. Hayakawa, J. et al. Inhibition of BAD phosphorylation either at serine 112 via extracellular signal-regulated protein kinase cascade or at serine 136 via Akt cascade sensitizes human ovarian cancer cells to cisplatin. Cancer Res. 60, 5988–5994 (2000).
50. Zha, J., Harada, H., Yang, E., Jockel, J. & Korsmeyer, S. J. Serine phosphorylation of death agonist BAD in response to survival factor results in binding to 14-3-3 not BCL-X(L). Cell 87, 619–628 (1996).
51. Polzien, L. et al. Identification of novel in vivo phosphorylation sites of the human proapoptotic protein BAD: pore-forming activity of BAD is regulated by phosphorylation. J. Biol. Chem. 284, 28004–28020 (2009).
52. Wu, K. K., Cheng, H.-H. & Chang, T.-C. 5-methoxyindole metabolites of L-tryptophan: control of COX-2 expression, inflammation and tumorigenesis. J Biomed Sci 21, 17–8 (2014).
53. Kaelin, W. G. & McKnight, S. L. Influence of metabolism on epigenetics and disease. Cell 153, 56–69 (2013).
54. Xia, J. & Wishart, D. S. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nat Protoc 6, 743–760 (2011).
55. Liew, K. L., Jee, J. M., Yap, I. & Yong, P. V. C. In Vitro Analysis of Metabolites Secreted during Infection of Lung Epithelial Cells by Cryptococcus neoformans. PLoS ONE 11, e0153356 (2016).
56. Rajman, L., Chwalek, K. & Sinclair, D. A. Therapeutic Potential of NAD-Boosting Molecules: The In Vivo Evidence. Cell Metab 27, 529–547 (2018).
57. Ma, Y. et al. Anti-Inflammation Effects and Potential Mechanism of Saikosaponins by Regulating Nicotinate and Nicotinamide Metabolism and Arachidonic Acid Metabolism. Inflammation 39, 1453–1461 (2016).
58. Peled, A. et al. Dependence of human stem cell engraftment and repopulation of NOD/SCID mice on CXCR4. Science 283, 845–848 (1999).
59. Horwitz, M. E. et al. Umbilical cord blood expansion with nicotinamide provides long-term multilineage engraftment. J. Clin. Invest. 124, 3121–3128 (2014).
60. Ruiz, M. et al. Altered glycolipid and glycerophospholipid signaling drive inflammatory cascades in adrenomyeloneuropathy. Hum. Mol. Genet. 24, 6861–6876 (2015).
61. Pan, Z.-Y. & Wang, H. Synergistic interaction between choline and aspirin against acute inflammation induced by carrageenan and lipopolysaccharide. Int Immunopharmacol 20, 229–237 (2014).
62. Ilcol, Y. O., Yilmaz, Z., Cansev, M. & Ulus, I. H. Choline or CDP-choline alters serum lipid responses to endotoxin in dogs and rats: involvement of the peripheral nicotinic acetylcholine receptors. Shock 32, 286–294 (2009).
63. Himanen, S. V. & Sistonen, L. New insights into transcriptional reprogramming during cellular stress. J. Cell. Sci. 132, jcs238402 (2019).
64. Hooper, A. T. et al. Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. Cell Stem Cell 4, 263–274 (2009).
65. Heissig, B. et al. Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of kit-ligand. Cell 109, 625–637 (2002).
66. Jung, Y. et al. Hematopoietic stem cells regulate mesenchymal stromal cell induction into osteoblasts thereby participating in the formation of the stem cell niche. STEM CELLS 26, 2042–2051 (2008).
67. Adams, G. B. et al. Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor. Nature 439, 599–603 (2006).
68. Arai, F. et al. Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. Cell 118, 149–161 (2004).
69. Calvi, L. M. et al. Osteoblastic cells regulate the haematopoietic stem cell niche. Nature 425, 841–846 (2003).
70. Sirin, O., Lukov, G. L., Mao, R., Conneely, O. M. & Goodell, M. A. The orphan nuclear receptor Nurr1 restricts the proliferation of haematopoietic stem cells. Nature Cell Biology 12, 1213–1219 (2010).
71. Min, I. M. et al. The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. Cell Stem Cell 2, 380–391 (2008).

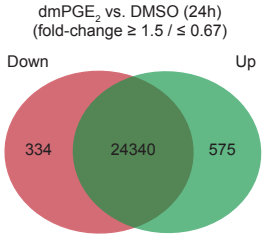
72. Essers, M. A. G. et al. IFNalpha activates dormant haematopoietic stem cells in vivo. Nature 458, 904–908 (2009).
73. Murai, J. et al. Chromatin Remodeling and Immediate Early Gene Activation by SLFN11 in Response to Replication Stress. Cell Rep 30, 4137–4151.e6 (2020).
74. Wilson, A. et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell 135, 1118–1129 (2008).
75. Yuan, M., Breitkopf, S. B., Yang, X. & Asara, J. M. A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. Nat Protoc 7, 872–881 (2012).
76. Norton, V. G., Imai, B. S., Yau, P. & Bradbury, E. M. Histone acetylation reduces nucleosome core particle linking number change. Cell 57, 449–457 (1989).
77. Struhl, K. Histone acetylation and transcriptional regulatory mechanisms. Genes & Development 12, 599–606 (1998).
78. Koury, M. J. Programmed cell death (apoptosis) in hematopoiesis. Experimental Hematology 20, 391–394 (1992).
79. Kollek, M. et al. Transient apoptosis inhibition in donor stem cells improves hematopoietic stem cell transplantation. J. Exp. Med. 214, 2967–2983 (2017).
80. George, R. J., Sturmoski, M. A., Anant, S. & Houchen, C. W. EP4 mediates PGE2 dependent cell survival through the PI3 kinase/AKT pathway. Prostaglandins & Other Lipid Mediators 83, 112–120 (2007).
81. Seyrek, K. et al. Controlling Cell Death through Post-translational Modifications of DED Proteins. Trends in Cell Biology 30, 354–369 (2020).
82. Datta, S. R. et al. Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. Cell 91, 231–241 (1997).
83. Sastry, K. S. R. et al. Epinephrine protects cancer cells from apoptosis via activation of cAMP-dependent protein kinase and BAD phosphorylation. J. Biol. Chem. 282, 14094–14100 (2007).
84. del Peso, L., González-García, M., Page, C., Herrera, R. & Nuñez, G. Interleukin-3-induced phosphorylation of BAD through the protein kinase Akt. Science 278, 687–689 (1997).

Supplemental Figures

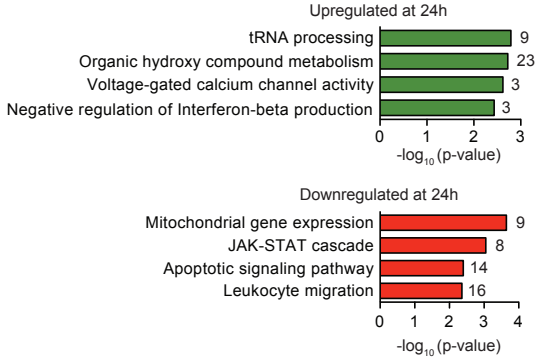
A

Biological function	Differentially expressed genes (2h)
Cell adhesion and migration	ADAM8, CCR7, CCRL2, CCL4, CSF1, CXCL1,CXCL2, CXCL3, CXCL8, CXCL16, MALT1, IL1R1, IL1RN, IL1RAP, IL10RA, IL4R, IL7R
Cell cycle and division	ANLN, AURKB, BORA, BUB1B, CCND2, CCNF, CDC25C, CDCA2, CENPA, DLGAP5, E2F8, KNSTRN,KIF20A, PIMREG,PLK1, TUBA8
Transciption factors	ATF3, EGR1, EGR3, FOS, FOSB, FOSL1, FOXO1, JUN, JUNB, JUND, NR4A1, NR4A2, NR4A2, RELB
Blood development	BCL11a, BTG1, CD86, EFNB2, HMOX1, LIF, NFE2L2, PDGFB, THBS1
Cell signaling	BMP6, CAMKK1, DUSP1, PDE4A, PDE4B, PDE4D, PLK2, PTGS2

B.



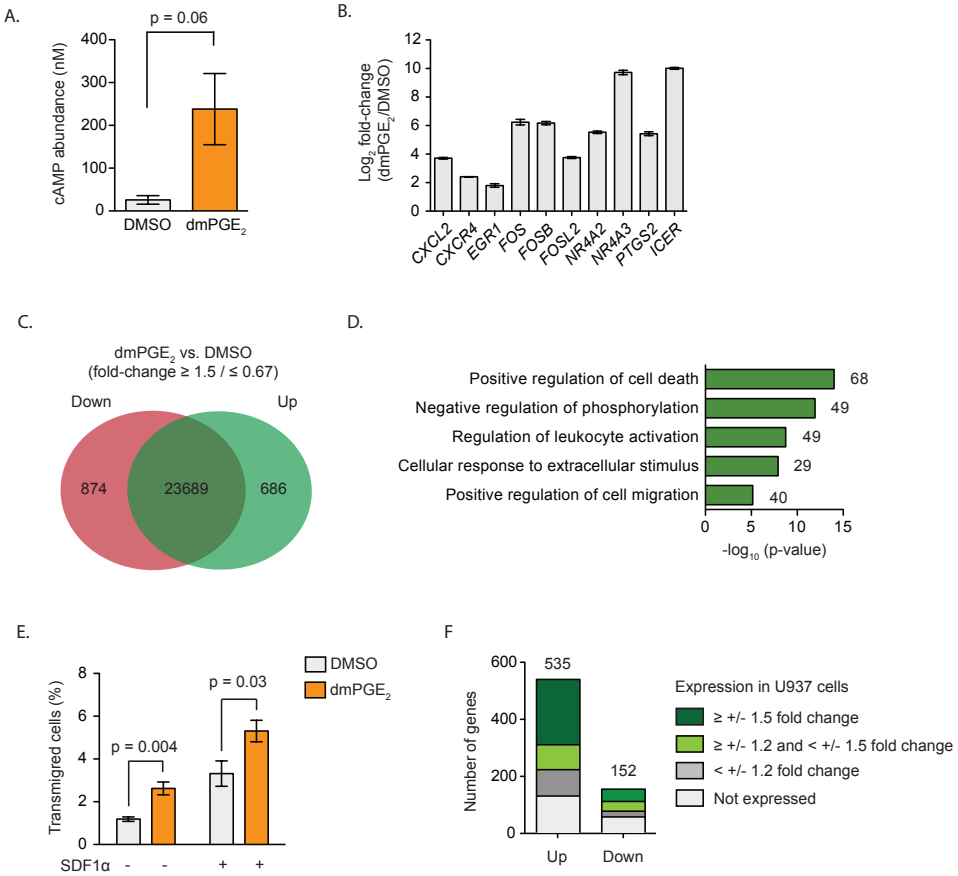
C.



D.

Biological function	Differentially expressed genes (24h)
tRNA processing	AARSD1, ANKRD16, CLP1, DTD1, RPP38, RPP40, TRMT12, TRMT11, TSEN34
Organic hydroxy compound metaboloism	ALDH2, APOA1, CYP7B1 CYP2D6, CRYM, GDPD1, HAGH, LCAT, LIPE, MAOA, NUDT10, OSBPL6, PLPP1, RDH5, RDH10
Mitochondrial gene expression	TEFM, MRPL55, MRPS12, MRPS22, MRPS28, MTERF1, MTRES1, RMND1
Apoptotic signaling	BCL3, BEX3, CD38, GSKIP, ISG15, PELI3, TNFSF8, TNFRSF10A, TRIM39, TICAM
Leukocyte migration	CCR2, GPR18, GYPB, FER, HOXA7, ITGA9, ITGAM, IL12

Supplemental Figure 1. Transcriptomic analysis in CD34⁺ HSPCs. (A) Examples of genes that are ≥ 1.5-fold or ≤ 0.67-fold differentially expressed in CD34⁺ HSPCs exposed to dmPGE$_2$ for 2 hours. (B) Venn diagram showing the number of genes (in green, 575) and downregulated (in red, 334) in CD34⁺ HSPCs 24 hours after dmPGE$_2$ stimulation, as determined by RNA-Seq analysis. Differentially expressed gene (DEG) criteria: FPKM ≥ 1 after treatment; fold change ≥ 1.5 or ≤ 0.67. (C) GO term enrichment analysis of genes upregulated (upper panel, in green) and down-regulated (lower panel, in red) in CD34⁺ HSPCs 24 hours post dmPGE$_2$ treatment. The number of genes associated with each GO term are shown at the end of each bar within the graph. P-values were calculates using hypergeometric test and Benjamini-Hochberg correction. (D) Examples of genes that are ≥ 1.5-fold or ≤ 0.67-fold differentially expressed in CD34⁺ HSPCs 24 hours after dmPGE$_2$ stimulation

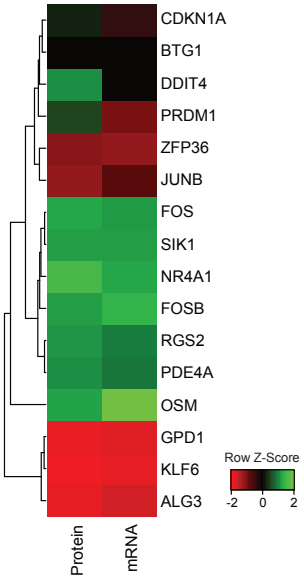


Supplemental Figure 2. U937 cells show a response to dmPGE$_2$ that is similar to CD34$^+$ HSPCs. (A) cAMP concentration in U937 cells exposed to 2 hours of DMSO or dmPGE$_2$. Levels were determined by bioluminescence cAMP-Glo assay. The data are shown as the mean ± SEM (B) RT-qPCR in U937 cells of genes identified as differentially expressed in CD34$^+$ HSPCs (n = 3 biologically independent experiments; mean values ± SEM). (C) Venn diagram showing the number of genes upregulated (in green, 686) and downregulated (in red, 874) in U937 cells after a 2-hour pulse dmPGE$_2$, as determined by RNA-Seq analysis. DEG criteria: FPKM ≥ 1 after treatment; fold change ≥ 1.5 or ≤ 0.67. (D) GO term enrichment analysis of upregulated genes (686) in U937 cells 2 hours post dmPGE$_2$ treatment. The number of genes associated with each GO term are shown at the bar within the graph. P-values were calculates using hypergeometric test and Benjamini-Hochberg correction. (E) Transwell migration assessment in U937 cells. Cells were exposed to dmPGE$_2$ or DMSO for 2h after which the stimuli were washed out. Cells were then placed in the top chamber of the transwell system, with or without recombinant human SDF1α in the bottom chamber. After 24 hours, cells migration to the bottom chamber was quantified as percentage of total cells seeded (n = 3 biologically independent experiments; mean values ± SEM). (F) Comparison of gene expression changes in U937 cells to genes identified as upregulated (fold change ≥ 1.5; 535 genes) or downregulated (fold change ≤ 0.67-fold; 152) in CD34$^+$ HSPCs.

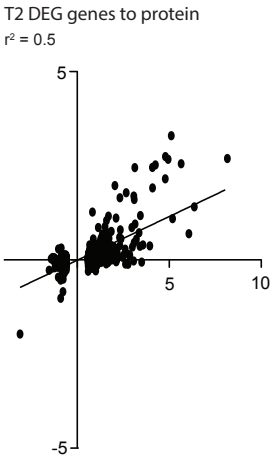
A.

Fold-change	Diffentially expressed proteins
≥ 1.5	BTG1, CDKN1A, DDIT4, FOS, FOSB, JUNB, NR4A1, OSM, PDE4A, PRDM1, RGS2, SIK1, ZFP36
≤ 0.67	ALG3, GPD1, KLF6

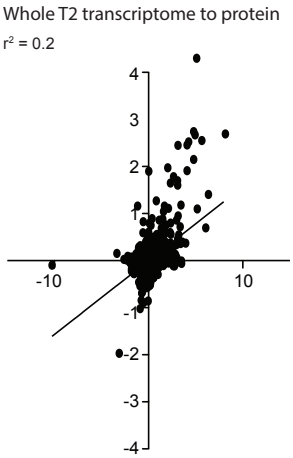
B.



C.



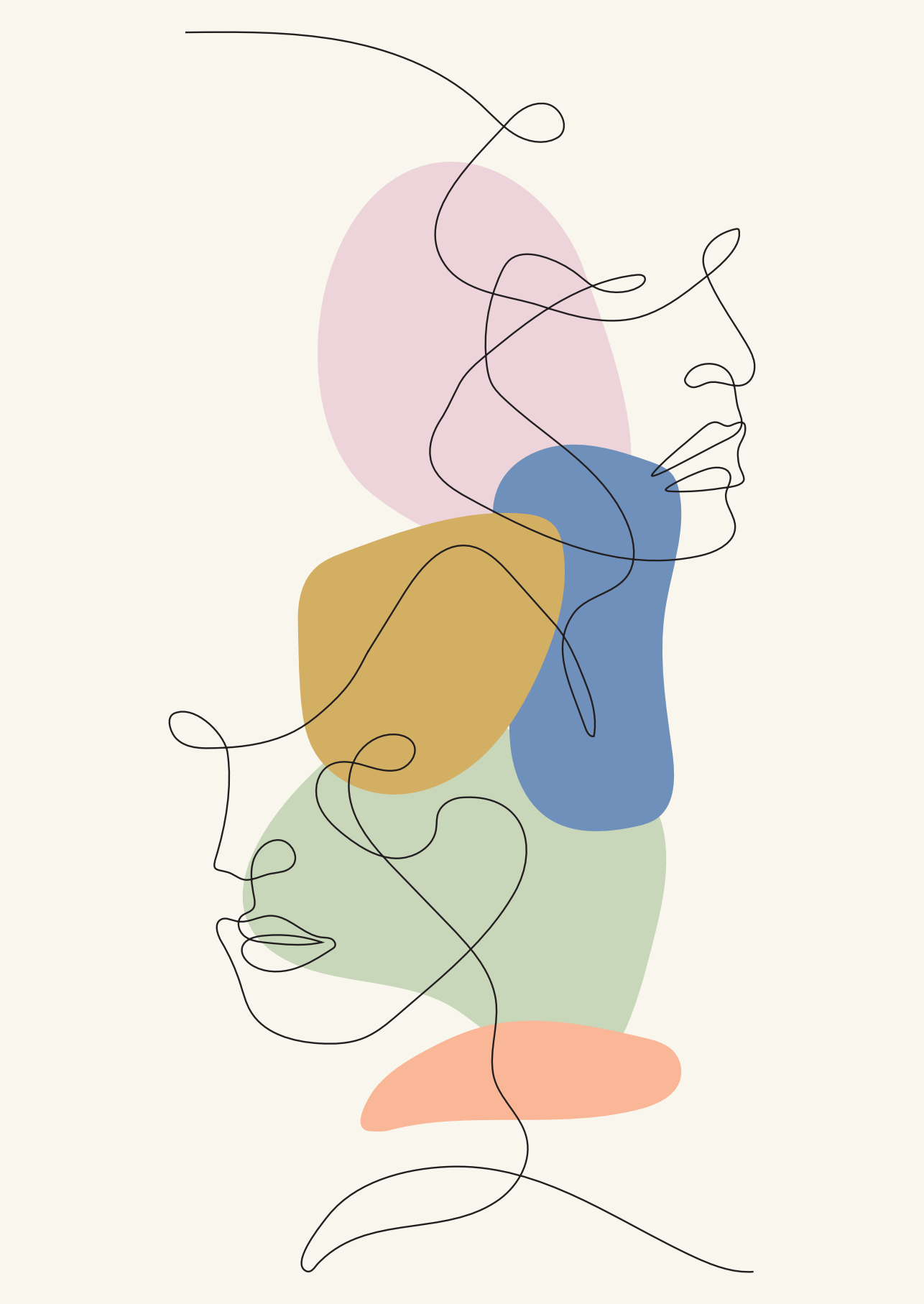
D.



E.

Fold-change	Diffentially expressed phosphoproteins
≥ 1.5	HIST1H1D, NUMA1, CLIP1, RANBP2, RANBP2, SQSTM1, RTN4, KCNAB2, KLC4, SIK1, CLIP1, CAMKK2, ATRX, NCBP1, TLE3, BAD, PRPF40A, REPS1, RCSD1, RCSD1, USP36, PLCB2, TPI1, CBX3, EIF5B, CBX3, RPRD2, ARHGAP30, WDR74, EIF3B, CAAP1, RNF31, DYNC1LI1, ARID4A, RAD23B, PPP1R10, SRRM1
≤ 0.67	SENP3 , TPR , RBBP6 , HDGFRP2, AKAP2, YBX1, EIF5B, HMGN1 , SETD1A, ZC3H14, TOP2B, NUP153, SEC62, NUP35, CD44, NIPBL, SAMD1, PRPF4B, STAU2, MARCKS, TERF2, SRPK1, CEP170, SRRM2, LIMA1, RAB44, BIN2, DLGAP5, TLE3, CLASRP, MYBBP1A, PCBP2, XRCC1, GIGYF2, INCENP, UBAP2L, TCOF1, TBC1D10B, ZC3H13, PPP1R18, SNAPIN, ACIN1, AKAP2, SH3PXD2A, AKAP2, EPS15, MYO9B, MARCKS, MAP7D1, CNOT3, EIF4G1, MAP7D1, AKAP2, KCNAB2, ATXN2L, CDCA2, PRPF38B, UBAP2L, PRPF38B, TBC1D10B, KPNA2, PATL1, STMN1, ZYX, HCLS1, IRF2BP2

Supplemental Figure 3. Proteomics and phosphoproteomics analysis in U937 cells after pulse exposure to dmPGE2. (A) Proteins with a fold change in abundance of ≥ 1.5-fold (13) or ≤ 0.67-fold (3) in U937 cells exposed to $dmPGE_2$ for 2 hours. (B) Hierarchical clustering heatmap of fold change in mRNA and protein levels from ≥ 1.5-fold or ≤ 0.67-fold differentially expressed proteins (16) after 2 hours of $dmPGE_2$ treatment in U937 cells. (C, D) Linear regression analysis of protein and mRNA fold changes in U937 cells. Values of genes identified as significant differentially expressed genes in $CD34^+$ HSPCs (687) (C) and the whole transcriptome (D) were plotted. (E) Phosphoproteins with a fold change in abundance of ≥ 1.5-fold (37) or ≤ 0.67-fold (66) in U937 cells exposed to $dmPGE_2$ for 2 hours.



4

Niche signals regulate continuous transcriptional states in hematopoietic stem cells

Manuscript submitted

Chapter 4

Niche signals regulate continuous transcriptional states in hematopoietic stem cells

Eva M. Fast, Audrey Sporrij, Margot E. Manning, Edroaldo Lummertz da Rocha, Song Yang, Yi Zhou, Jimin Guo, Ninib Baryawno, Nikolas Barkas, Leonard I. Zon#

Correspondence to: leonard.zon@enders.tch.harvard.edu

Abstract

Hematopoietic stem cells (HSCs) ensure adequate blood cell production upon exposure to external stressors. A comprehensive understanding on the heterogeneity and specificity of responses to distinct external stimuli in HSCs and more differentiated progenitor is currently lacking. We performed single cell RNA sequencing (scRNA-Seq) on functionally validated HSCs and multipotent progenitors (MPPs) after *in vivo* perturbation of the niche signals interferon, G-CSF and prostaglandin. We identified that HSCs reside in six distinct states with individual cell clusters showing enriched, but not exclusive, expression of specific marker genes. Two-hour pulse stimulation with 16,16-dimethyl Prostaglandin E_2 (dmPGE$_2$) or poly(I:C) induced rapid transitions between HSCs states, shifting 55% and 41% of HSCs, respectively. Furthermore, depletion of endogenous prostaglandins by indomethacin (Indo) induced immediate early gene (IEG) turnover. Unlike the rapid HSC state transitions seen in response to dmPGE$_2$ and poly(I:C), G-CSF induced specific effects within HSC states, specifically within a metabolically active cell cluster, while not changing cell distributions between HSC states. Comparison of chromatin accessibility using scATAC-Seq in unperturbed HSCs and LSKs revealed HSC specific chromatin heterogeneity that could be indicative of intrinsic determinants that predispose for specific cellular responses to niche signals. Overall, we provide the first comprehensive dissection, on a single cell level, of HSC and LSK transcriptional states during perturbations in the niche. The HSC state specific features may represent important determinants of regenerative potential during stress hematopoiesis.

Introduction

Stem cell therapy holds promises for numerous indications including blood diseases, autoimmune diseases, neurodegeneration and cancer[1]. Despite being used in the clinic for over 30 years, HSC transplantation is still a highly risky procedure. To improve strategies for the enrichment of functional HSCs, recent efforts used scRNA-Seq to discover novel cell markers[2-4]. Yet, a consensus on the markers that identify the most purified HSC population has not been reached. The endeavor might be hindered by the extensive functional heterogeneity present within HSCs[5]. Both intrinsic and extrinsic factors regulate HSC function[6,7]. The stem cell niche forms an important extracellular regulator of HSCs as it not only anchors the cells but also maintains stem cell fate[8]. Release of soluble signals from the niche such as interferons, prostaglandins, and growth factors including SCF and G-CSF were shown to influence HSC function during homeostasis and injury[9,10] (reviewed by[11]). While known to be affected by a wide variety of extracellular signals, little is known about the heterogeneity and specificity of HSC responses to these external stimuli nor is it understood how differential responses relate the functional diversity between HSCs. Chromatin state is a crucial determinant of cell identity and behavior[12]. Hematopoietic differentiation is a prime example of cell fate change being concomitant with drastic remodeling of the epigenetic landscape[13]. Few studies have assessed chromatin states of purified, *in vivo* derived HSC populations due to technical limitations such as cell number[14,15]. Recent advancements in single cell chromatin accessibility sequencing (scATAC-Seq) provide a methodological framework for studying the diversity and uniqueness of HSC chromatin features at homeostasis and following stimulation[16,17].

Here, we performed comprehensive scRNA-Seq and scATAC-Seq profiling on functionally validated mouse HSCs and examined *in vivo* transcriptional responses to extracellular stimulation by mimicking signals from the niche. We found that unperturbed HSCs exist in distinct transcriptional states. Niche signals can alter the cell distribution between HSC states to various degrees depending on the stimulant as well as induce specific changes within distinct states. By comparing HSC to multipotent progenitor subpopulations, we determined the specificity of transcriptional responses to niche signals in HSCs and progenitor populations. Finally, analysis of chromatin states revealed an intrinsic cellular heterogeneity that may prime HSCs for particular transcriptional responses. Overall, this work provides the first comprehensive description of the transcriptomic and epigenetic landscape of HSCs and multipotent progenitors on a single cell levels during homeostasis and in response to niche signals.

Results

In vivo stimulation of HSCs and MPPs for functional, transcriptomic and epigenetic profiling

To investigate transcriptional responses to external signals, we profiled HSCs after *in vivo* perturbation with four distinct niche signals (Figure 1A). Male and female mice were treated with one of three activators, that is $dmPGE_2$, poly(I:C), or G-CSF, for two hours prior to downstream assessment or exposed to the Cox1/2 inhibitor Indomethacin (Indo) for one week to deplete endogenous prostaglandins. After the respective drug treatments, HSC and MPP populations within the Lin^-, $c\text{-}Kit^+$, $Sca1^+$ (LSK) compartment were isolated via FACS (Supplemental Figure 1A). Through limit dilutions transplantation assays (LDTA) and ELDA analysis[18], we determined a HSC frequency of 1 in 8 (Supplemental Figure 1B-1D). This confirmed that our isolation procedure allows for the profiling of functional, highly purified HSCs. Cell cycle analysis further verified that HSCs were mostly quiescent, in contrast to MPP populations[19] (Supplemental Figure 1E). Phenotypic marker composition within LSK cells remained largely consistent following different stimulations (Supplemental Figure 1F). An exception was the reduction of cells within the HSC compartment after $dmPGE_2$ treatment, decreasing from 1.9% to 0.85% of LSK cells upon $dmPGE_2$ stimulation (p-value of differential proportion analysis (DPA)[20] = 6.4×10^{-4}). To account for a potential phenotypic shift in HSC surface marker expression due to CD34 externalization, which would move functional HSCs into the MPP1 population, we compared the contribution of the transcriptional 'stem cell' state defined by scRNA-Seq in HSCs and MPP1s. We found no difference in the proportion of the 'stem cell' state in $dmPGE_2$ treated MPP1s when compared to control MPPs (Supplemental Figure 2G). After cell sorting, we subjected a total of 46,344 cells to scRNA-Seq using the 10x Genomics platform. We obtained an average of 37,121 ± 14,308 reads per cell and 2,994 ± 480 genes per cell (Supplementary Table 1). These numbers are indicative of a rich dataset that contained functionally validated HSCs.

Niche signals induce rapid transitions between transcriptional HSC states

To determine how niche stimulants affects HSCs, we analyzed cells from control and treatment exposed subjects by scRNA-Seq. Male versus female analysis of HSCs revealed sexual dimorphism during steady state and following activation (Supplemental Figure 3, Supplementary Table 2 and 3). We therefore regressed out any sex specific effects from further downstream analyses (see Methods). In the aggregated dataset, we detected a total of six HSC clusters (Figure 1B). To ensure optimal clustering parameters, we used a data driven approach (Silhouette Coefficient and Davies–Bouldin index) that was validated by comparison of two independent biological scRNA-Seq replicates

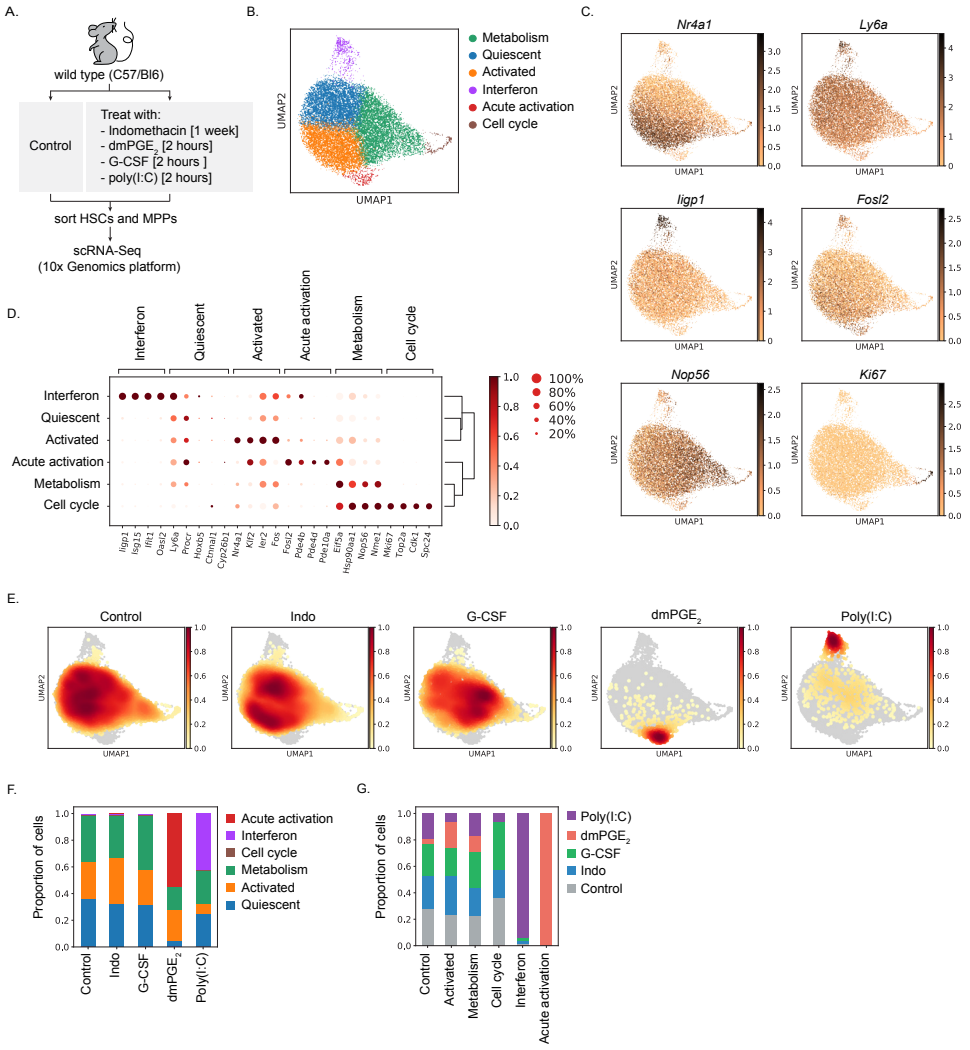


Figure 1. HSCs are transcriptionally heterogeneous and niche perturbations rapidly shift cells into different states. (A) Schematic overview of stimulant treatment before HSC and MPP isolation. (B) UMAP plot of HSC clusters. (C) UMAP plot with expression of representative genes for each cluster. (D) Dot plot of enriched genes for each HSC cluster. Expression is scaled across columns. (E) UMAP density graphs of HSC distribution upon treatment with each niche stimulant. (F) Proportion of HSCs within clusters for each niche perturbation. (G) Proportion of HSCs from each niche perturbation within a cluster normalized for the total cell number per treatment.

of control HSCs sorted from different mouse strains (Supplemental Figure 2A-2D, Supplementary Table 4). The absence of a clear separation into highly distinct clusters in UMAP space (Figure 1B), together with fact that most marker genes are not exclusively expressed but rather enriched in a given cluster (Figure 1C), suggests that HSC reside in transcriptional states with continuous transitions as opposed to discrete subtypes of HSCs. Reactome pathway enrichment analysis (Supplemental Figure 2E) in combination with manual curation of enriched genes (Figure 1D, Supplementary Table 4) allowed us to

assign tentative labels to these states. Three HSC clusters comprised 98% of control cells (Figure 1F). The remaining 2% of control HSCs found in two additional clusters. Consistent with the FACS results (Supplemental Figure 1E), the proportion of HSCs residing in a 'Cell cycle' cluster marked by genes such as *Ki67* was very low (1%; Figure 1C, 1F). A prominent subpopulation contained cells that were defined by various IEGs including *Nr4a1*, *Ier2* and *Fos* (Figure 1B-1D). Therefore, we named this cluster 'Activated'. HSCs have been tightly linked to dormancy and quiescent states[21,22]. The cluster adjacent to the 'Activated' state was termed 'Quiescent' because cells within this population showed the highest expression of previously described dormancy markers[2-4,23-25] (Supplemental Figure 2F). Furthermore, 'Quiescent' HSCs did not express IEGs (Figure 1C) and were located most distal to the 'Cell cycle' cluster in UMAP space (Figure 1B-1D). The 'Metabolism' cluster was comprised of the highest metabolically active HSCs and showed enrichment of transcripts involved in translation initiation (*Eif5a*, *Eif4a1*), nucleotide metabolism (*Nme1*, *Dctpp1*), and ribosome assembly (*Ncl*, *Nop56*, *Nop10*, *Npm1*). The 'Metabolisms' state was positioned adjacent to the 'Cell cycle' state in UMAP space (Figure 1B-1D). We next evaluated whether treatment with niche stimulants shifted the distribution of cells within these distinct states (Figure 1E). Interferons induced by poly(I:C) treatment increased the proportion of HSCs within the 'Interferon' cluster from 1% to 42% (Figure 1F, p-value (DPA) < 10^{-5}). The 'Interferon' cluster was characterized by expression of interferon response genes such as *Iigp1*, *Isg15*, *Ifit1*, and *Oasl2* (Figure 1D). *In vivo* treatment with dmPGE$_2$ gave rise to a novel cluster named 'Acute activation' (Figure 1B) that contained 55% of dmPGE$_2$ treated HSCs (Figure 1F). The cluster itself was entirely composed of dmPGE$_2$ treated HSCs (Figure 1G) and genes marking this state include known cAMP-response genes such as *Fosl2* and the phosphodiesterases *Pde10a*, *Pde4b*, and *Pde4d* (Figure 1C, 1D). G-CSF and indomethacin induced slight shifts in cell distribution compared to the control (Figure 1E) but did not significantly alter cell proportions between different states (Figure 1F; p-value (DPA) > 0.05 for all clusters;). In conclusion, HSCs were equally distributed between three distinct transcriptional states, here defined as 'Quiescent', 'Activated', and 'Metabolism' during homeostasis. Few HSCs reside in 'Interferon' and 'Cell cycle' states during unperturbed conditions. A two-hour *in vivo* pulse with poly(I:C) or dmPGE$_2$ significantly alters the distribution of HSCs between preexisting transcriptional states and allows for novel transcriptional states to surface.

HSC specific single cell phenotypes can be identified by comparison to LSK progenitors

To investigate heterogeneity within HSCs and assess specificity of transcriptional responses to niche signals, we compared HSCs to LSK populations. The LSK compartment encompasses both HSCs and MPPs (Supplemental Figure 1A, 1G and Figure 2A). To evaluate transcriptional response and cell states in phenotypically defined MPP populations known as MPP, MPP1, MPP2, and MPP3/4[22], we barcoded cells with

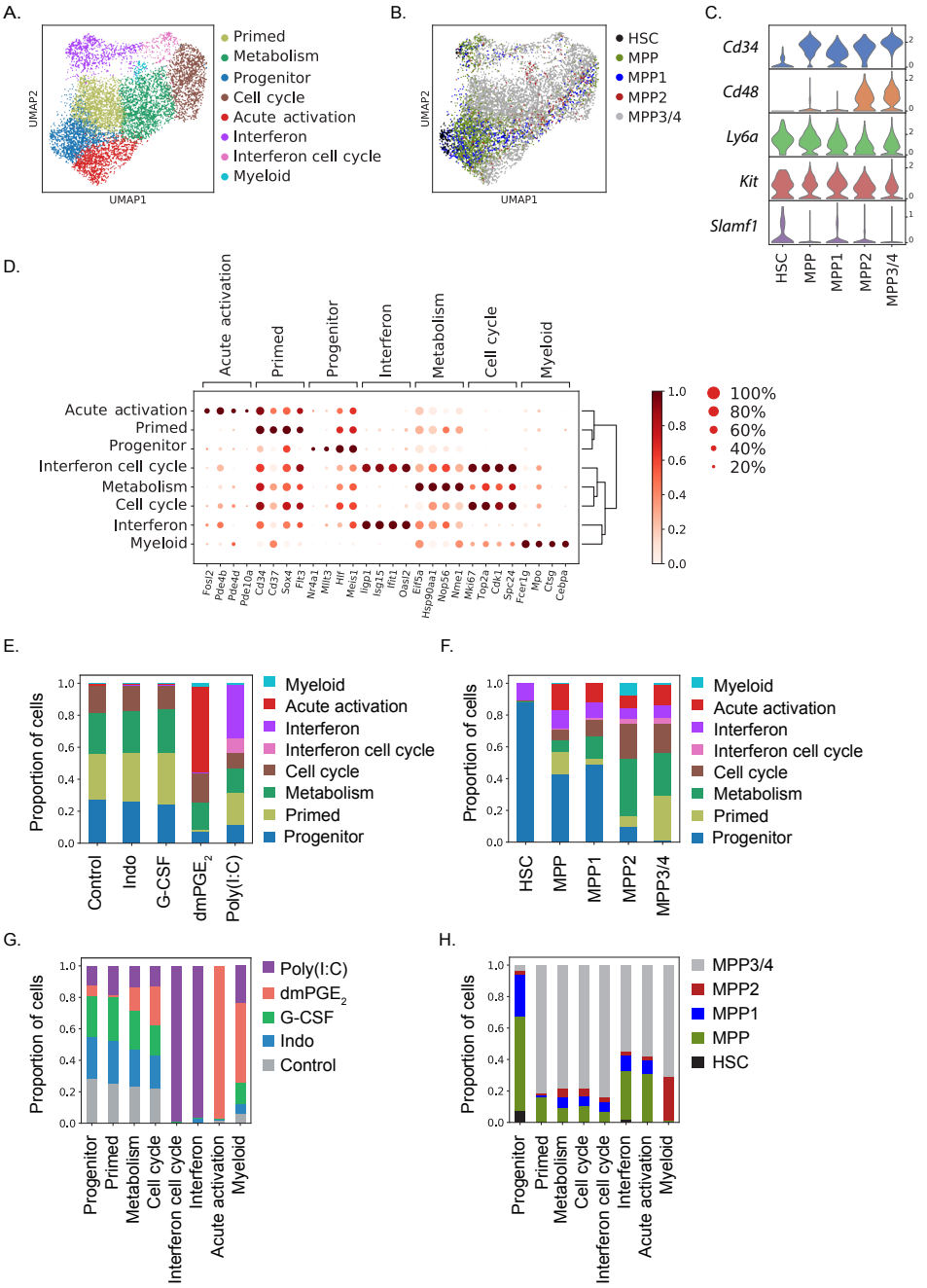


Figure 2. Response to niche signals in MPP surface marker validated LSK is similar to HSCs. (A) UMAP plot of LSK clusters. (B) UMAP plot of surface receptor phenotypes in LSK cells. (C) Stacked violin plots of RNA expression within MPP surface receptor phenotypes. (D) Dot plot of enriched genes for each LSK cluster. Expression is scaled across columns. (E) Proportion of LSK cells within clusters for each niche perturbation. (F) Proportion of LSK cells belonging to different clusters for each surface phenotype. (G) Proportion of LSK cells from each niche perturbation within a cluster normalized for total cell number per treatment. (H) Proportion of MPP surface phenotypes within each LSK cluster.

specific surface phenotypes using a Hashtag Oligonucleotide (HTO) labelling strategy that is part of Cellular Indexing of Transcriptomes and Epitopes Sequencing (CITE-Seq)[26] (Figure 2A, 2B and Supplemental Figure 1A). CITE-Seq enabled us to track specific cell surface phenotypes in our scRNA-Seq data. scRNA-Seq gene expression of markers genes such as *Cd34*, *Cd48* and *Cd150 (Slamf1)* matched the surface phenotypes used for sorting of CITE-Seq barcoded MPPs (Figure 2C, Supplemental Figure 4B). Similar to the approach used for HSCs, we analyzed transcriptomes of LSK cells as an aggregated set consisting of the stimulation- and control conditions together. We discovered a total of eight LSK clusters (Figure 2A). Cell clusters were labelled through analysis of enriched genes (Figure 2D, Supplementary Table 4) and by comparison to earlier defined HSC states (Supplemental Figure 4C). LSK clusters most similar to the 'Quiescent' and 'Activated' HSC state were named 'Progenitor' and 'Primed', respectively. The 'Progenitor' cluster encompassed the majority of phenotypic HSCs and was significantly depleted of MPP3/4s compared to other LSK states (Figure 2H, p-value (DPA) < 0.02). Conversely, phenotypic HSCs were almost exclusively composed of cells in the 'Progenitor' state (Figure 2F). Their location at the edge of the UMAP plot could indicate the origin of HSC differentiation (Figure 2B, Supplemental Figure 4F). LSK cells in the 'Primed' cluster likely represent a more progenitor committed population given their *Cd34* and *Cd135 (Flt3)* expression profile[22] and enrichment in *Cd37* and *Sox4* which suggested commitment to the lymphoid fate[27,28] (Figure 2D). The 'Cell cycle' and 'Metabolism' states contained LSK cells that are mitotically active or on the verge of entering the cell cycle, respectively. The LSK 'Metabolism' cluster shared a similarity of 33% and 41% in top 100 most enriched genes with the HSC 'Metabolism' and 'Cell cycle' clusters, respectively (Figure 2D and Supplemental Figure 4B, 4C). Consistent with previous reports[29], MPP2s contained the highest proportion (7.6%) of myeloid cells and comprised 28% of the 'Myeloid' cluster (Figure 2F, 2H). Control treated LSKs were distributed amongst four main clusters, that is 'Primed', 'Progenitor', 'Metabolism' and 'Cell cycle', that together encompass 99% of all control LSK cells (Figure 2E).

Similar to our observations in HSCs, G-CSF and indomethacin treatment did not alter cellular distributions between LSK cluster (Figure 2E, Supplemental Figure 4G). $dmPGE_2$ and poly(I:C) gave rise to novel clusters that were absent in control LSKs (Figure 2E, 2G). These treatment-induced states displayed transcriptional profiles that were similar to the HSC equivalents (Supplemental Figure 4C). Surprisingly, and in contrast to HSCs, no 'Interferon' responsive cell state was present in LSKs at baseline (Figure 2E). Poly(I:C) treatment induced two interferon responsive clusters in LSKs, of which one showed higher mitotic activity hence named 'Interferon cell cycle' (Figure 2D, Supplemental Figure 4C). In summary, single cell transcriptome analysis of CITE-Seq validated LSK populations revealed an increased proportion of lineage-committed and mitotically active cells when compared to HSCs. This is consistent with previous findings that that HSCs are the most quiescent LSK population while others within this compartment are more actively

cycling[22]. The transcriptional responses to niche perturbation as well as the changes in distribution between cell states seen upon niche stimulation were highly similar between HSCs and the LSK compartment.

Within cluster analysis detects cell state specific effects of niche perturbations

We next used the MAST framework to study transcriptional heterogeneity within cell states[30]. This compared stimulated versus control cells by performing differential expression analysis within cell clusters. We then compiled differentially expressed genes (DEGs) from every clusters per treatment condition (Figure 3A-3D). DEG were assessed at three levels, those being a genes expression change ≥ 1.5-fold (1), ≥ 1.2-fold (2) or without a fold change cutoff applied (3). Additionally, all DEG were required to have an FDR < 0.1 (Supplementary Table 5). We then aggregated genes based their common or unique expression profiles in HSCs or LSKs (Figure 3A-3D).

G-CSF treatment perturbed gene expression more strongly within LSKs (green bars, Figure 3A) whereas interferon stimulation by poly(I:C) predominantly affected HSCs (Figure 3B, purple bars). Receptor expression does not fully explain this difference since both the G-CSF receptor Csf3r and the type I Interferon receptors *Ifnar1* and *Ifnar2* were expressed in a higher proportion of LSK cells compared to HSCs (Figure 3E, 3F). Indomethacin was found to selectively affect HSCs but only very few genes showed a substantial differential expression (Figure 3C). $dmPGE_2$ led to a balanced effect on HSCs and LSKs, with neither compartment dominating the differential gene expression (Figure 3D). We further analyzed DEGs that were unique for specific cell clusters within either HSCs or LSKs. $dmPGE_2$ treatment decreased expression of genes that promote the cell cycle such as *Aurka*, *Plk1*, and *Ki67* specifically within the LSK 'Cell cycle' cluster (Figure 3I, 3G; red bar, and Supplemental Figure 6C). To mimic a dataset that could be obtained by bulk RNA-Seq after sorting of MPPs, we analyzed DEGs of MPPs based on surface phenotypes rather than transcriptional clusters. DEG analysis within MPP surface phenotypes failed to recover $dmPGE_2$ mediated regulation of cell cycle genes (Figure 3H, in red). This is likely because the 'Cell cycle' cluster only contains 6% (MPP) to 22% (MPP2) of cells across all MPP populations which results in dilution of the effect in pseudo-bulk MPP analysis (Figure 2F). Pseudo-bulk analysis also failed to capture the effect of G-CSF on the 'Cell cycle' state (Supplemental Figure 4D, 4E). These results illustrate that cluster based scRNA-Seq analysis can identify genes that differentially regulated by specific niche signals in particular transcriptional states within a cellular compartment. Furthermore, the data revealed cell-state specific transcriptional responses to niche signals in HSCs compared to LSKs that were not identified by earlier approaches, with the cell cycle state being prominently affected by various stimuli.

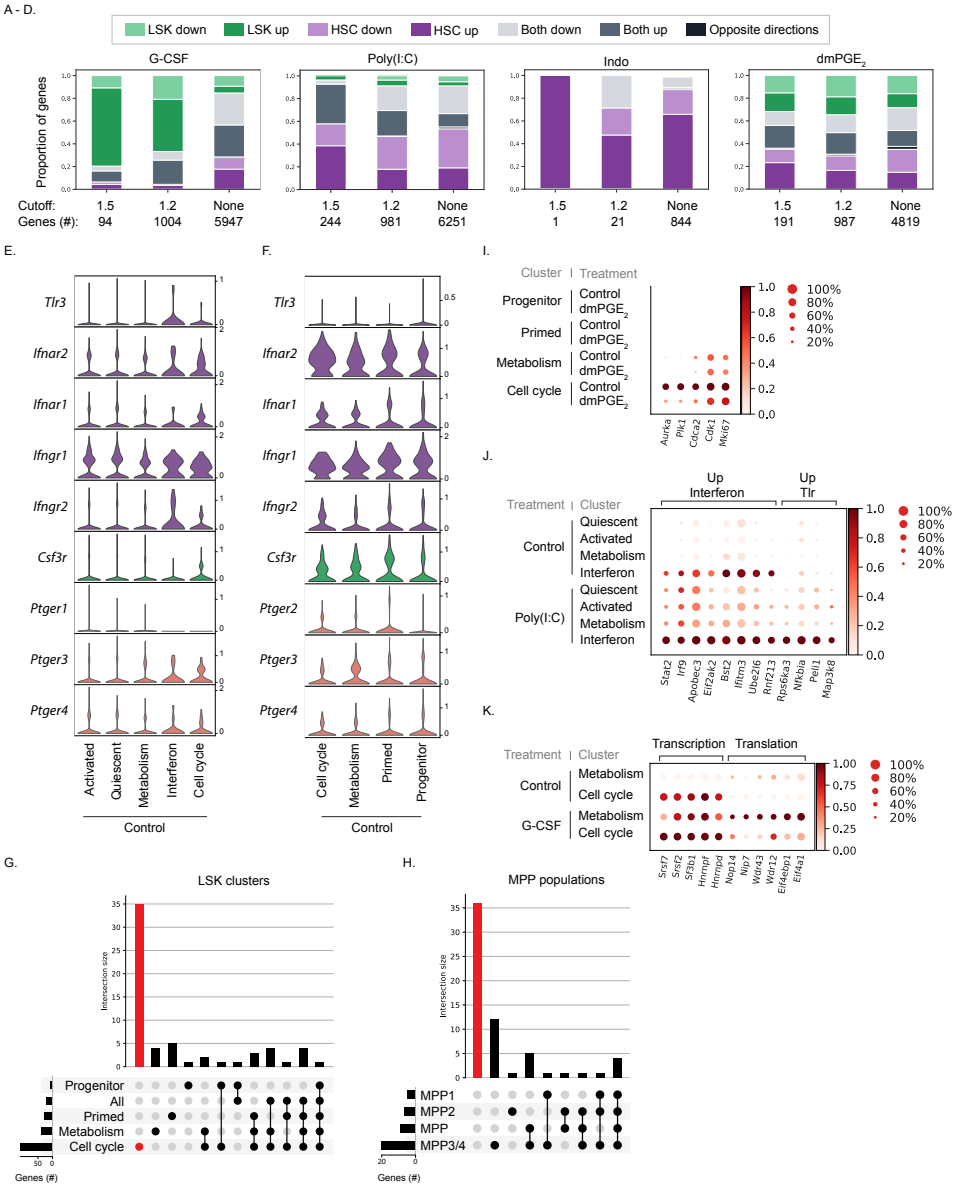


Figure 3. Cluster specific differential gene expression analysis reveals specific effects on HSC and LSK subpopulations. (A-D) Stacked bar graphs with proportion of differentially expressed genes that are unique to HSCs, MPPs, or overlapping in both compartments for G-CSF (A), poly(I:C) (B), indomethacin (C), and dmPGE$_2$ (D) treatment. The total number differentially expressed genes at a given fold change cutoff is listed below each bar graph. (E-F) Violin plots of receptor expression in control HSCs (E) and LSKs (F) split by cluster. (G) Upset plot visualizing differentially expressed genes between dmPGE$_2$ and control for each LSK cluster (horizontal bars) and intersection of gene sets between clusters (vertical bars). Red bar and dot indicate specific genes in LSK 'Cell cycle' cluster. (H) Upset plot visualizing differentially expressed genes between dmPGE$_2$ and control cells for each MPP surface phenotype. Red bar without red dot indicates genes that are absent from any of the MPP surface phenotype conditions. (I) Dot plot of representative cell cycle genes and their expression in LSK clusters in control and dmPGE$_2$. Expression is scaled across columns. (J) Dot plot of representative poly(I:C) response genes in treated and control HSC clusters. Expression is scaled across columns. (K) Dot plot of representative G-CSF response genes from the G-CSF treated and control HSC clusters. Expression is scaled across columns.

Endogenous cell states distinguish TLR- and IFN-specific response of poly(I:C) treatment

To identify distinct patterns of regulation for different stimulations, we selected genes that were differentially expressed within at least one HSC or LSK cluster. We then performed hierarchical clustering for the respective DEGs in HSCs (Supplemental Figure 5, Supplementary Table 6) and LSKs (Supplemental Figure 6, Supplementary Table 7). Poly(I:C) binds to Toll-like receptor 3 (Tlr3)[31] which leads to expression of type 1 interferons (IFNα and IFNβ) that in turn signal via IFNα/β receptor 1 *(Ifnar1)* and 2 *(Ifnar2)* heterodimers, all of which are expressed in HSCs (Figure 3E). We identified two expression patterns in poly(I:C) treated HSCs that are consistent with Toll-like receptor and interferon receptor signaling. The first expression pattern is characterized by induction of poly(I:C) response genes across all cell states (Figure 3J). In addition, we found poly(I:C) response genes to be specifically enriched in the 'Interferon' cluster already prior to stimulation (Figure 3J; Up interferon). Genes following this expression pattern were either directly downstream of type 1 interferon receptors, such as *Stat2*, and *Irf9*, or act as effector proteins in the viral interferon response, such as *Apobec3*, and *Eif2ak2* (Figure 3J, Supplemental Figure 5A). The high expression of interferon response genes *(e.g. Bst2, Ifitm3, Ube2l6,* and *Rnf213)* in the control 'Interferon' cluster might point to a state of general surveillance for viral infection at baseline (Figure 3J). The second expression pattern results from genes that are strongly induced by poly(I:C) treatment, especially in the 'Interferon' cluster, but show low expression at baseline in all control cell states. (Figure 3J, Supplemental Figure 5A). Genes within this signature include *Nfkbia*, *Peli1*, *Map3k8*, and *Rps6ka3* and are part of TNFα and Toll-like signaling pathways. Therefore, this expression profile might represent a more direct response to poly(I:C) interaction with Tlr3. The comparison of differential expression patterns across cell states allowed us to distinguish between poly(I:C)-mediated TLR- and interferon based signaling responses.

G-CSF triggers changes within metabolically active HSCs and LSKs

G-CSF has been identified as a potent enhancer of granulocyte and neutrophil differentiation[32]. It is widely used as a mobilizing agent for hematopoietic stem cells in a therapeutic setting[33]. In line with its clinical use, we found niche adhesion receptors such as *cKit* and *Cd9*[34] to be downregulated after *in vivo* G-CSF treatment (Supplemental Figure 5B). The transcriptional downregulation of adhesion receptors observed here is consistent with the use of G-CSF as HSC mobilization agent. Hierarchical clustering analysis of DEGs indicated that HSCs in the 'Metabolism' state are most responsive to G-CSF. G-CSF treatment alters the expression profile of the 'Metabolism' cluster to become more similar to the 'Cell cycle' states (Supplemental Figure 5B). Indeed, genes related to transcription, such as RNA binding proteins *(Hnrnpd, Hnrnpf, Hnrnpa2b1)*, as

well as splicing factors *(Srsf7, Sf3b1, Srsf2)* were induced by G-CSF. Induction of these genes led to resemblance of the 'Cell cycle' state, where their expression was high both prior and after G-CSF stimulation (Figure 3K). G-CSF also increased expression of transcripts involved in translation, including genes involved in ribosome biogenesis *(Nop14, Nip7, Wdr43, Wdr12)* and translation initiation *(Eif4a1, Eif4ebp1)*. These translational genes were not as strongly expressed in the 'Cell cycle' states (Figure 3K). Overall, a two-hour pulse of G-CSF pushed HSCs towards a more metabolically active state. This is consistent with the function of G-CSF as a growth factor that regulates myeloid differentiation[32].

Endogenous prostaglandins regulate immediate early genes within 'Activated' cell states

To investigate loss of niche signaling in a more physiological setting compared to genetic perturbations, we treated mice orally for one week with indomethacin to deplete endogenous prostaglandins. Differential expression analysis revealed 21 genes with a 1.2-fold change in expressed after indomethacin treatment (Figure 3A). The majority of genes can be classified as IEGs *(Fos, Fosb, Jun, Klf4, Klf6)* and ribosomal proteins *(Rps21, Rpl36)* (Supplemental Figure 5D). While indomethacin did not alter cell proportions between the HSC clusters (Figure 1F), cells distribution shifted slightly towards the periphery within the UMAP plot (Figure 1E). This change in distribution was also seen when examining individual cell cluster marker genes such as *Fos* and other IEGs (Figure 4A, Supplemental Figure 7A). To further investigate the influence of indomethacin on cell distribution while taking the entire transcriptional landscape into account, we computed diffusion pseudotime (DPT)[35] between the 'Activated' and 'Quiescent' cluster in HSCs. The cell with the highest combined expression of top enriched 'Activated' cell cluster markers (Figure 4B; left panel) was set as root cell and DPT was calculated based on similarity to the root cell (Figure 4B; right panel). Indomethacin-treated cells displayed a significant shift to the left in overall pseudotime kernel density distribution (KDE), which is indicative of overall lower pseudotime (Figure 4C, p-value (Mann–Whitney U-test) = 5.8×10^{-12}). No shift was observed when comparing control HSCs to G-CSF treated HSCs (Figure 4C, p-value = 0.18). Ranking cells for each treatment condition according to pseudotime and averaging gene expression in 10 equally sized bins (quantile rank 1-10) further illustrated the change in expression of *Fos* and other IEG genes after indomethacin, especially at lower pseudotime (Figure 4D, Supplemental Figure 7B; indicated by asterisks). Genes that were not part of the activated gene signature, such as *Ly6a*, did not show such a pattern (Figure 4D) nor was this trend observed in response to G-CSF (Supplemental Figure 7C). This indicates that lower pseudotime is specific to indomethacin treatment. The pseudotime analysis indicated a shift in transcriptional states in indomethacin treated HSCs which could point to a change in turnover of IEG expression. To confirm the effect of indomethacin on IEG turnover in

an orthogonal assay, we measured single cell protein levels of FOS by intracellular flow cytometry. A seven-day *in vivo* indomethacin treatment led to a 34% ± 8.2% reduction in FOS mean fluorescent intensity (MFI) in HSCs (p-value of T-test with Welch correction = 0.0062) and a 35% ± 8.6% decrease in LSKs (p = 0.0066, Figure 4E, 4F). Overall, endogenous prostaglandin levels impact both transcript and protein levels of FOS and potentially other IEGs.

HSC specific chromatin architecture is an intrinsic regulator of differential responses to niche signals

To understand intrinsic factors that regulate the transcriptional receptiveness to niche signals, we assessed chromatin states of HSCs and MPPs by scATAC-Seq. We clustered cells based on chromatin accessibility in HSCs (Figure 5B) and LSKs (Figure 5E, 5F). To gain insight into differential accessible chromatin regions, we computed a transcription

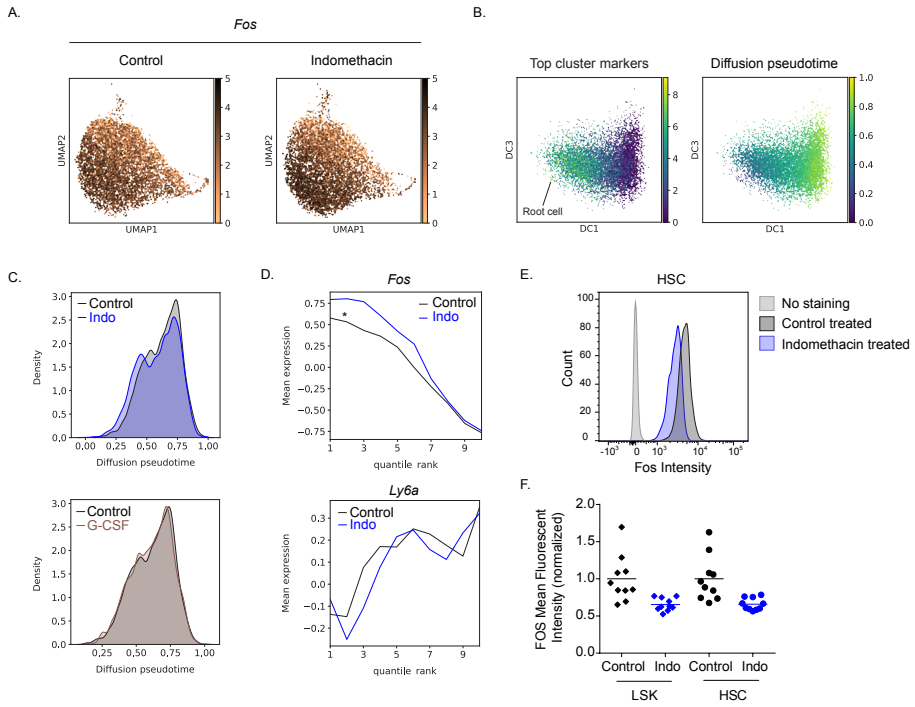


Figure 4. Indomethacin treatment induces changes in IEG transcriptional state in HSCs. (A-H) Diffusion pseudotime analysis. (A) UMAP plot of *Fos* expression in control and upon indomethacin treatment. (B) Diffusion map embedding with combined expression of activated genes to select root cell (left panel) and cells colored by pseudotime (right panel). (C) Kernel density of cell distribution across pseudotime comparing indomethacin and control (upper panel) and G-CSF and control (lower panel). (D) Average expression of *Fos* and *Ly6a* across cells ranked by pseudotime. Cells were split into 10 bins to decrease noise. The change in gradient indicated by the asterisk. (E) Histogram of intracellular FACS in HSCs for FOS. No stain indicates the FACS negative control, Control indicates FOS expression in untreated mice. (F) Normalized mean fluorescence intensity (MFI) for FOS in control and indomethacin treated HSCs (p-value = 0.0062) and LSK cells (p = 0.0066) across two independent experiments. Statistical analysis was performed by T-test with Welch correction.

factor (TF) motif activity score on a per cell basis using ChromVar[36] and evaluated motif enrichment across clusters. The motif activities of the TFs STAT3, NF-κB, and CREB1 that are immediately downstream of G-CSF, Poly(I:C), and Prostaglandins (Figure 5A), respectively, were homogeneously distributed in HSCs and LSKs (Supplemental Figure 8A, 8B, and Supplementary Table 8). This suggested that cells have an equal potential to respond to these signals, based on chromatin states of the TF motifs. We did, however, detect differential motif enrichment of TFs that are further downstream in the signaling response. Specifically, we found differential enrichment of interferon signaling response element (ISRE) motif activities in cluster 1 (Figure 5C). Interferon-regulatory factors (IRFs) that bind ISREs are induced by NF-κB signaling and also direct targets of intracellular poly(I:C) binding[37] (Figure 5A). In addition to ISRE, HSC cluster 1 is characterized by motifs for transcriptional regulators that exert roles in cellular metabolism, cell growth, and differentiation such as CTCF, YY1, and NRF1 (Figure 5G, Supplemental Figure 8C). LSK cluster 5 contains a combination of motif activities most similar to HSC cluster 1, which suggests that this motif activity combination marks cells that are committing to or undergoing differentiation. (Figure 5H, Supplemental Figure 8G).

HSC cluster 0 showed enrichment of AP-1 motif activities (Figure 5D). AP-1 motifs can be bound by FOS and JUN which are downstream effectors of Prostaglandin/CREB1 signaling[38] (Figure 5A). SMAD motifs, elements binding the SMAD family of signal-responsive TFs, is also enriched in HSC cluster 0 (Supplemental Figure 8E, 8F). We furthermore found motif activity enrichment for lineage specific master TFs (RUNX, GATA and PU.1/SPI1) in HSC cluster 0 (Figure 5I, Supplemental Figure 8D). In contrast to HSCs, LSK cells did not contain a corresponding cluster where signaling TF- and lineage specific master factor motifs co-occur (Figure 5J, Supplemental Figure 8H). The chromatin features of HSC cluster 0 indicates an HSC specific chromatin state.

We found that chromatin heterogeneity related to factors further downstream in the response, such as AP-1 and IRFs, may underlie cell intrinsic variability to stimulation. Comparison of chromatin accessibility in unperturbed HSCs and LSKs revealed HSC specific chromatin heterogeneity that could be indicative of intrinsic determinants that predispose for specific cellular responses to niche signals.

Discussion

Here, we provide the first transcriptional and epigenetic single cell analysis of highly purified, functionally validated HSCs. Our work reveals that HSCs exist in fluent transcriptional and epigenetic states rather than in distinctly separated cell populations. Niche perturbations shift distributions between HSC states within hours of signaling. This provide evidences that the transcriptional states are highly dynamic, allowing HSCs to rapidly transition between states in response to different stimuli. Furthermore, we identify cell intrinsic chromatin heterogeneity predisposes specific HSC subpopulations for certain

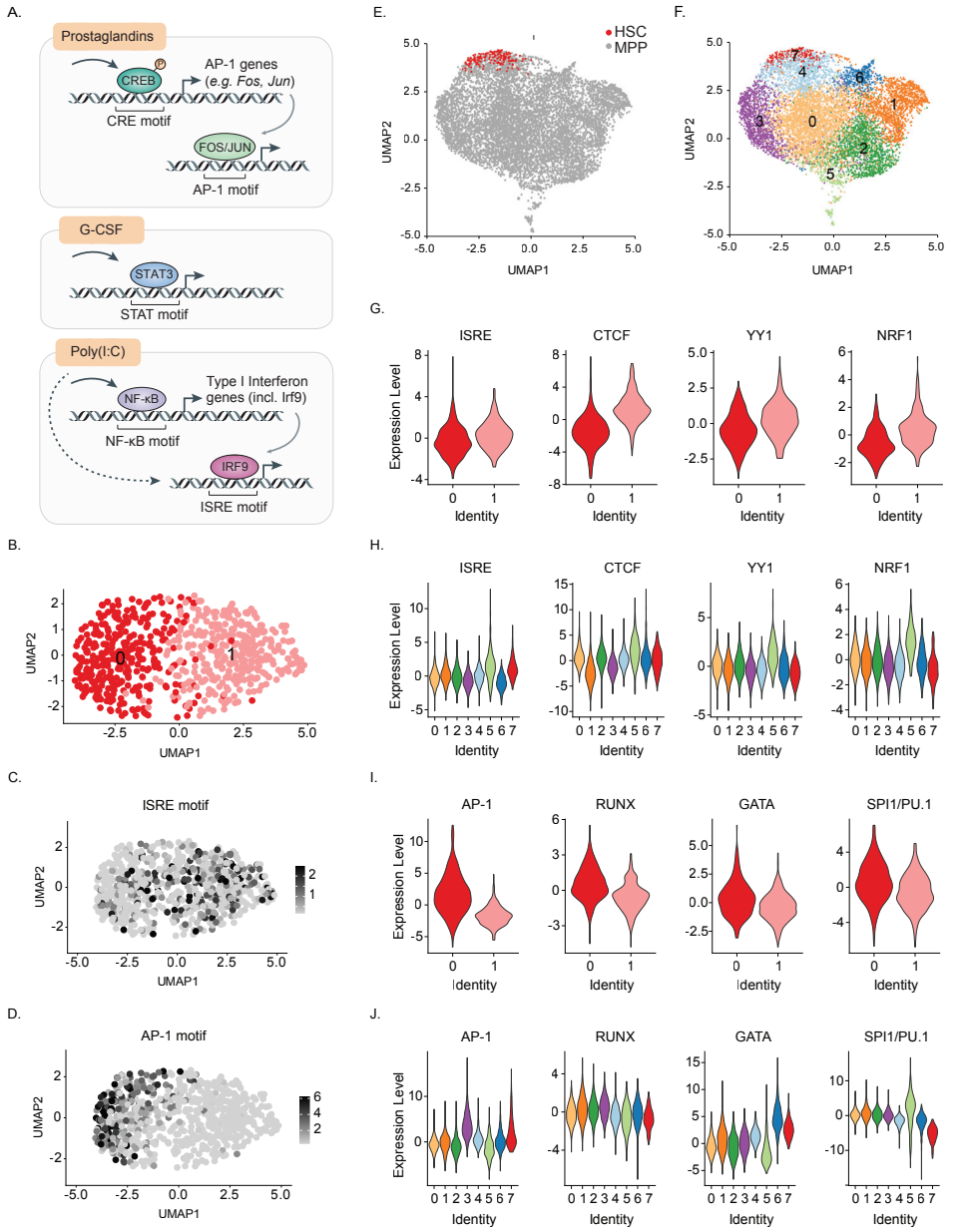


Figure 5. Specific transcription factor binding motif co-ocurrences in HSCs. (A) Schematic of downstream transcriptional signaling pathways for niche stimulants. (B) UMAP plot of HSC scATAC-Seq clusters. (B, C) UMAP plot of ISRE (C) or AP-1 motif activity. (E, F) UMAP plot of LSKs colored by surface phenotype (E) and scATAC-Seq clusters (F). (G-J) Violin plots of motif activity enriched in HSC (G, I) and LKS (H, J) clusters.

niche signals. We detected HSC specific co-occurrence of signaling and lineage specific transcription factor motif activities, which is consistent with our previous observation in human hematopoietic progenitors[39].

Our combined scRNA-Seq and CITE-Seq approach uncovered insights into the transcriptional landscape of HSCs and phenotypically defined MPP populations at steady state and following niche perturbations. Furthermore, we assessed found sexual dimorphism using *Xist* expression. We observed only week sex differences in response to the niche signals studied here. Since sexual dimorphisms in HSCs to other signals such as hormones has been described, these sex differences are may highly specific to the type of signal[40,41]. The combinatorial method of scRNA-Seq coupled with CITE-Seq and gender deconvolution uncovers deeper insights into cell populations than enabled by scRNA-Seq alone.

We used a two-pronged strategy to evaluate the specificity of niche perturbations on HSCs or LSKs. First, we assessed stimuli-induced transitions between cell states. Second, we evaluated differential expression within particular cell states. The strength of transcriptional perturbation could not solely be explained by the distribution of cells among cell states alone. G-CSF, for example, did not change cell proportions between clusters but rather elicited strong transcriptional responses within specific cell states that were defined prior to signaling. Analysis of DEGs within clusters also helped tease apart interferon- versus Toll-like receptor response genes induced by poly(I:C) treatment. In addition, we uncovered specific effects of dmPGE$_2$ that within the 'Cell cycle' states. The findings on dmPGE$_2$ described here corroborate results from a recent study describing that dmPGE$_2$ affects cell cycle states in the bone marrow[42]. Thus, we found that niche perturbations induce transition of HSCs between cell states and within specificity states.

There is a tradeoff between the strength of a perturbation required for experimental robustness versus the study of signals that are of higher physiologically relevance but lead to subtle changes within and between cells. We evaluated responses of HSCs to three different activators that were 2-4 orders of magnitude higher than what an animal would encounter during actual injury or infection[43-45]. To evaluate niche derived signals in a more physiological setting, we depleted endogenous prostaglandins using the Cox1/2 inhibitor indomethacin. As expected, the changes in gene expression in response to indomethacin were much weaker than those observed after injection with prostaglandin E$_2$, G-CSF, and poly(I:C). scRNA-Seq analysis empower us to evaluate gene expression changes in response to weak, but physiologically relevant, perturbations. Indomethacin leads to a small but significant shift in HSC transcriptional states, especially affecting expression of IEGs. The effect of indomethacin on IEGs such as *Fos* was further validated in orthogonal assays. How exactly the discrepancy between increased transcription levels of *Fos* observed in certain single cell subpopulations can be reconciled with decreased protein levels determined by FACS analysis will need to be addressed in future studies. This observation, however, illustrates an important implication and potential caveat that

transcript and protein levels may not always positively correlate, even on a single cell level. Additional mechanisms such as mRNA and protein stability, translation rates and protein turnover may play an important role during the physiological responses to niche signals. Regardless, scRNA-Seq technologies provides is sensitive to evaluate subtle changes in cellular states.

In summary, we showed that HSCs exist in dynamics transcriptional states and niche signals can induce rapid transitions between, as well changes within, these cell states. We also revealed differential motif enrichment of TFs that are further downstream in the signaling response that may underlie cell intrinsic variability to stimulation. While our work does not reveal whether the transcriptional states are associated with specific niches *in vivo*, novel spatial transcriptomic approaches provide exiting new opportunities to address such questions[46]. Additionally, recently developed barcoding strategies enable combined assessment of niche induced transcriptional changes and functional potential of single cells[47]. Understanding endogenous levels of niche derived factors and the associated transcriptional and epigenetic responses will advance our basic understanding of stem cells and their potential applications in the clinic.

Methods

Mice and chemical treatment

Male and female mice (8-10 weeks) were ordered from Jackson labs (strain CD45.2 (Ly5.2), #00664). Mice were kept for at least 1 week in the animal facility before initiating experiments. Indomethacin (Sigma, 6mg/l) was administered for 7 days in acidified drinking water to maintain stability[48,49]. Indomethacin-supplemented drinking water was changed every other day. Mice were injected with the following drugs and euthanized after 2 hours: poly(I:C) HMW (Invivogen) through intraperitoneal (IP) injection at 10mg/kg[50], G-CSF Recombinant Human Protein (Thermo fisher) by IP injection at 0.25mg/kg[51] and dmPGE_2 (Cayman) through subcutaneous injection at 2mg/kg[45]. Mice were weighed before injection and injection volume was adjusted to ensure equal dose between individual mice. For the pilot experiment, we used the following mouse strain #016617 that was obtained from Jackson labs but bred in-house. All animal procedures were approved by the Harvard University Institutional Animal Care and Use Committee.

Bone marrow preparation and Fluorescence activated cell sorting (FACS)

Whole bone marrow (WBM) was isolated from femur, tibia, hip and vertebrae via gentle crushing using a mortar and pestle. Stem and progenitor cells were enriched via lineage depletions (Miltenyi Biotech, 130-090-858). Antibodies, dilutions and vendors are listed in Supplementary Table 9. Cells were stained for 1.5 hours based on published best practice protocols for assessing CD34 labelling[52]. Cells were sorted on a FACSAria (Becton Dickinson) and representative sorting scheme is shown in Supplemental Figure 1A. Purity of >80% was ensured by reanalyzing each sorted population.

Sample batching

For each chemical treatment, 5 female mice and 5 male mice were treated. Because of sample processing times, a maximum of two conditions could be performed on the same day which resulted in three separate days of experiments. To mitigate batch effects resulting from different experimental days, the following precautions were taken: (1) All mice included in the chemical treatment were ordered from the same batch from JAX, (2) Control mice were administered acidified water as well as injected with DMSO to control for both unspecific perturbations that might result from the chemical treatments, (3) All experiments were performed within less than one week and all single-cell libraries were prepared at the same time after the initial droplet reaction was frozen, (4) FACS gates were set up initially but left constant for each experiments. Single color controls as well as fluorescence minus one (FMO) controls ensured that there was minimal technical drift between days on the FACS instrument.

Intracellular staining for FACS

WBM extraction, lineage depletion and surface marker staining were performed as described above. Cells were fixed and permeabilized for intracellular staining according to manufacturer's instructions (BD biosciences, 554714). Intracellular staining was performed for 30 minutes on ice. Samples were analyzed on an LSRII FACS analyzer.

Limit dilution transplantation assay

Recipient CD45.2 (Jax #00664) mice were gamma-irradiated (Cs-137 source) with a split dose of 5.5Gy each one day before transplantation. HSCs were isolated from CD45.1 (Jax #002014) donors and transplanted with 200,000 (CD45.2) WBM cells via retro-orbital injection. Donor cell engraftment was monitored monthly for 16 weeks using a LSRII FACS analyzer (Becton Dickinson). Flow cytometry data were analyzed with FlowJo (Tree Star). HSC frequency was calculated using http://bioinf.wehi.edu.au/software/elda/.

Single-cell RNA and ATAC library preparation and sequencing

Male and female cells were sorted separately but pooled in equal ratios before further downstream processing. For CITE-Seq HTO labelling of MPP populations, 0.25µg of TruStain FcX™ Blocking reagent (Biolegend) was added for 10 minutes on ice. Each MPP populations was labelled with 1µg of TotalSeq™ antibody cocktail (Biolegend, see Supplementary Table 9) and incubated for 30 minutes on ice. After washing, the cells were resuspended in small amounts, counted and pooled in equal ratios. Each drug treatment condition resulted in one pooled MPP and one HSC sample that was processed separately for scRNA-Seq. Single cell libraries were prepared according to manufactures recommendations (10X Genomics, 3' V2 for pilot experiment and V3 for chemical treatment experiments). Briefly, for pooled MPPs no more than 10,000 cells were loaded, for HSCs all sorted cells (between 2,222 sorted events for $dmPGE_2$ and 12,017 sorted events for control) were loaded on the 3' library chip. For preparation of HTO labelling, surface libraries manufacturers recommendations (Biolegend) were followed. For ATAC-Seq, HSCs and MPPs (pooled MPP, MPP1, MPP2 and MPP3/4) were sorted as described above from 5 male and 5 female mice (strain CD 45.2 (Ly5.2), JAX strain #00664). Nuclei were isolated and libraries were prepared using manufacturers recommendations (10x Chromium Single Cell ATAC). Libraries were sequenced on a Nextseq (Pilot experiment, scRNA-Seq) and NOVAseq (chemical treatments, scRNA-Seq, scATAC-Seq).

Computational quantification and statistical analysis

All code and a detailed description of the analysis is available in the following Github repository: https://github.com/evafast/scrnaseq_paper. To ensure reproducibility in the entire analysis (except for cellranger and CITE-Seq count) was performed in Docker containers. Containers used for the analysis are indicated in the Jupyter notebooks and corresponding images are available on dockerhub (repository: evafast1). The interactive

cell browser web application is available here: http://gbweb.tchlab.org/cells/app/.

Demultiplexing and generation of count matrices

Cellranger (v3.0.1) command 'mkfastq' was used to demutliplex raw base call (BCL) files into individual samples and separate mRNA FASTQ files and HTO surface fastq files. The cellranger 'count' command was used with default options to generate gene by cell matrices from mRNA FASTQ files. CITE-Seq count (version 1.4.3) was used to generate surface count by cell matrices from the HTO surface FASTQ libraries. For the fresh HSC, the Replicate1 experiment cellranger (v2.1.0) was used for demultiplexing and count matrix generation. The mm10 reference genome was used for all alignments. For scATAC-Seq, cellranger-atac mkfastq and count (1.2.0) was used for demultiplexing, alignment and generation of the fragment file. To generate the count matrix, MACS2 was run with default parameters (keeping duplicates) on the aligned reads. Resulting peak summits were extended to 300bp and counts were extracted from Fragment file using a custom script (see Github repository) to generate a count matrix.

Quality control and Filtering and dimensionality reduction of scRNA-Seq data

The main parts of the bioinformatic analysis of scRNA-Seq data was performed using the python package scanpy[53]. For filtering and quality control, best practice examples were followed[54]. Count matrices were filtered on a gene and cell level. Cells were excluded with either less than 3,000 counts, less than 1,500 (LT) or 2,000 (MPPs) genes or more than 20,000 (LT) or 30,000 (MPPs) counts. A cutoff of no more than 10% mitochondrial genes was applied. Genes expressed in less than 20 cells were excluded from the analysis. Counts were normalized to 10,000 per cell and log transformed. Features (genes) were scaled to unit variance and zero mean before dimensionality reduction. To reveal the structure in the data, we built a neighborhood graph and used the leiden community detection algorithm[55] to identify communities or clusters of related cells (see also below). The UMAP algorithm was used to embed the high dimensional dataset in a two-dimensional space[56]. Differential proportion analysis (DPA) was used for comparing cell proportions between clusters as previously described[20].

Demultiplexing of CITE-Seq hashtag data

We used the DemuxEM[57] implementation in pegasuspy (https://github.com/klarman-cell-observatory/pegasus/tree/0.17.1) to assign MPP surface identities and demultiplex to the pooled MPP sample. First background probabilities ('pg.estimate_background_probs') were estimated using default settings and 'pg.demultiplex' was run adjusting the alpha and the alpha_noise parameter to maximize cell retrieval by singlet classification. Assignments were validated by plotting count matrix in UMAP space and observing four distinct clusters indicative for the four HTO labels that were pooled. The proportion of demultiplexed cells matched the original pooling ratio. Analysis of coexpression of sex

specific genes allowed to further validate doublet rate. Proportion of cells classified by DemuxEM as doublets exceeded doublet rate estimated by coexpression of sex specific genes.

Batch correction

Because of timing required for sort and sample prep, it was not possible to sort HSCs from all conditions on the same day. To correct for potential batch effects, we used ComBat[58] with default settings on the $\log_2$ expression matrix, allowing cells to be clustered by cell type or cell state. Batch correction was similar when we used Scanorama[59]. To correct for potential sex, specific differences Xist counts were regressed out. Raw data was used for all differential expression analysis and plotting of single cell gene expression values. Batch corrected counts were used for clustering and diffusion pseudotime analysis.

Optimal cluster parameter selection

Since HSCs and MPPs are highly purified cell population we did not observe any clearly separated clusters in UMAP space. To aid the optimal choice of hyperparameters for Leiden clustering, we used a combination of Silhouette Coefficient and Davies–Bouldin index. We first validated this approach using the PBMC3K (from 10x genomics, scanpy. datasets.pbmc3k()) silver standard dataset. We iterated through a range of KNN nearest neighbors and Leiden resolution combinations measuring average Silhuette coefficient and Davies-Bouldin index in PCA space for each combination. Plotting the optimal value for Silhouette score and Davies-Bouldin index versus increasing numbers of clusters allowed to determine appropriate cluster number for the dataset. For the PBMC dataset, there was a clear dropoff in optimal value after 8 cluster which is corroborated by most single cell tutorials that also report 8 clusters for this dataset. After validation of this approach on PBMCs, we used assessed Silhouette Coefficient and Davies–Bouldin index for different clustering results of our own HSC and MPP datasets. This allowed us to select the optimal hyper parameters for each cluster number. The approach was validated by using data driven parameters to compare two independent biological replicates of control HSCs ('Replicate 1' and 'Replicate 2').

Differential expression using MAST

Differential expression analysis was performed using MAST ("Model-based Analysis of Single cell Transcriptomics[30]. This method is based on a Hurdle model that takes both the proportion of cells expressing a given transcript as well as transcript levels themselves into account while being able to control for covariates. Based on previous reports, differential expression cutoff was set at 1.2-fold[60] and a more stringent cutoff of 1.5-fold was also included. Only genes that were expressed in at least 5% of the cells were considered for differential expression analysis. FDR cutoff was set at 1%. For

drug treatments, differential expression between treatment and control was assessed within the entire MPP or LT dataset and within each cluster controlling for number of genes per cell and sex. For differential expression analysis between male and female cells at baseline, control datasets were analyzed with clusters and number of genes as a covariates. For sex specific effects of drug treatments, samples were split by sex and analyzed separately. Resulting differential expression coefficients were compared between male and female. To identify gene signatures with common patterns, for each treatment average expression of differentially expressed genes was extracted per cluster, scaled (z-score) and grouped together by similarity using hierarchical clustering (seaborn clustermap, Euclidian distance, single linkage).

Diffusion pseudotime analysis

For diffusion pseudotime analysis[35] cells from the 'Quiescent' and 'Activated' cluster were selected for the following treatments: control, indomethacin and G-CSF. We recalculated PCA and UMAP embeddings in this reduced dataset. Reclustering using the Leiden algorithm was used to exclude outlier cells and assess top enriched genes within the new 'Activated' cluster. Raw expression of the three top enriched genes (Nr4a1, Nr4a1, Hes1) was summed to robustly select the most highly 'activated' cell as a root cell. Diffusion pseudotime was calculated with the following function ('sc.tl.dpt') and using default settings. Cells were ranked according to pseudotime and kernel density distribution was plotted using a bandwidth of 0.02. The Mann-Whitney U-test was used to assess if cells from different samples are drawn from the same distribution. To analyze gene expression across pseudotime, for each sample cells were split into ten equally sized bins according to ascending pseudotime. Bin 1 contained the first 10% of cells with the lowest pseudotime and bin 10 contained the 10% of cells with the highest pseudotime. Average gene expression for representative genes were plotted for each bin and sample.

Single-cell ATAC-Seq

The R package Signac (https://github.com/timoast/signac, version 0.2.5), an extension of Seurat[61], was used for quality control, filtering of ATAC-Seq peaks counts and plotting. Quality of scATAC dataset was ensured by presence of nucleosomal banding pattern and enrichment of reads around transcription start sites (TSS). Cells with < 1000 or > 20000 peak read fragments were removed. Male and female cells were classified according to absence or presence of Y chromosome reads. Since distribution of male and female cells appeared uniform across all analyses, no sex correction was carried out. Term frequency-inverse document frequency (TF-IDF) was used for normalization and dimensionality reduction was performed by singular value decomposition (SVD). Cells were clustered using the Louvain community finding algorithm and neighborhood graphs were built with k = 20 (HSCs) or k = 30 (LSK) nearest neighbors. ChromVAR[36] was run with default parameters using the JASPAR 2018 motif database.

Acknowledgments

We thank E. Hagedorn and S. Avagyan for critical reading of our manuscript; members of the Zon laboratory for discussions on the project. A.S. was supported by a Boehringer Ingelheim PhD Fellowship. This work was supported by the following grants from L.I.Z.: R01 HL04880, P015PO1HL32262-32, 5P30 DK49216, 5R01 DK53298, 5U01 HL10001-05, R24 DK092760, 1R24OD017870-01.

Author contributions

E.M.F., A.S., and M.E.M., performed experiments. E.M.F. performed single-cell computational analysis. E.L.R, J.G., and N.B. assisted with experiments. N.B., Y.Z., S.Y. provided input on computational analysis. E.F., A.S. and L.I.Z. wrote the manuscript.

Conflict of interests

L.I.Z. is founder and stockholder of Fate, Inc., Scholar Rock, Camp4 therapeutics and a scientific advisor for Stemgent. The other authors declare no competing interests.

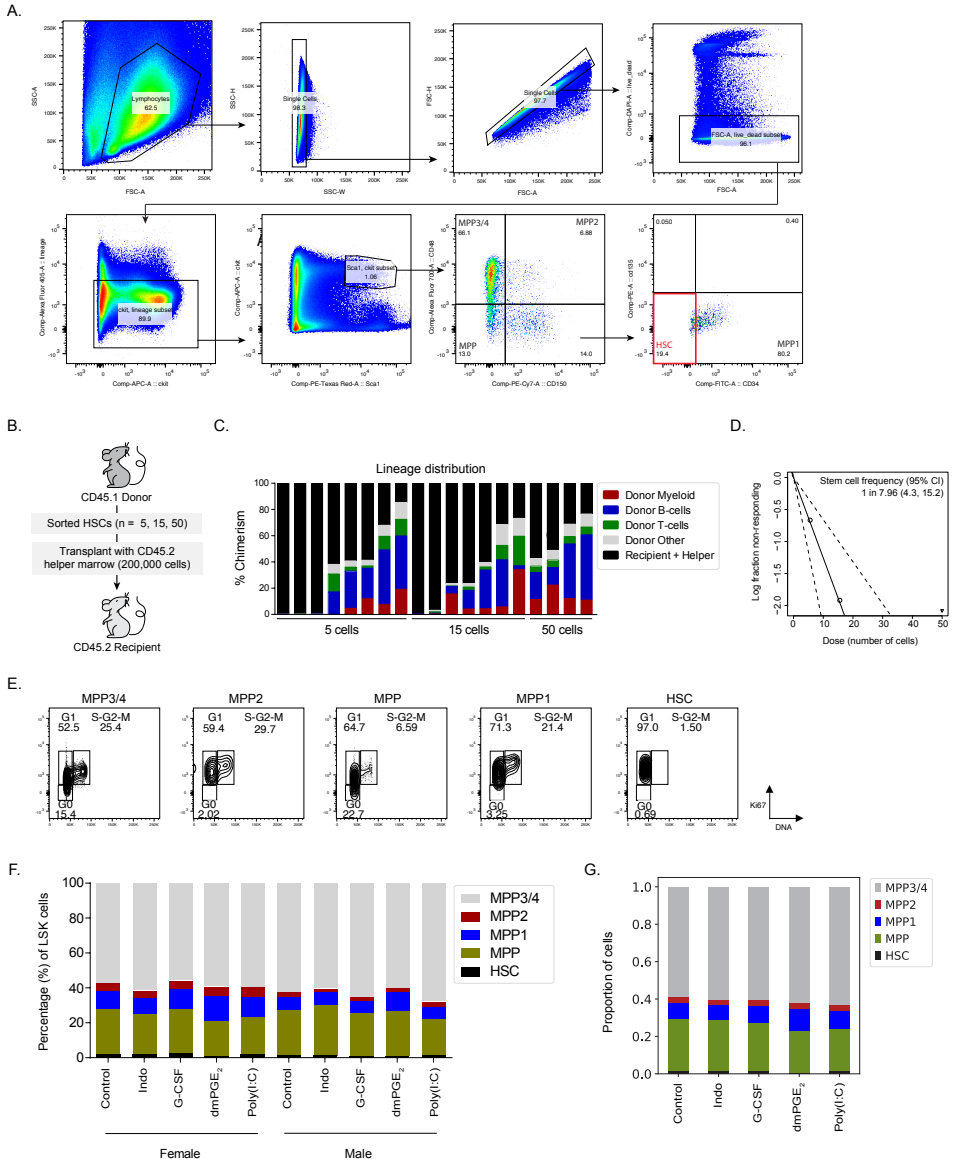
References

1. Blau, H. M. & Daley, G. Q. Stem Cells in the Treatment of Disease. N Engl J Med 380, 1748–1760 (2019).
2. Chen, J. Y. et al. Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. Nature 530, 223–227 (2016).
3. Cabezas-Wallscheid, N. et al. Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. Cell 169, 807–823.e19 (2017).
4. Wilson, N. K. et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell 16, 712–724 (2015).
5. Haas, S., Trumpp, A. & Milsom, M. D. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. Cell Stem Cell 22, 627–638 (2018).
6. Orkin, S. H. & Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell 132, 631–644 (2008).
7. Morrison, S. J., Wandycz, A. M., Akashi, K., Globerson, A. & Weissman, I. L. The aging of hematopoietic stem cells. Nature Medicine 2, 1011–1016 (1996).
8. Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. Cell 132, 598–611 (2008).
9. Zhao, M. et al. Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. Nature Medicine 20, 1321–1326 (2014).
10. Pietras, E. M. et al. Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal. Nature Cell Biology 18, 607–618 (2016).
11. Pinho, S. & Frenette, P. S. Haematopoietic stem cell activity and interactions with the niche. Nature Reviews Molecular Cell Biology 20, 303–320 (2019).
12. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. 20, 207–220 (2019).
13. Avgustinova, A. & Benitah, S. A. Epigenetic control of adult stem cell function. Nature Reviews Molecular Cell Biology 17, 643–658 (2016).
14. Yu, V. W. C. et al. Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. Cell 167, 1310–1322.e17 (2016).
15. Lara-Astiaso, D. et al. Immunogenetics. Chromatin state dynamics during blood formation. Science 345, 943–949 (2014).
16. Buenrostro, J. D. et al. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell 173, 1535–1548.e16 (2018).
17. Lareau, C. A. et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat Biotechnol 37, 916–924 (2019).
18. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods 347, 70–78 (2009).
19. Seita, J. & Weissman, I. L. Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip Rev Syst Biol Med 2, 640–653 (2010).
20. Farbehi, N. et al. Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. eLife 8, 1241 (2019).
21. Foudi, A. et al. Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat Biotechnol 27, 84–90 (2009).
22. Wilson, A. et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell 135, 1118–1129 (2008).
23. Acar, M. et al. Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. Nature 526, 126–130 (2015).
24. Gazit, R. et al. Fgd5 identifies hematopoietic stem cells in the murine bone marrow. J. Exp. Med. 211, 1315–1331 (2014).
25. Balazs, A. B., Fabian, A. J., Esmon, C. T. & Mulligan, R. C. Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow. Blood 107, 2317–2321 (2006).
26. Stoeckius, M. et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet

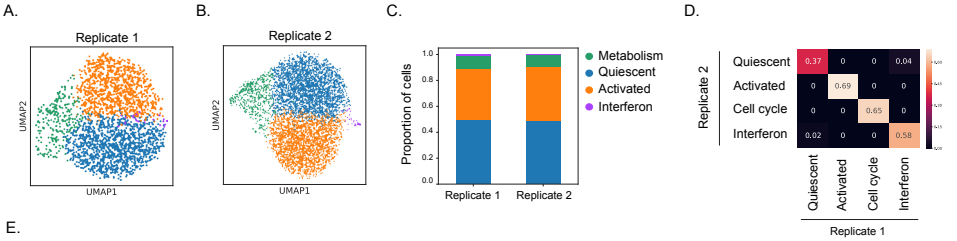
detection for single cell genomics. Genome Biol. 19, 224–12 (2018).
27. Sun, J. et al. Clonal dynamics of native haematopoiesis. Nature 514, 322–327 (2014).
28. Laurenti, E. et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. Nat Immunol 14, 756–763 (2013).
29. Pietras, E. M. et al. Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. Cell Stem Cell 17, 35–46 (2015).
30. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 16, 278–13 (2015).
31. Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. Nature 413, 732–738 (2001).
32. Metcalf, D. & Nicola, N. A. Proliferative effects of purified granulocyte colony-stimulating factor (G-CSF) on normal mouse hemopoietic cells. J. Cell. Physiol. 116, 198–206 (1983).
33. Bendall, L. J. & Bradstock, K. F. G-CSF: From granulopoietic stimulant to bone marrow stem cell mobilizing agent. Cytokine Growth Factor Rev 25, 355–367 (2014).
34. Leung, K. T. et al. The tetraspanin CD9 regulates migration, adhesion, and homing of human cord blood CD34+ hematopoietic stem and progenitor cells. Blood 117, 1840–1850 (2011).
35. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. Nature Methods 13, 845–848 (2016).
36. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nature Methods 14, 975–978 (2017).
37. Negishi, H., Taniguchi, T. & Yanai, H. The Interferon (IFN) Class of Cytokines and the IFN Regulatory Factor (IRF) Transcription Factor Family. Cold Spring Harb Perspect Biol 10, a028423 (2018).
38. Luan, B. et al. CREB pathway links PGE2 signaling with macrophage polarization. Proc. Natl. Acad. Sci. U.S.A. 112, 15642–15647 (2015).
39. Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell 147, 577–589 (2011).
40. Nakada, D. et al. Oestrogen increases haematopoietic stem-cell self-renewal in females and during pregnancy. Nature 505, 555–558 (2014).
41. Gal-Oz, S. T. et al. ImmGen report: sexual dimorphism in the immune system transcriptome. Nature Communications 10, 4295–14 (2019).
42. Patterson, A. M. et al. A Single Radioprotective Dose of Prostaglandin E2 Blocks Irradiation-Induced Apoptotic Signaling and Early Cycling of Hematopoietic Stem Cells. Stem Cell Reports 15, 358–373 (2020).
43. Eyles, J. L. et al. A key role for G-CSF-induced neutrophil production and trafficking during inflammatory arthritis. Blood 112, 5193–5201 (2008).
44. Porter, R. L. et al. Prostaglandin E2 increases hematopoietic stem cell survival and accelerates hematopoietic recovery after radiation injury. STEM CELLS 31, 372–383 (2013).
45. Hoggatt, J. et al. Differential stem- and progenitor-cell trafficking by prostaglandin E2. Nature 495, 365–369 (2013).
46. Rodrigues, D. et al. Drug-induced gene expression profile changes in relation to intestinal toxicity: State-of-the-art and new approaches. Cancer Treat Rev 77, 57–66 (2019).
47. Rodriguez-Fraticelli, A. E. et al. Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nature 583, 585–589 (2020).
48. Praticò, D., Tillmann, C., Zhang, Z. B., Li, H. & FitzGerald, G. A. Acceleration of atherogenesis by COX-1-dependent prostanoid formation in low density lipoprotein receptor knockout mice. Proceedings of the National Academy of Sciences 98, 3358–3363 (2001).
49. Curry, S. H., Brown, E. A., Kuck, H. & Cassin, S. Preparation and stability of indomethacin solutions. Can J Physiol Pharmacol 60, 988–992 (1982).
50. Pietras, E. M. et al. Re-entry into quiescence protects hematopoietic stem cells from the killing effect of chronic exposure to type I interferons. J. Exp. Med. 211, 245–262 (2014).
51. Morrison, S. J., Wright, D. E. & Weissman, I. L. Cyclophosphamide/granulocyte colony-

stimulating factor induces hematopoietic stem cells to proliferate prior to mobilization. Proceedings of the National Academy of Sciences 94, 1908–1913 (1997).
52. Ema, M. et al. Primitive erythropoiesis from mesodermal precursors expressing VE-cadherin, PECAM-1, Tie2, endoglin, and CD34 in the mouse embryo. Blood 108, 4018–4024 (2006).
53. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15–5 (2018).
54. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 15, e8746 (2019).
55. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 9, 5233–12 (2019).
56. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 37, 38–44 (2018).
57. Gaublomme, J. T. et al. Nuclei multiplexing with barcoded antibodies for single-nucleus genomics. Nature Communications 10, 2907–8 (2019).
58. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127 (2007).
59. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat Biotechnol 37, 685–691 (2019).
60. Smillie, C. S. et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell 178, 714–730.e22 (2019).
61. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21 (2019).

Supplemental Figures



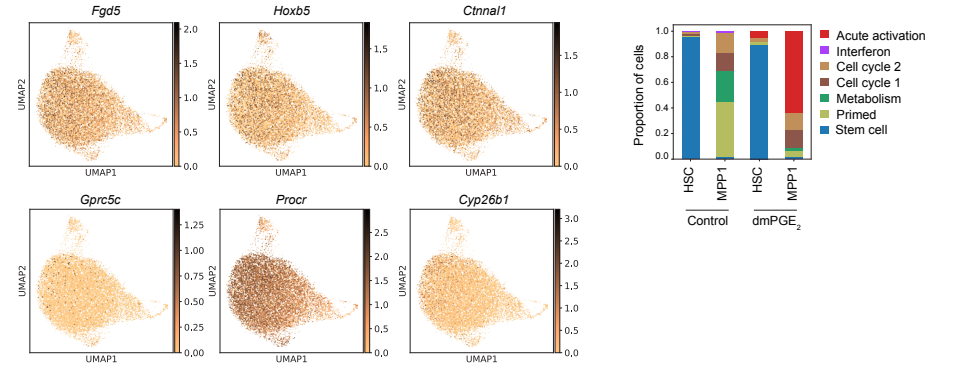
Supplemental Figure 1. Functional characterization of HSC populations confirms high regenerative capacity. (A) Overview of sorting scheme of MPPs and HSCs. Cells were lineage depleted prior to cell sorting. (B) Schematic representation of limit dilution transplant experiment. (C) Chimerism and lineage distribution per mouse at 4 months post-transplant. (D) Limit dilution analysis transplant experiment to determine HSC frequency. (E) Cell cycle FACS in MPP and HSC populations. (F) Percentage of MPPs and HSCs within the LSK compartment at baseline and after niche stimulation. (G) Proportion of MPP surface phenotypes in all experimental conditions after computational reassembly of the LSK compartment.



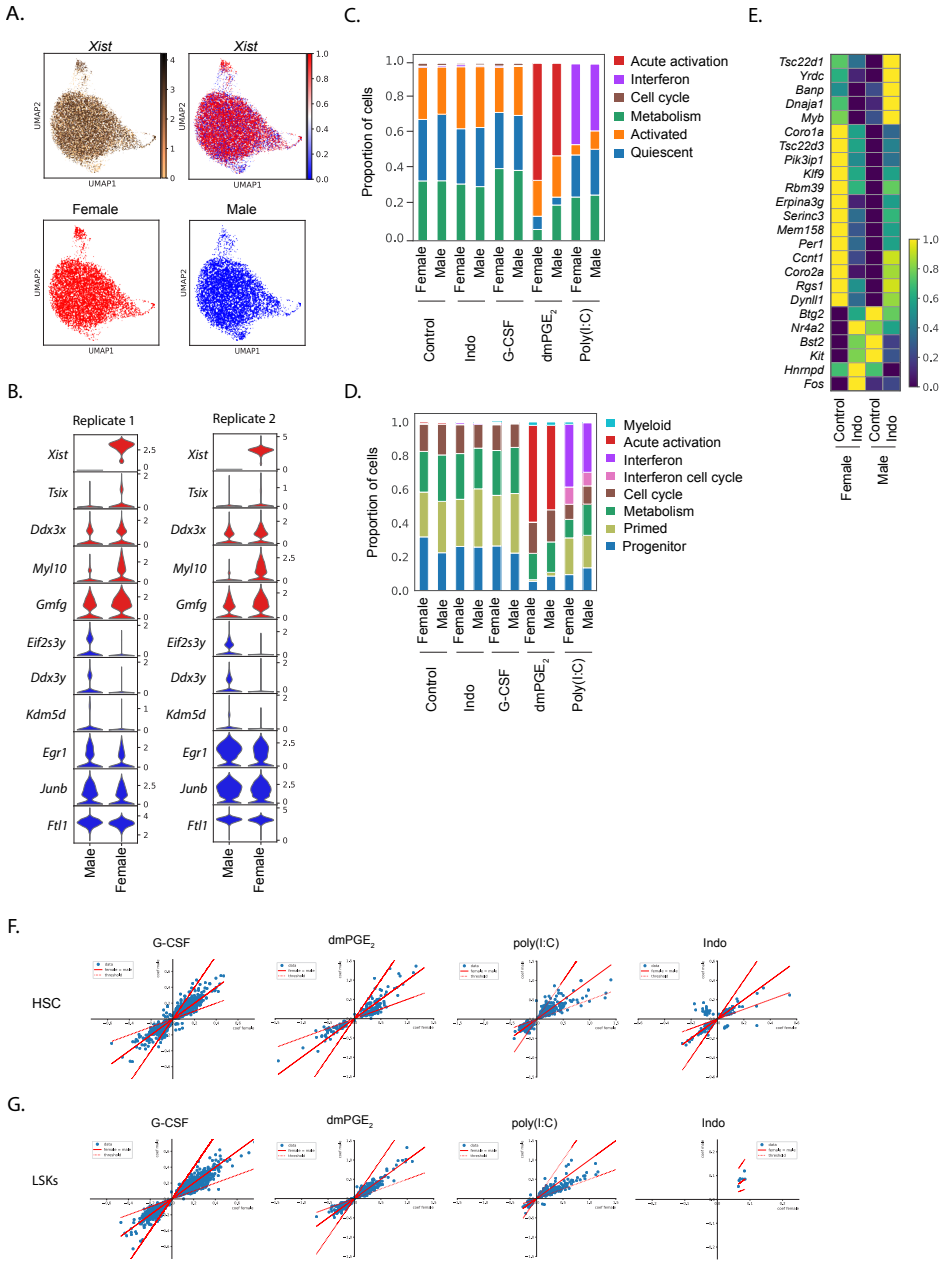
E.

Cluster	Enriched Reactome Pathways
Metabolism	RNA metabolism mRNA splicing Processing of capped intron-containing pre-mRNA HSP90 chaperon cycle for steroid hormone receptors Attenuation phase G2/M progression after G2 checkpoint Cell cycle rRNA processing
Interferon	Antiviral mechanism by IFN-stimulated genes Interferon signaling ISG15 antiviral mechanism DDX58/IFIH1-mediated induction of interferon alpha/beta Immune system OAS antiviral response Cytokine signaling in immune system Nicotinamide salvaging Interferon alpha/beta signaling
Cell cycle	Cell cycle Mitotic prometaphase Resolution of sister cromatid cohesion M phase Cell cycle checkpoints Rho GTPases activate formins Aplication of signal from the kinetochores Amplication of signal from unatteched kinetochores Rho GTPase effectors

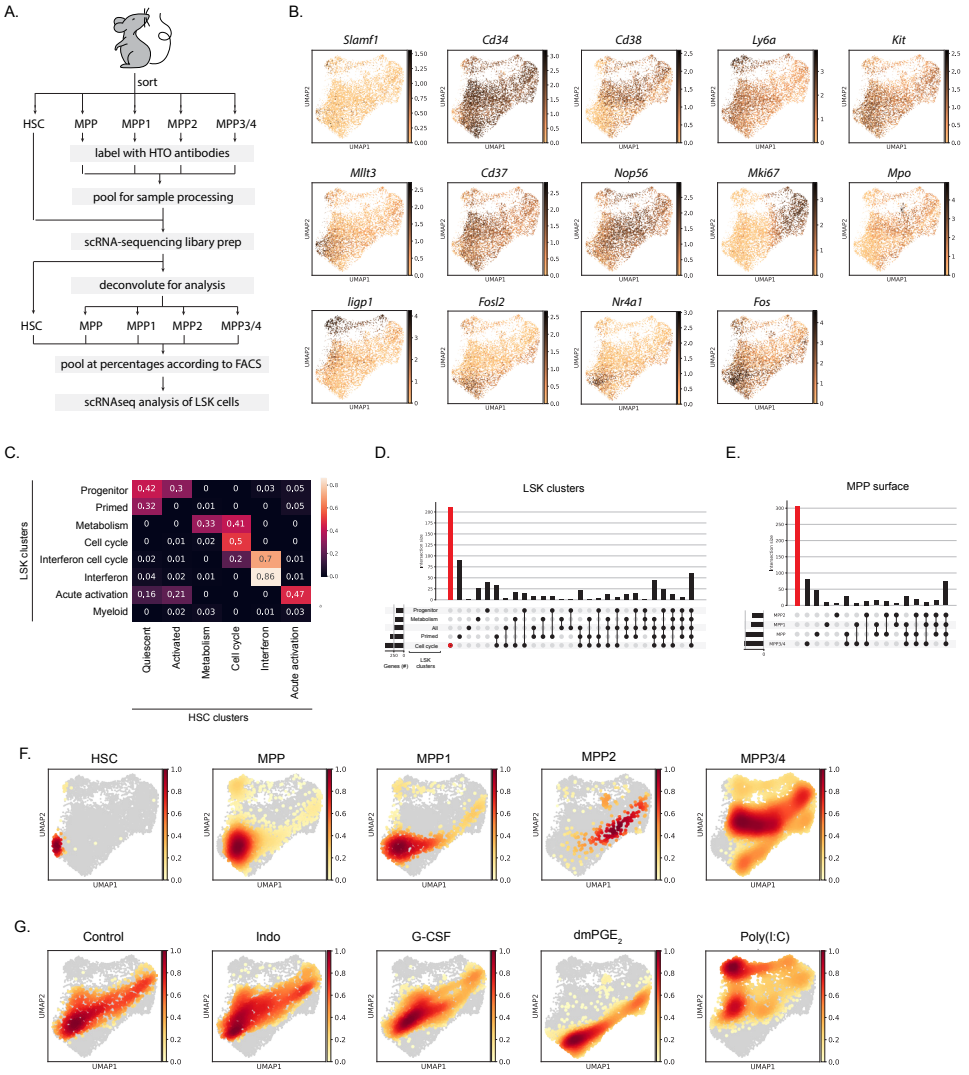
Cluster	Enriched Reactome Pathways
Quiescent	L13a-mediated translational silencing of Ceruloplasmin expression GTP hydrolysis and joining of the 60S ribosomal subunit Formation of a pool of free 40S subunits SRP-dependent cotranslational protein targeting to membrane Cap-dependent translation initiation Eukaryotic translation initiation Nonsense mediated decay Translation
Activated	Cellular senescence Cellular responses to external stimuli Cellular responses to stress Transcriptional activity of SMAD Negative regulation of MAPK pathway RAF-independent MAPK activation Signal transduction Signaling by TGF-beta
Acute activation	SRP-dependent cotranslational protein targeting to membrane Nonsense mediated decay Formation of a pool of free 40S subunits GTP hydrolysis and joining of the 60S ribosomal subunit L13a-mediated translational silencing of Ceruloplasmin expression Eukaryotic translation initiation Cap-dependent translation initiation Translation



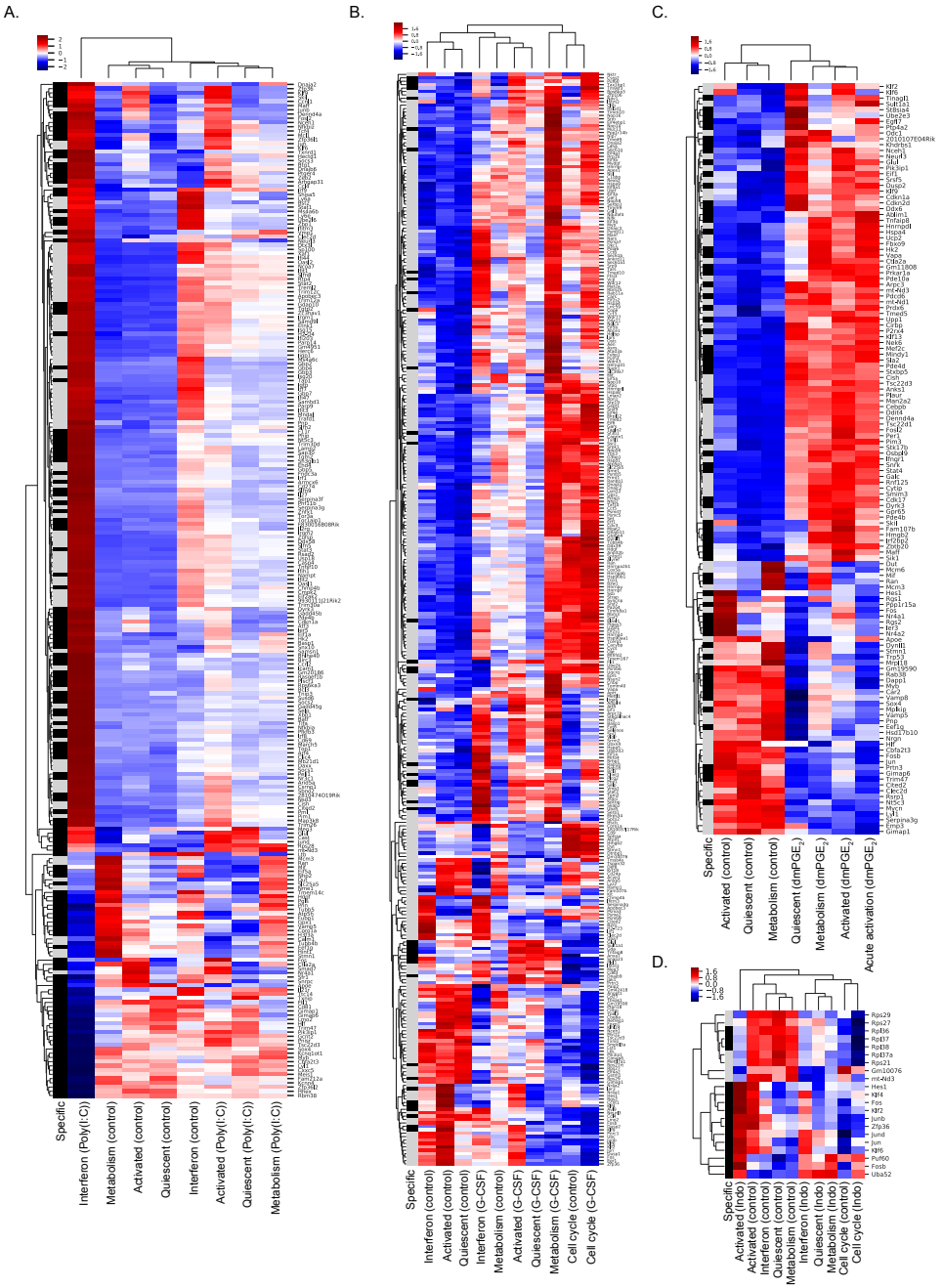
Supplemental Figure 2. Validation of scRNA-Seq clustering with independent replicates, pathway enrichment and candidate genes. (A-D) Comparison of clustering of control HSCs in two independent biological scRNA-Seq replicate experiments. UMAP plots for replicate 1 (A) and replicate 2 (B) and cell proportions in each cluster (C). (D) Pairwise comparison of the proportion of overlap of the top 100 enriched genes for each HSC cluster. (E) Reactome pathway analysis with enriched genes for each HSC cluster. (F) UMAP plots with expression of previously described HSC markers. (G) Proportion of $dmPGE_2$ and control cells within clusters split by surface phenotype for HSCs and MPP1s.



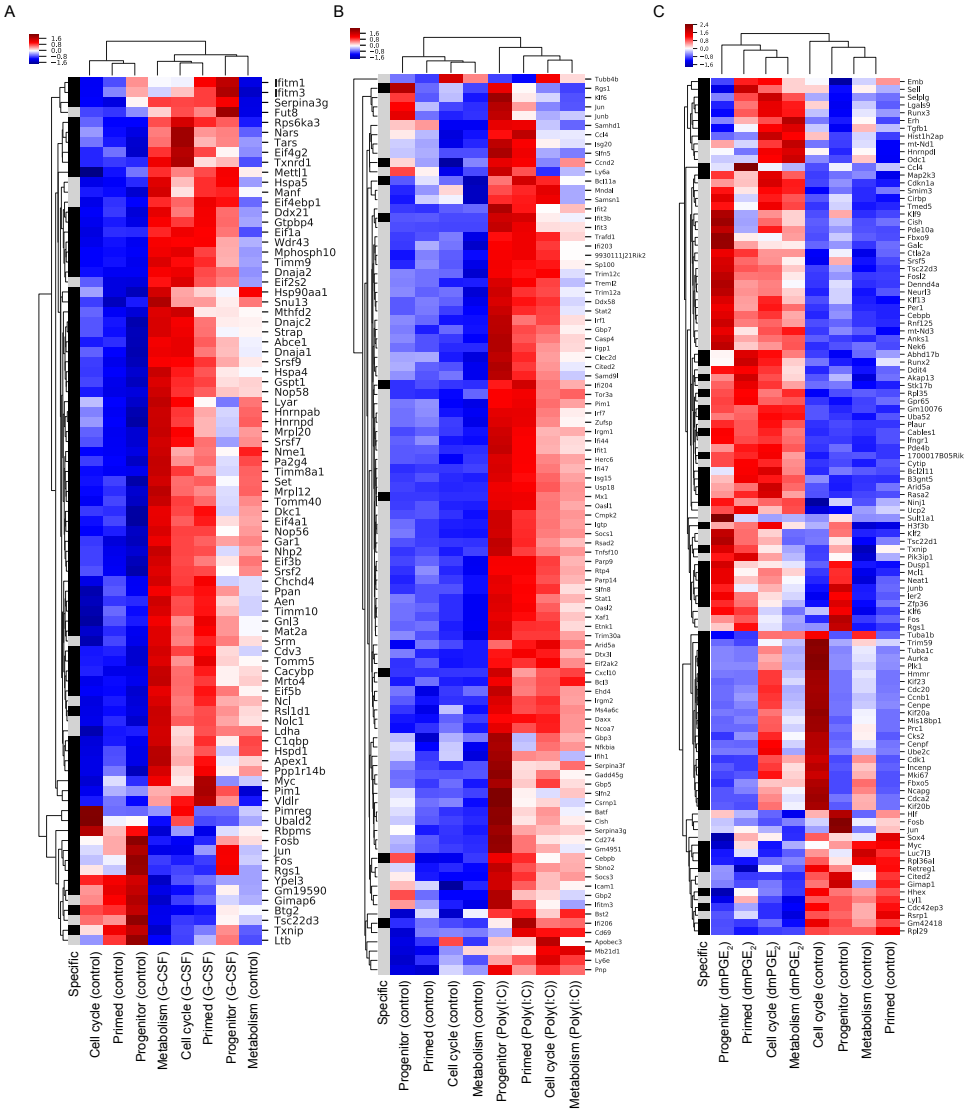
Supplemental Figure 3. Little sexual dimorphism in HSCs in steady state and upon stimulation. (A) *Xist* expression and classification of male (blue) and female (red) cells. (B) Stacked violin plots of all consistent sexual dimorphic genes for two independent biological replicates. Female expression in red, male expression in blue. (C-D) Proportions of male and female HSCs (C) and LSK cells (D) within clusters for each drug treatment (p-value of (DPA) > 0.05 for all comparisons). (E) Heatmap of average expression for genes with opposite directionality in HSCs for indomethacin and control HSCs. (F, G) Scatter plot of differential expression coefficient (converted to $\log_2$ scale) induced by stimulants in HSCs (F) and LSKs (G) between male (y-axis) and female (x-axis). Solid red line indicates equal expression coefficients. Dashed line indicates a two-fold deviation between male and female expression.



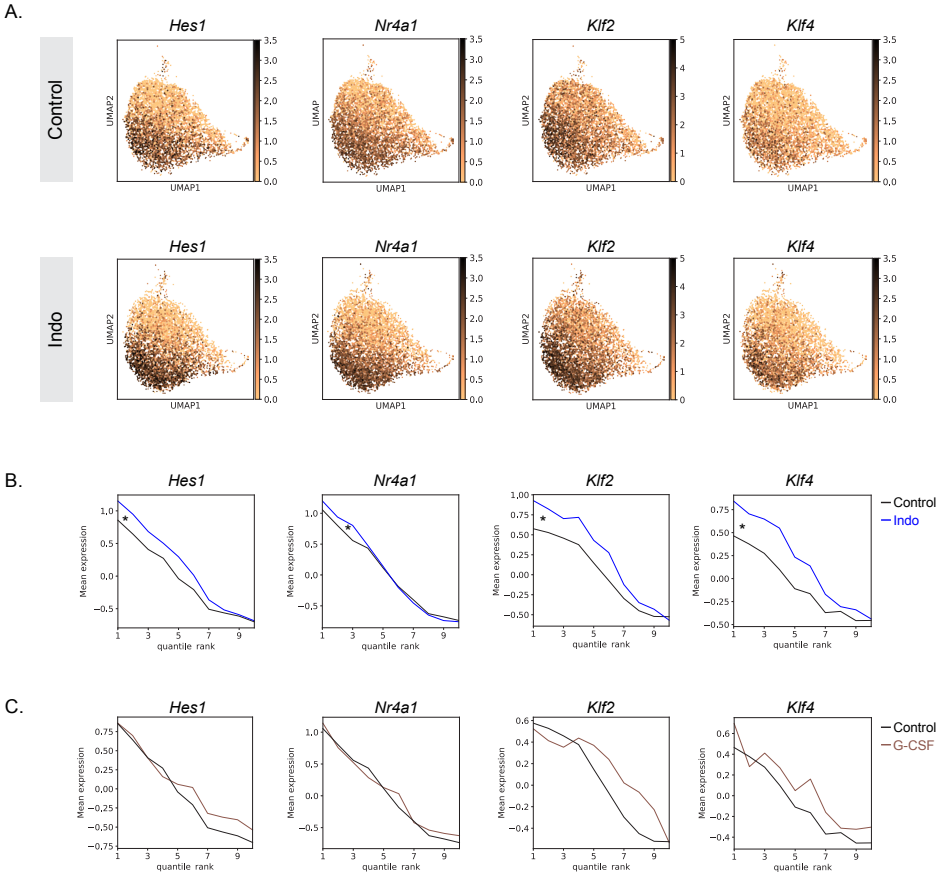
Supplemental Figure 4. Gene expression in LSKs allows for the evaluation of the specificity of niche stimulations in different cell populations. (A) Schematic of LSK pooling and CITE-Seq with surface hashtag (HTO) labelling. (B) UMAP plots of expression of selected genes in LSK cells. (C) Heatmap of proportion of overlap between the 100 top enriched genes for LSK (rows) and HSC (columns) clusters. (D, E) Upset plot visualizing overlap of differentially expressed genes between G-CSF and control for LSK clusters (D) and MPP surface phenotypes (E). (F, G) UMAP density graphs visualize distribution of cells by MPP surface phenotype (F) or niche stimulation (G).



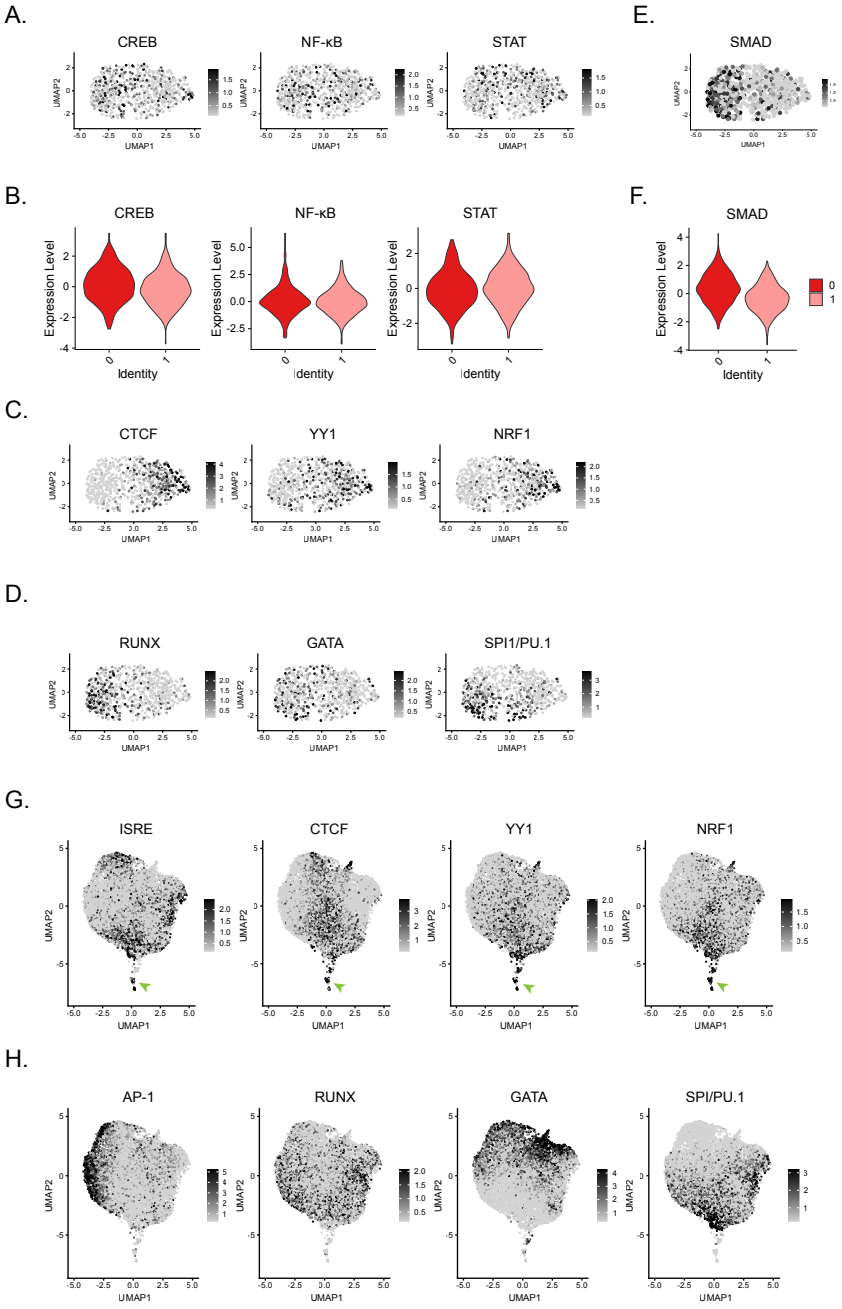
Supplemental Figure 5. Clustered heatmaps of differentially expressed genes in HSCs. (A-D) Heatmap of differentially expressed genes (DEGs) between niche stimulated and control in HSCs. Expression is averaged within a single cell cluster, scaled to z-scores and similar genes (rows) and clusters (columns) are aggregated by hierarchical clustering. Black row label indicates genes that are specific for HSCs, grey label marks genes overlapping in both HSCs and LSKs. DEGs hierarchical cluster was performed at 1.5-fold change cutoff for poly(I:C) (A), 1.2-fold change cutoff for G-CSF (B), 1.5-fold change cutoff for dmPGE$_2$ (C), and 1.2-fold change cutoff for indomethacin (D).



Supplemental Figure 6. Clustered heatmaps of differentially expressed genes in LSKs. (A-D) Heatmap of differentially expressed genes between niche stimulated and control in LSKs. Expression is averaged within a single cell cluster, scaled to z-scores and similar genes (rows) and clusters (columns) are aggregated by hierarchical clustering. Black row label indicates genes specific for HSCs, grey label marks genes overlapping in both HSCs and LSKs. DEGs hierarchical cluster was performed at 1.5-fold change cutoff for poly(I:C) (A), 1.2-fold change cutoff for G-CSF (B), 1.5-fold change cutoff for dmPGE$_2$ (C).



Supplemental Figure 7. Indomethacin affects transcriptional state of IEGs. (A) UMAP plot with expression of selected 'activated' cell state genes in control and indomethacin treated HSCs. (B, C) Average expression of the 'activated' cell state genes across cells ranked by pseudotime comparing indomethacin and control (C) or G-CSF and control (D). The change in gradient indicated by the asterisk. Cells were split into 10 bins to decrease noise



Supplemental Figure 8. Motif activities of effectors immediately downstream and secondary to niche stimulants in HSCs and LSKs. (A, B) UMAP plots (A) and violin plots (B) of transcription factor motif activities immediately downstream of prostaglandin E_2, poly(I:C) and G-CSF signaling in HSCs. (C, D) UMAP plots of motifs with enriched activities in HSC cluster 1 (C) and in HSC cluster 0 (D). (E, F) UMAP plot (E) and violin plot (F) of SMAD motif activities in HSCs. (G, H) UMAP plots of motif activities in LSK cells for motifs that were found to be enriched in HSC cluster 1 (G) and HSC cluster 0 (H). Overlapping motif activities in LSK cluster 5 are indicated with a green arrowhead in G.

Supplementary Tables

Supplementary Table 1:
Sequencing metric output from cellranger

Supplementary Table 2:
Overlap of differentially regulated genes in male and female HSCs and LSKs

Supplementary Table 3:
Differential expression result (MAST) by sex of a treatment conditions versus the control

Supplementary Table 4:
Marker gene enrichments in scRNA-Seq clusters

Supplementary Table 5:
Differential gene expression result (MAST)

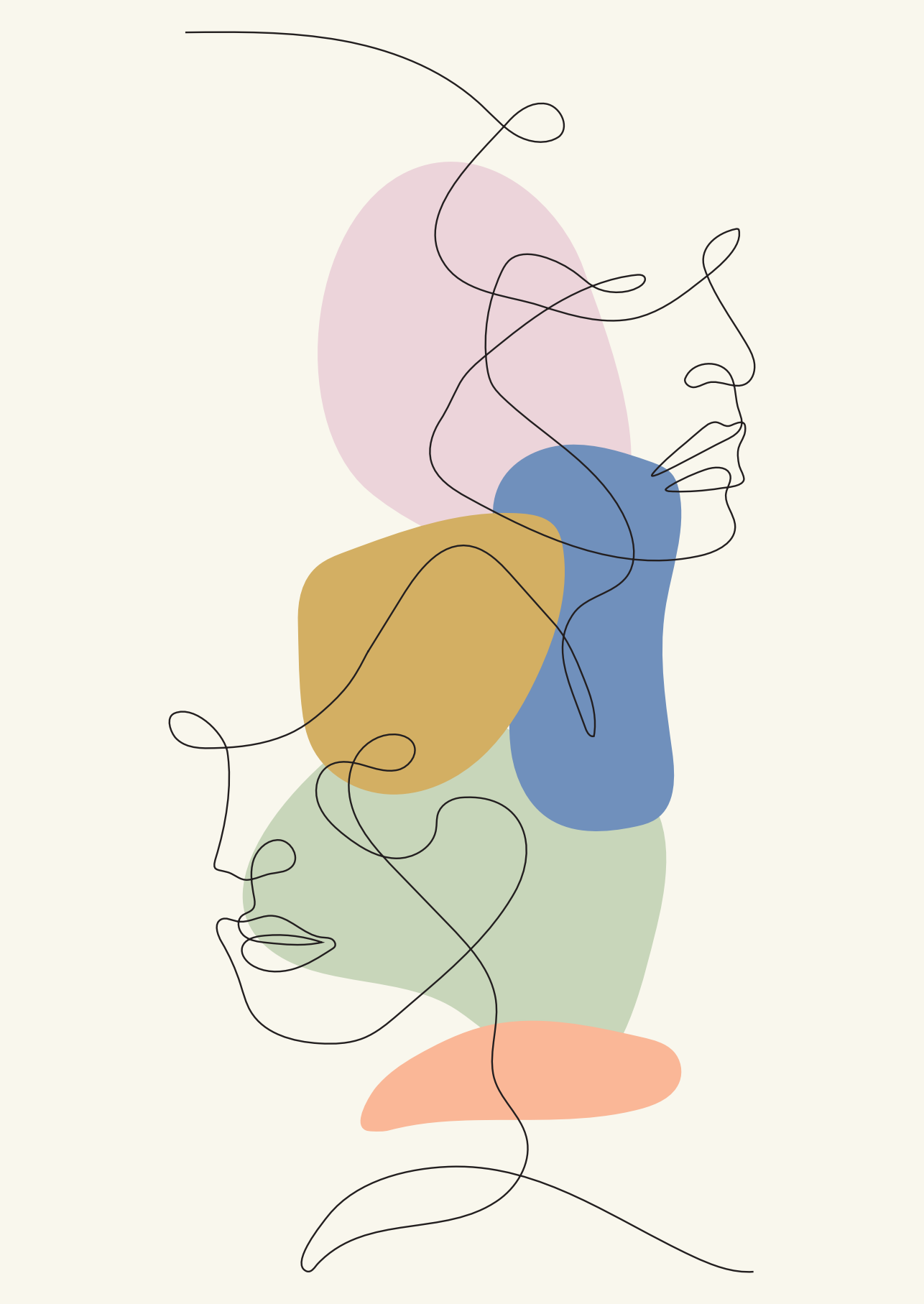
Supplementary Table 6:
Average expression per cluster of differentially regulated genes in HSCs

Supplementary Table 7:
Average expression per cluster of differentially regulated genes in LSKs

Supplementary Table 8:
ChromVar motif activity enrichment in LSK and HSC scATAC clusters

Supplementary Table 9:
List of antibodies used

Supplementary tables are available upon request



5

Retention and modification of H2A.Z-variant accessible nucleosomes at stimuli-responsive enhancers for inflammatory gene activation

Chapter 5

Retention and modification of H2A.Z-variant accessible nucleosomes at stimuli-responsive enhancers for inflammatory gene activation

Audrey Sporrij, Meera Prasad*, Brejnev Muhire*, Eva M. Fast, Margot E. Manning, Avik Choudhuri, Song Yang, Robert E. Kingston, Michael Y. Tolstorukov, Leonard I. Zon#

* These authors contributed equally

Correspondence to: leonard.zon@enders.tch.enders.edu

Abstract

Prostaglandin E2 (PGE_2) and 16,16-dimethyl-PGE_2 ($dmPGE_2$) are important regulators of hematopoietic stem and progenitor cell (HSPC) fate and offer potential to enhance stem cell therapies[1,2]. However, the mechanism of gene regulation in response to $dmPGE_2$ is poorly understood. Here, we show that $dmPGE_2$ regulates the chromatin architecture and activity of enhancer elements. We reveal the specific genomic reorganization at stimuli-responsive enhancers that permits rapid transcriptional activation. We found that $dmPGE_2$-inducible enhancers retain MNase accessible, H2A.Z-variant nucleosomes that are permissive to binding of the transcription factor CREB. CREB binding to enhancer nucleosomes is concomitant with deposition of the histone acetyltransferase p300. Subsequent H2A.Z acetylation improves chromatin accessibility at stimuli-responsive enhancers. Our findings support a model where histone variant nucleosomes retained within inducible enhancers facilitate TF binding. Acetylation of histone variant nucleosomes by TF-associated nucleosome remodelers creates the accessible nucleosome landscape required for immediate transcriptional induction.

Introduction

Hematopoietic stem cells are characterized by their unique ability to self-renew and differentiate into all mature blood cell lineages. During normal homeostasis and in conditions of stress such as injury or inflammation, HSPCs maintain an appropriate balance of the hematopoietic system. HSPCs sense and respond to a variety of extrinsic signals that regulate their quiescence, proliferation and differentiation[3]. A main mechanism of adaptation involves the activation of TFs that are downstream of signal transduction pathways to ensure appropriate transcriptional responses[4,5]. In higher eukaryotes, gene expression is regulated by the coordinated action of enhancers and promoters[6]. Stimuli-responsive TFs (STFs) tend to operate within the cis-regulatory repertoire that is established during cell fate specification and maintained by constitutive binding of lineage specific master TFs (MTFs)[7]. The access of STFs to these regulatory elements and their interaction with cofactors, such as transcriptional activators and chromatin remodeling complexes, depends largely on the local chromatin architecture[6,8].
Generally accepted features of active regulatory regions include open chromatin conformation, histone modifications and TF binding[9]. While promoters consist of a nucleosome depleted region that is established by chromatin remodelers, general TFs, and the basal transcription machinery, the typical chromatin organization and nucleosome configuration at enhancers remains unclear[10]. Moreover, the principles through which inductive signals regulate the activity of distal regulatory sites to ensure immediate gene expression changes in specific environmental contexts are largely unknown.
Prostaglandins are physiologically active lipids produced in response to mechanical, chemical or immunological stimuli. They sustain a variety of homeostatic and pathogenic functions. This includes roles in the inflammatory response. PGE_2 is one of the most abundant prostaglandins produced in the body[11]. PGE_2 and its stable derivative $dmPGE_2$ act as important regulators of vertebrate HSPC development and homeostasis[1,12]. *Ex vivo* pulse exposure of HSPCs to $dmPGE_2$ enhances engraftment and self-renewal in mice and clinical studies indicate benefits for hematopoietic stem cell transplantation (HSCT) outcomes in humans[2,13]. $dmPGE_2$ predominantly exerts its effects by binding to the PGE_2 receptor (EP) subtypes EP2 and EP4 on HSPCs[14]. Interaction with these G-coupled protein receptors enhances intracellular cAMP levels which activates signaling cascades and downstream effectors, for instance Wnt and β-Catenin[12]. Improved engraftment presumably results from upregulation of genes implicated in HSPC homing, such as *CXCR4*. As enhancement of HSPC function by external stimuli supports a strategy to improve HSCTs, understanding the mechanism of gene regulation in response to inductive signals can provide a significant clinical opportunity.
Here, we sought to address how transcriptional induction is regulated during the HSPC response to $dmPGE_2$. We exploited inducible TF binding to chromatin and identified, and then mechanistically dissected, enhancers controlling inflammatory gene expression

changes in HSPCs. We assessed the chromatin accessibility and nucleosome organization of regulatory regions responsive to dmPGE$_2$ and studied how the higher order chromatin structures changed following induction. We found that, contrary to predominant assumptions in the field, stimuli-induced enhancers retained MNase-accessible nucleosomes. These enhancer nucleosomes were enriched in non-canonical histone variant H2A.Z and remodeled but not evicted during acute stimulation. Rather than prohibiting TF binding, we observed enrichment of the dmPGE$_2$-responsive TF CREB at accessible nucleosomes within inducible enhancers. CREB binding is concomitant with deposition of the chromatin remodeler p300 that acetylates histone variant H2A.Z following dmPGE$_2$ stimulation. This may further improve nucleosomes accessibility at stimuli-responsive enhancers and allow for additional TFs and co-activator complexes to bind. We suggest the nucleosome organization at enhancers is not exclusively repressive to gene regulation but favors STF binding which enables for rapid enhancer activation and inflammatory gene induction.

Results

CREB regulates gene expression changes through binding at enhancer elements

We sought to define the molecular mechanisms that underlie the transcriptional response of HSPCs to dmPGE$_2$. We exposed human mobilized peripheral blood CD34$^+$ HSPCs to 10μM dmPGE$_2$ for 2 hours and performed extensive gene expression and chromatin profiling (Figure 1A). Using RNA-sequencing (RNA-Seq), we identified a total of 687 consistent differentially expressed genes (DEGs) after 2 hours of dmPGE$_2$ treatment, when compared to vehicle (DMSO) treated control cells (Figure 1B). The effect of dmPGE$_2$ on gene expression was overwhelmingly stimulatory. More specifically, 535 genes were ≥ 1.5-fold upregulated. This includes a significant number of genes involved in cell migration and cell cycle regulation (Supplemental Figure 1A, 1C). Among the upregulated set were genes representative of dmPGE$_2$/cAMP/PKA signaling, including *PDE4B* and *PTGS2*[15]; several chemokines and cytokines, such as *CXCL2* and *CXCL8*[16]; and genes involved in proliferation and differentiation, for instance *CCND2*, *NR4A1* and *JUNB*[17,18] (Figure 1C, Supplementary Table 1). We validated the gene expression signature observed by RNA-Seq through RT-qPCR in HSPCs for representative genes (Supplemental Figure 1D). 152 genes showed a ≤ 0.67-fold decrease in expression (Supplemental Figure 1B, Supplementary Table 1). Among the repressed genes, we found enrichment of genes that regulate cell division, such as *HOXB4* and *CCNF*[19]. To determine functional conservation of the engraftment phenotype *in vitro*, we assessed cell migration after dmPGE$_2$ treatment in a transwell migration assay. We observed greater migration of HSPCs after exposure to dmPGE$_2$ (Supplemental Figure 1E). These data showed that enhanced engraftment *in vivo* after dmPGE$_2$ stimulation is driven by

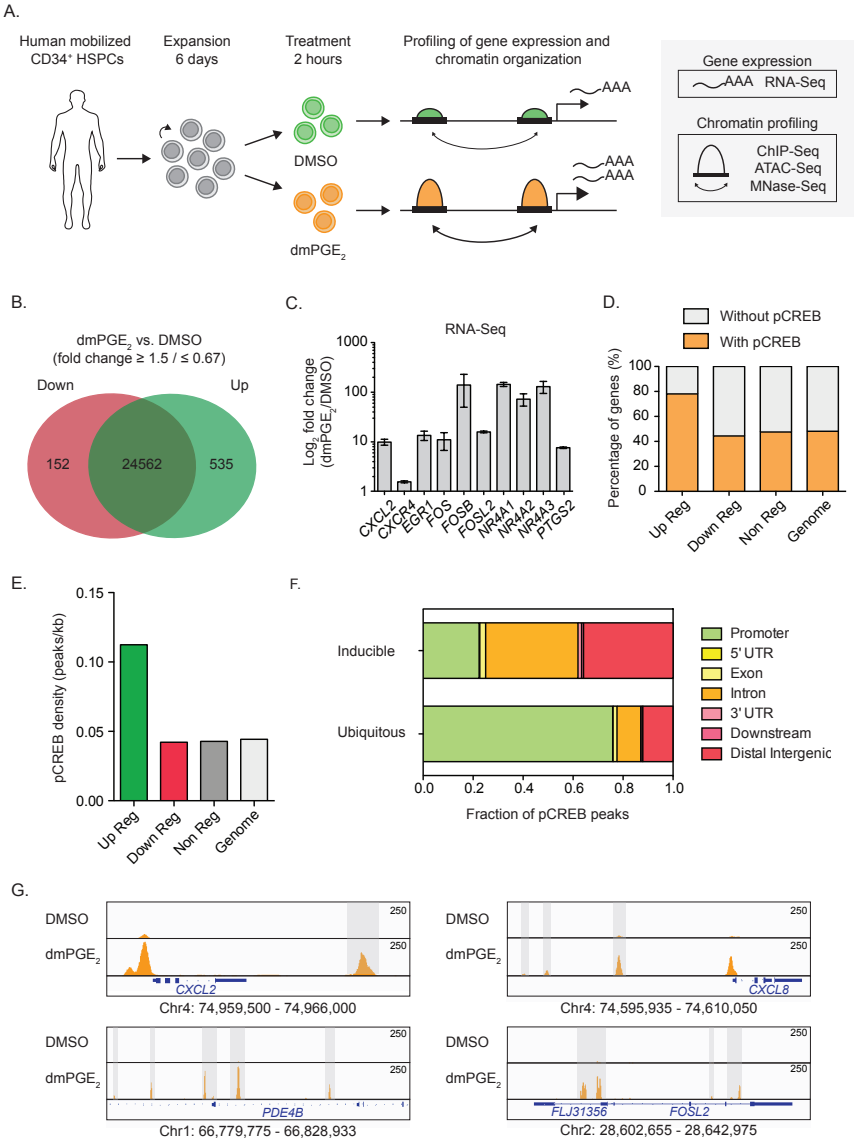


Figure 1. pCREB regulates gene expression changes through binding at enhancers. (A) Schematic representation of the experimental approach in this study. (B) Venn diagram showing the number of genes upregulated in green (535) and downregulated in red (152) in CD34⁺ HSPCs after 2 hours of dmPGE$_2$ stimulation, as determined by RNA-Seq. DEG criteria: FPKM ≥ 1 after treatment; fold change ≥ 1.5 or ≤ 0.67 (n = 3 biologically independent experiments). (C) Examples of differentially expressed genes (mean values ± SEM). (D) Percentage of genes containing at least one pCREB peak within a window from -5kb to +5kb downstream of the TTS after dmPGE$_2$ stimulation (n = 2 biologically independent ChIP-Seq experiments). (E) Correlation between pCREB binding and gene expression. pCREB density was calculated by dividing the total number of pCREB peaks associated with each gene category by the total amount of base pairs that this category occupies in the genome. Peak density was calculated by considering random distribution of pCREB sites in the genome. (F) Genomic distribution of unique pCREB peaks (inducible; present only after dmPGE$_2$ stimulation) versus ubiquitous pCREB peaks. (G) Enrichment of pCREB binding at 4 representative dmPGE$_2$ response genes: *CXCL2* (promoter, intergenic), *CXCL8* (promoter, intergenic), *PDE4B* (intronic), and *FOSL2* (promoter, intronic). Gray bars indicate intronic and intergenic pCREB peaks. Genomic location of presented window is indicated at the bottom of the panels.

transcriptional induction of migration and cell cycle genes.

Although previous studies implicated Wnt in the response to $dmPGE_2$[12], the TF CREB has also been associated with $dmPGE_2$ signaling[20]. Indeed, our transcriptomic analysis revealed that 30% (158/535) of upregulated genes are known targets of the CREB (Supplemental Figure 2A). CREB binds to cyclic-AMP response elements (CREs) near its target genes where it upon phosphorylation at serine 133 (S133) by protein kinases promotes the recruitment of co-activator proteins[21]. This increases transcription of CREB-dependent genes[22]. One of protein kinases that phosphorylates CREB at S133 is protein kinase A (PKA). We assessed S133 phosphorylation of CREB in HSPCs after $dmPGE_2$ treatment. Western blot analysis revealed increased abundance of S133 phosphorylated CREB (pCREB) while total TF protein levels remain unaltered (Supplemental Figure 2B). To correlate pCREB with gene induction, we performed chromatin immunoprecipitation sequencing (ChIP-Seq) with an antibody against pCREB. We found 31,198 binding sites in $CD34^+$ HSPCs treated for 2 hours with $dmPGE_2$ compared to 8,332 sites in DMSO (Supplemental Figure 2C, Supplementary Table 2). Correlation of pCREB occupancy to DEGs showed enrichment of pCREB at genes induced by $dmPGE_2$ (Figure 1D). In fact, 78% of the upregulated genes contained at least one pCREB peak in the proximity, that is a window from -5kb upstream of the transcription start site (TSS) to +5kb downstream of the TTS (Figure 1D). This value increased to 85% by extending the window up to -100kb from the TSS to +25kb from TTS (Supplemental Figure 2C). Besides a higher percentage of upregulated genes containing pCREB, the TF density was also higher at upregulated genes compared to noninduced genes (Figure 1E, Supplemental Figure 1D). We observed no clear correlation between the magnitude of the transcriptional response and the density of pCREB (Supplemental Figure 1F). Genes downregulated by $dmPGE_2$ showed no additional enrichment in pCREB compared to noninduced genes. This data demonstrated that $dmPGE_2$ regulates pCREB activity and genomic binding near transcriptionally induced genes.

We next assessed the genomic location of pCREB bound regions. 23,386 (75%) sites displayed unique pCREB enrichment ('inducible' pCREB regions) in $dmPGE_2$ treated HSPCs compared to control treated cells. The other 7,812 (25%) pCREB sites were present in both control and stimulated HSPCs ('ubiquitous' pCREB regions). The majority of inducible pCREB sites were located distal to the TSS of genes, with a strong representation in intronic and intergenic sequences (Figure 1F, 1G). Ubiquitous pCREB sites presented most profoundly in promoter regions. This indicated that pCREB plays an important role in upregulating $dmPGE_2$ response genes via binding at putative distal regulatory elements.

Inducible enhancers gain master and signaling transcription factor binding

To understand the epigenetic impact of $dmPGE_2$ on distal regulatory elements, we

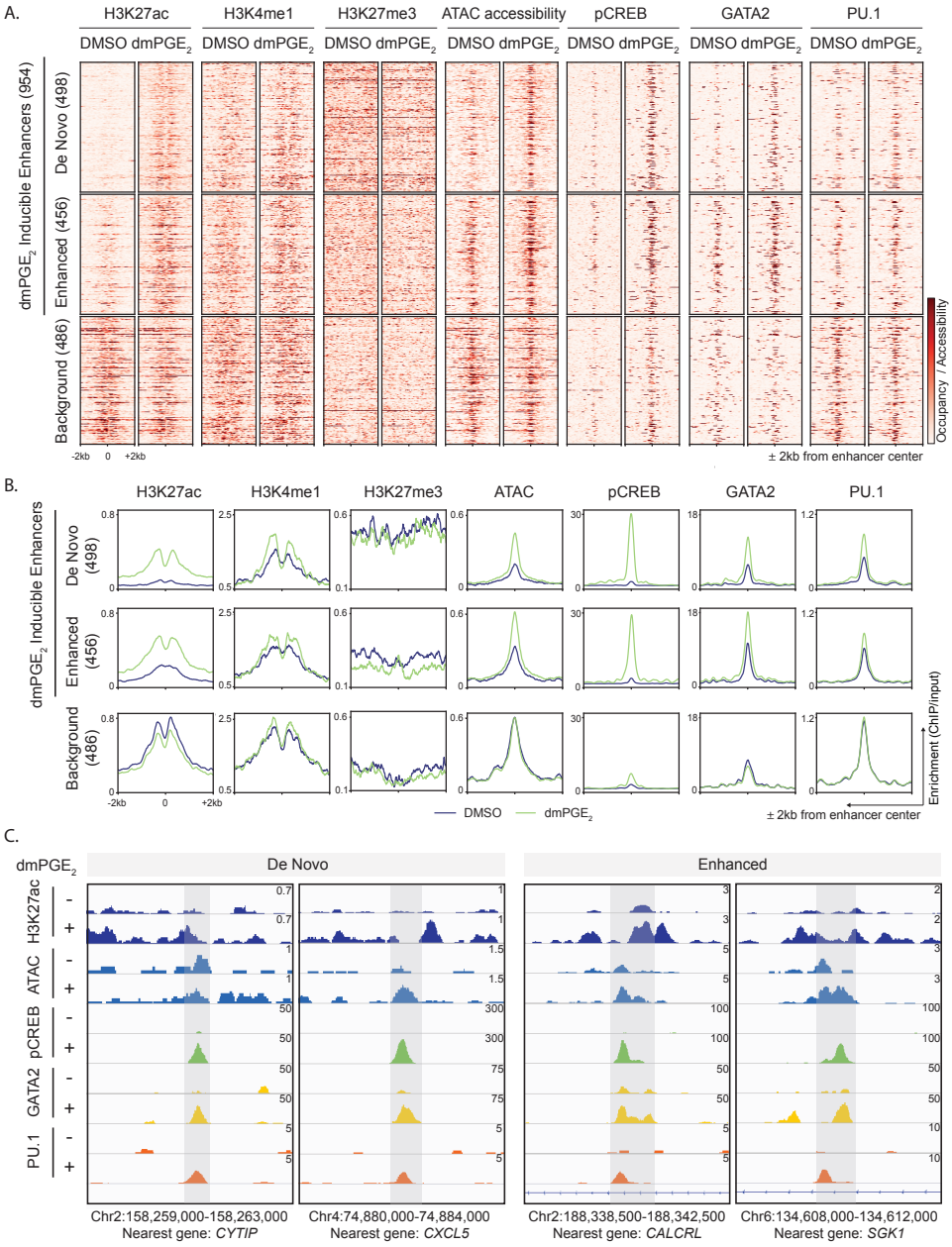


Figure 2. Stimuli-responsive enhancers gain chromatin accessibility and transcription factor binding after dmPGE$_2$ stimulation. (A, B) Heat maps (A) and average enrichment profiles (B) of histone marks, ATAC accessibility and transcription factors at enhancers before and after dmPGE$_2$ treatment. H3K27ac enriched regions identified by ChIP-Seq are classified as De Novo, Enhanced or Background enhancers according to the change in H3K27ac levels observed following dmPGE$_2$ stimulation (n = 2 biologically independent experiments). A randomly sampled, comparable number of background enhancers (486) is shown. (C) Enrichment of histone mark, ATAC accessibility and transcription factor binding in response to dmPGE$_2$ at 4 representative stimuli-response enhancers. Genomic location of presented window and nearest gene are indicated at the bottom of the panel.

determined the effects on the state of enhancers. H3K27ac distinguishes active enhancers from primed and poised enhancer elements that are marked by H3K4me1 alone or with H3K27me3, respectively[23]. We performed chromatin immunoprecipitation sequencing (ChIP-Seq) for these classifying histone modifications. We assessed H3K27ac enrichment in control and $dmPGE_2$ stimulated HSPCs and identified a total of 25,998 active distal regulatory elements (Supplementary Table 3). Enhancers that regulate stimulus responsive gene programs can be distinguished from other active enhancers by their specific increase in H3K27ac upon receipt of the stimulus[24-26]. Comparison of H3K27ac enrichment between control and $dmPGE_2$ treated HSPCs allowed us to identify putative enhancers involved in the response to $dmPGE_2$. We identified a total of 954 (3.7%) enhancers that gained significant enrichment in H3K27ac following $dmPGE_2$ treatment (Supplemental Figure 3A, 3B, and Methods). A subset of these stimuli-inducible enhancers (498/954, 52%) were only detected as active regulatory regions after $dmPGE_2$ stimulation. While these 'de novo' enhancers were depleted of H3K27ac prior to stimulation, their $H3K4me1^{high}/H3K27me3^{low}$ state indicated that de novo enhancers reside in a primed state prior to activation (Figure 2A). The other fraction of $dmPGE_2$-inducible enhancers (456/954, 48%) displayed significant enhancement in H3K27ac enrichment after $dmPGE_2$ treatment ('enhanced', Figure 2A, Supplemental Figure 3C). Since the number of non-inducible enhancers (25,044 regions, 96.3%) is much larger than the set of inducible enhancers, we randomly sampled a comparable number of non-inducible enhancers ('background', 486) to ensure that observed differences are not due to variations in size of enhancer categories. To assess epigenetic transitions at inducible enhancers, we profiled genome accessibility by assay for transposase-accessible chromatin sequencing (ATAC-Seq). We found a profound increase in DNA accessibility at inducible enhancers compared to background enhancers after $dmPGE_2$ (Figure 2A, 2B), which is a sign of active chromatin reorganization[27]. In addition, the mean ATAC-Seq signal prior to activation of de novo enhancers suggested preexisting, yet minimal, accessibility before stimulation. This revealed that $dmPGE_2$ stimulation results in rapid activation of stimuli-responsive enhancers that is concomitant with increased chromatin accessibility.

It is well known that TFs act as anchors to recruit chromatin remodelers to regulate gene expression[6]. Most STFs do not possess pioneering activity and therefore preferentially bind DNA elements located within nucleosome depleted regions. Moreover, STFs often land at regulatory elements predefined by lineage specific MTF that have pioneer functions[25,28]. Given these insights, we assessed both pCREB occupancy and binding of the HSPC MTFs GATA2 and PU.1 at enhancers (Figure 2A, 2B, Supplemental Figure 3D). We found that pCREB colocalized with GATA2 and PU.1 at stimuli-responsive enhancers. We moreover observed that inducible enhancers not only gain pCREB but also GATA2 and PU.1 enrichment after $dmPGE_2$ stimulation (Figure 2C). No enrichment in MTF occupancy was observed in background enhancers (Figure 2A, 2B, Supplemental

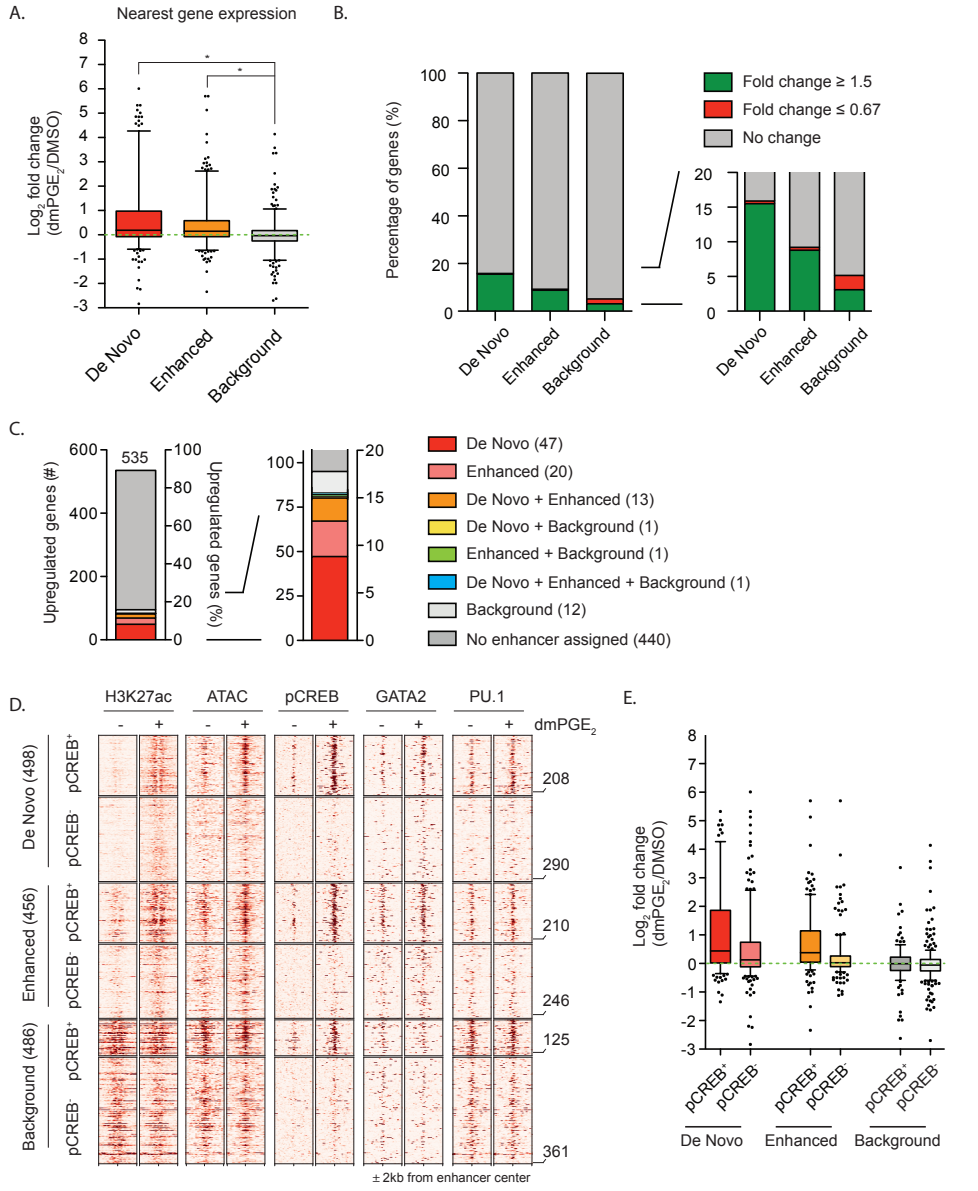


Figure 3. Stimuli-responsive enhancers mediate gene expression changes. (A) Gene expression changes of genes associated with enhancers. Enhancers were assigned to an individual nearest gene within a 15kb range. Box plots shows median, 25th and 75th percentiles, whiskers are from 5th and 95th percentiles. Dots indicate outliers. (B) Percentages of enhancer nearest genes with changes in expression ≥ 1.5-fold or ≤ 0.67-fold. (C) Upregulated genes with a fold change in expression ≥ 1.5 (535) and their associated enhancers categories. (D) Heat maps of histone marks, ATAC accessibility and transcription factors at enhancers subset by presence or absence of pCREB after $dmPGE_2$ stimulation. Numbers of enhancers within each subset is indicated on the right of the heatmap. (E) Gene expression changes of genes associated with enhancers subset by pCREB presence. Enhancers were assigned to an individual nearest gene within a 15kb range. Box plots shows median, 25th and 75th percentiles, whiskers are from 10th and 90th percentiles. Dots indicate outliers. For all analysis a randomly sampled set of background enhancers (486) was used.

Figure 3E). These data showed that $dmPGE_2$-inducible enhancers gain both STF and MTF deposition following stimulation.

The importance of distal regulatory regions for activation of $dmPGE_2$ target genes is illustrated by the changes in expression of genes regulated by the enhancers (Figure 3A, Supplemental Figure 4A). We tentatively assigned enhancers to individual nearest genes. To reasonably limit arbitrariness in gene assignment, we only considered genes with a mapped TSS within 15kb of an enhancer. The genes nearest to stimuli-inducible enhancers showed a greater transcriptional response than genes annotated to background enhancers (Figure 3A, Supplemental Figure 4A). This indicates that differential enhancer activity is directly reflected on gene expression levels. Importantly, genes associated with inducible enhancers belonged to several pathways, including cell signaling and blood cell migration such as *SGK1*, *CALCRL*, *CXCL2*, *CXCL5* and *ITGA4* (Figure 2C, Supplementary Table 3). We found a clear correlation between the number of genes showing differential expression and the changes in enhancer activity. 16% of genes nearest to de novo enhancers display ≥ 1.5-fold induction after $dmPGE_2$ treatment, compared to 9% of enhanced and 3% of background enhancers (Figure 3B, Supplemental Figure 4B). In total, 83 (16%) of the 535 upregulated genes were regulated by at least one stimuli-responsive enhancer (Figure 3C, Supplemental Figure 4C). This demonstrated that $dmPGE_2$ mediated activation of stimuli-responsive enhancers correlates with modulation of gene expression.

To gain a better understanding of the role of pCREB at inducible enhancers, we further segmented enhancers based on the presence or absence of pCREB after $dmPGE_2$ treatment. We found that 208/498 (42%) de novo, 210/456 (46%) enhanced and 125/486 (26%) background enhancers to contain at least 1 pCREB ChIP-Seq peak after $dmPGE_2$ treatment. Chromatin accessibility and MTF binding increased more profoundly, but not exclusively at $pCREB^+$ stimuli-inducible enhancers (Figure 3D). The results support a prominent, but not restricted, role for pCREB in regulating gene expression through binding at inducible enhancers. Additionally, $pCREB^+$ stimuli-responsive enhancer showed greater transcriptional changes than those without pCREB (Figure 3E). Together, our data suggested that gene induction is a downstream result of increased enhancer chromatin accessibility and the cooperative binding MTFs and STFs at these distal regulatory regions.

Inducible enhancers retain accessible nucleosomes after inflammatory stimulation

Many studies showed that TFs preferentially bind to sites of 'open' chromatin[29]. However, recent work indicated that accessible chromatin may not necessarily represent nucleosome depleted regions[30,31]. To test whether greater chromatin accessibility at stimuli-responsive enhancers is due to nucleosome depletion, we performed micrococcal nuclease sequencing (MNase-Seq). This allowed us to map nucleosome positions and

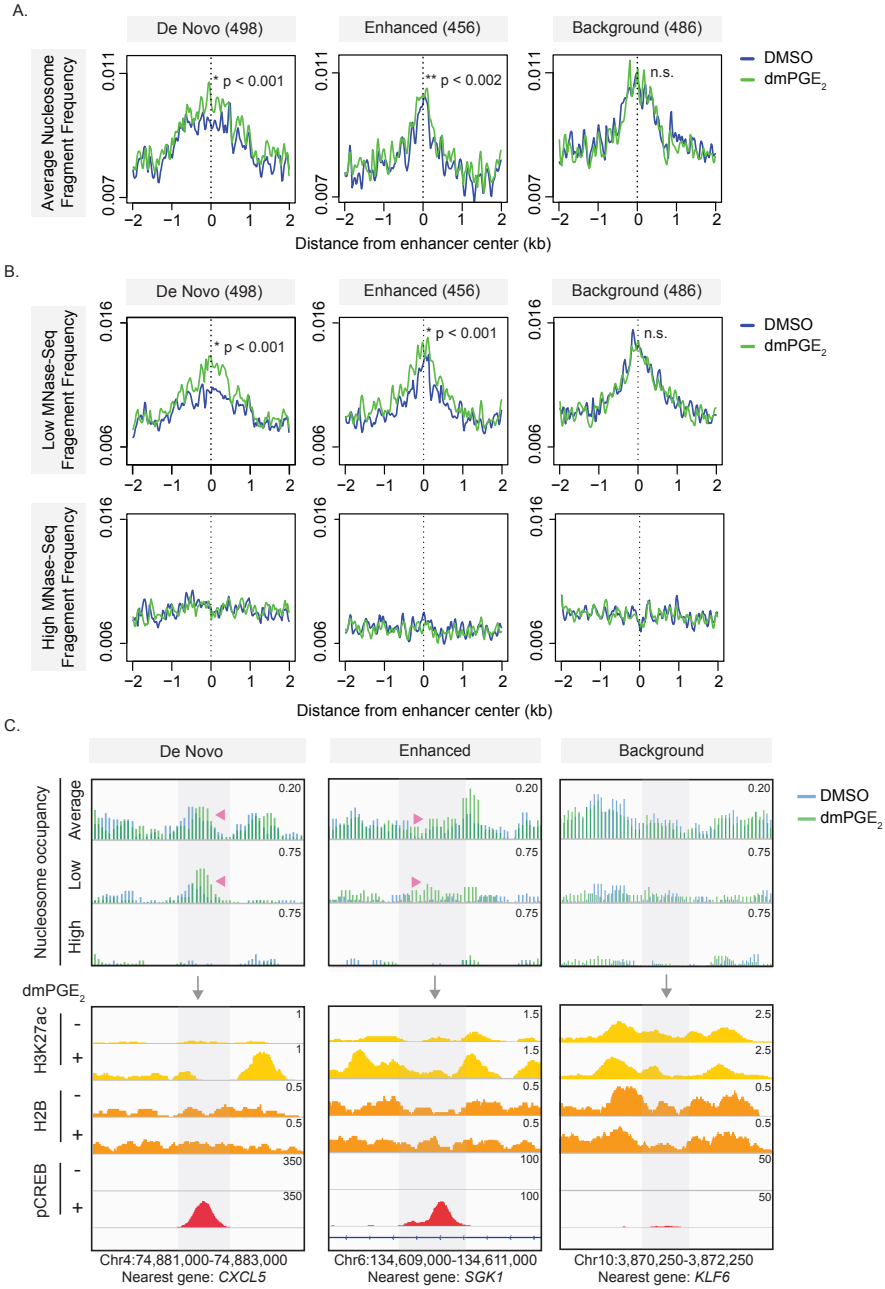


Figure 4. dmPGE$_2$-responsive enhancers retain accessible nucleosomes after stimulation. (A) Average nucleosome occupancy profiles at stimuli-responsive and background enhancers from 4 MNase titration points. (n = 3 biologically independent MNase-Seq experiments). (B) Nucleosome profiles of low- and high MNase-Seq at stimuli-responsive and background enhancers. P-values by Wilcoxon rank-sum test. (C) Nucleosome fragment frequency and enrichment of H3K27ac, H2B, and pCREB at 3 representative stimuli-responsive and background enhancers. Genomic location of presented window and nearest gene are indicated at the bottom of the panel. Genomic location of presented window and enhancer nearest gene are indicated at the bottom of the panel. For all analysis a randomly sampled set of background enhancers (486) was used.

nucleosome occupancy changes after dmPGE$_2$ stimulation. Because nucleosomes have varying sensitivities to enzymatic digestion, MNase titrations are performed to obtain a comprehensive map of the nucleosome landscape within a genome[30-32]. While comparing occupancy profiles between individual titration points provides insights on nucleosomal accessibility, the combinatorial analysis of all titration points within a given condition generates a complete view of the nucleosome organization[30] (Supplemental Figure 5A). We prepared native nuclei from HSPCs treated with either DMSO and dmPGE$_2$ and exposed them to increasing units of MNase. We selected 4 digestion points that generated increasingly larger fractions of mononucleosomal-size DNA. Mononucleosomal fractions comprised around 10%, 25%, 50%, and 75% of the input chromatin, respectively (Supplemental Figure 5B). Individual MNase titration profiles within a condition, as well as pooled average nucleosome occupancy profiles, revealed a TSS pattern similar to those reported previously[30] (Supplemental Figure 5C, 5D). We observed lower average nucleosome occupancy at TSS-proximal regions of dmPGE$_2$-responsive genes after treatment (Supplemental Figure 5E). This inverse correlation aligns with former studies which described that high transcriptional levels are concomitant with nucleosome eviction at the TSS, as elongation by RNA Polymerase II is thought to disrupt the nucleosomal organization[33,34].

When assessing the nucleosome organization at inducible enhancers, we found that these regions presented in high nucleosome occupancy states when compared to active promoters (Figure 4A, Supplemental Figure 5E). Although high occupancy regions are traditionally thought of as 'closed', novel work indicated that accessible regulatory regions can retain nucleosomes[31,32]. We determined whether nucleosomes were repositioned or evicted at dmPGE$_2$-responsive enhancers following stimulation. No decrease in average nucleosome occupancy was observed at stimuli-responsive enhancers. Rather, inducible enhancers demonstrated higher nucleosome occupancy after dmPGE$_2$ treatment (Figure 4A). This was not observed at background enhancers, indicating specificity of the phenomena to dmPGE$_2$-inducible enhancers. Our data suggests that stimuli-responsive enhancers retained nucleosomes upon activation, rather than remodeled to a nucleosome free organization via nucleosome eviction.

Nucleosomes profiles from individual MNase titration points can be leveraged to determine how MNase sensitivity changes in specific regions (Supplemental Figure 5A). We found greater low MNase-Seq signal at inducible enhancers after dmPGE$_2$ stimulation (Figure 4B, upper panel) whereas high MNase digestion degraded the nucleosomal DNA fragments (Figure 4B, bottom panel). No changes in low MNase sensitivity were observed in background enhancers. Greater low MNase sensitivity after dmPGE$_2$ treatment indicates a higher presence of MNase accessible nucleosomes. This shows that nucleosomes retained within inducible enhancers gained low MNase sensitivity of upon stimulation.

To ensure that the low MNase-Seq signal represents nucleosomes, we performed ChIP-

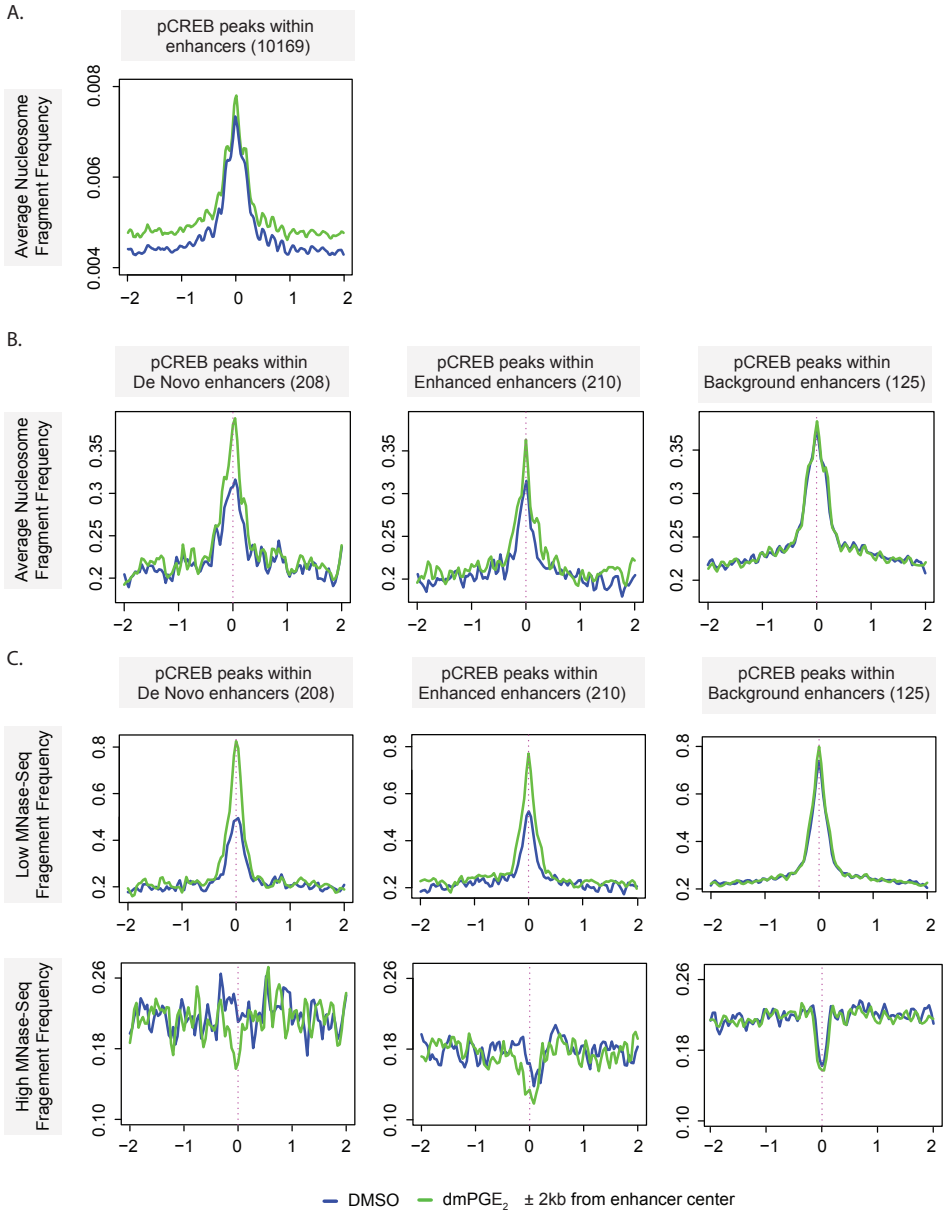


Figure 5. pCREB engages accessible nucleosomes in stimuli-responsive enhancers. (A) Average nucleosome frequency at pCREB peaks located within enhancers before and after dmPGE$_2$ stimulation (n = 3 biologically independent MNase-Seq experiments). (B) Averaged nucleosome occupancy profiles at pCREB-bound sites within stimuli-responsive and background enhancers from 4 MNase titration points (n = 3 biologically independent experiments). (C) Nucleosome profiles of low- and high MNase-Seq at pCREB-bound sites within stimuli-responsive and background enhancers. For all analysis a randomly sampled set of background enhancers (486) was used.

Seq for the core histones H2B and H4 (Figure 4C, Supplemental Figure 6A). Inducible enhancers were not depleted of core histones prior to or after $dmPGE_2$ stimulation. This is in line with our MNase-Seq data and suggested that nucleosomes are indeed present, and retained at, $dmPGE_2$-responsive enhancers. Thus, in contrast to our expectations that increased chromatin accessibility at inducible enhancers (Figure 2B) results from nucleosome displacement or eviction, our MNase-Seq and ChIP-Seq data revealed that stimuli-responsive enhancers retained accessible nucleosomes upon activation.

Intrigued by the observation of accessible nucleosomes at enhancers, we next evaluated the relationships between nucleosomes and binding of pCREB at inducible enhancers. 10,169/23,386 (43%) of pCREB binding sites uniquely present after $dmPGE_2$ treatment are located within the 25,998 active enhancers identified in HSPCs. We observed enrichment of phased nucleosomes at the summit of $dmPGE_2$-unique pCREB peaks located within enhancers, both before and after $dmPGE_2$ treatment (Figure 5A). To exclude the possibility that the MNase-Seq fragments observed at pCREB sites within enhancers represent non-histone proteins protecting from MNase digestion, we assessed fragment size distribution (Supplemental Figure 6B). We found a fragment periodicity that is characteristic for MNase-digested nucleosomes[35,36]. The majority of the MNase-Seq fragments that coincide with pCREB binding was 148 base pairs (bp) in length. This corresponds to precisely trimmed nucleosome core particles. Subnucleosomal peaks showed a clear ~10bp periodicity that reflects the accessibility of DNA as it is wound along the surface of the histone octamer[37]. This analysis suggested that MNase-Seq fragments mapping to pCREB sites within enhancers represent nucleosomes. Looking specifically at $pCREB^+$ stimuli-responsive enhancers, we observed that the effects of $dmPGE_2$ on nucleosome occupancy and low MNase sensitivity described earlier are further amplified at $pCREB^+$ regions (Figure 5B, 5C). Assessment of individual loci confirmed pCREB binding at stimuli-responsive enhancers that show greater low MNase signal and increased nucleosome occupancies after $dmPGE_2$ treatment (Figure 4C). Together, the data demonstrated that accessible nucleosomes are retained at $dmPGE_2$-responsive enhancers and that these nucleosomes are not prohibitive of pCREB binding.

Modification of H2A.Z-variant accessible nucleosomes at stimuli-responsive enhancers

We hypothesized that retention of nucleosomes may exert important roles in pCREB binding. Nucleosome-driven TF binding has been observed for several other stress-responsive TFs[38]. To understand the conformational changes that underlie greater low MNase sensitivity of enhancer nucleosomes after $dmPGE_2$ stimulation, we investigated other mechanisms that influence accessibility to nucleosomal DNA. Weakening of internucleosomal interactions by covalent modifications of histones as well as the introduction of histone variants increases DNA accessibility[39]. The histone variants H2A.Z and H3.3 can be incorporated in replication independent manners and are

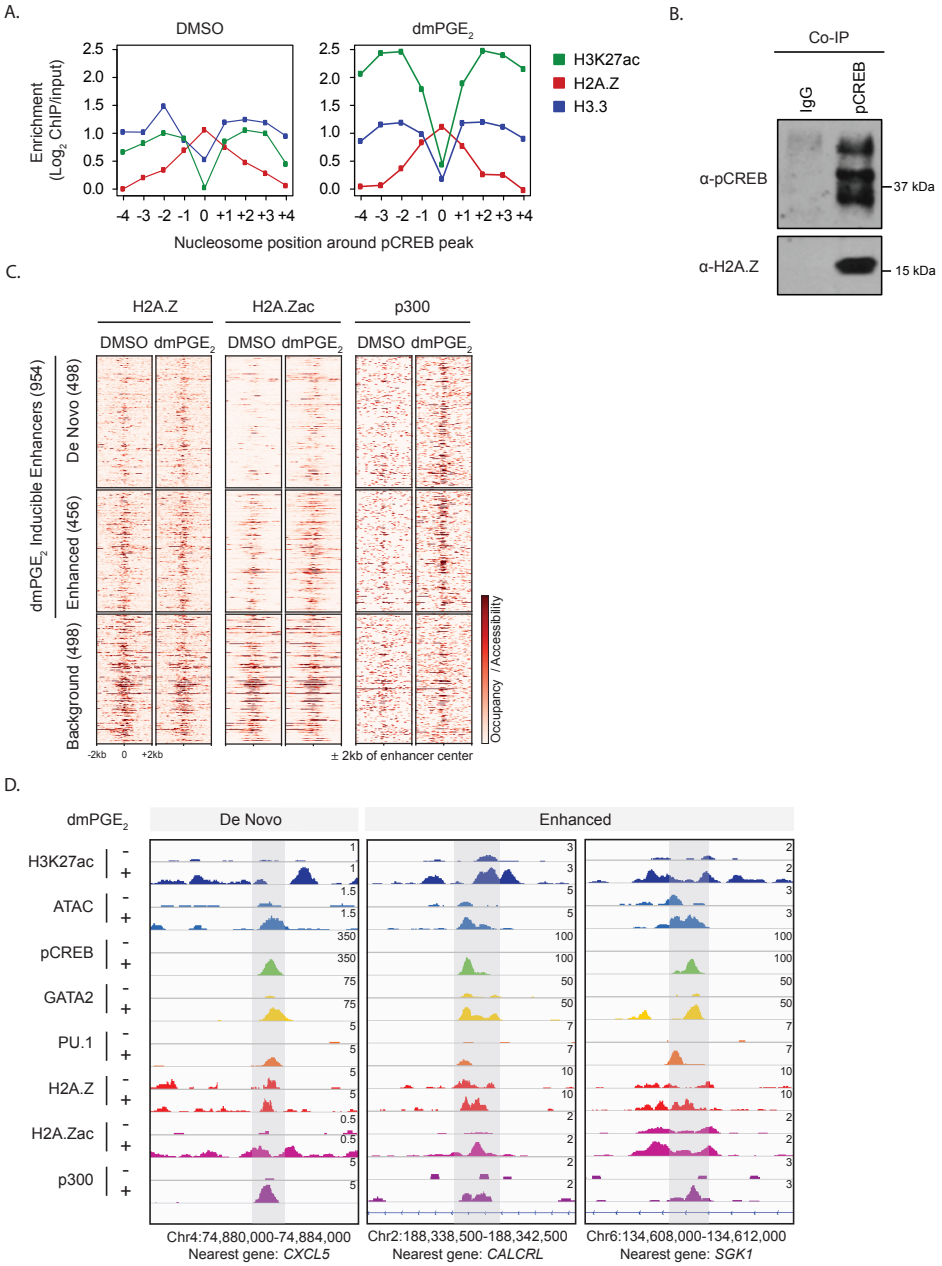


Figure 6. Modification of H2A.Z-variant accessible nucleosomes at stimuli-responsive enhancers. (A) Histone variant Enrichment at nucleosomes positions surrounding pCREB peaks within enhancers before and after dmPGE$_2$ stimulation. Position 0 indicates the nucleosome overlapping with pCREB peak centers. (B) Co-IP showing that pCREB associated with H2A.Z in U937 myeloid leukemia cells (n = 3 biologically independent experiments). (C) Heat maps of H2A.Z, H2A.Zac and p300 binding at enhancers before and after dmPGE$_2$ treatment. De Novo, Enhanced or Background enhancers were subset based on the presence or absence of pCREB after dmPGE$_2$. (D) Enrichment of chromatin binding factors and transcription factors in response to dmPGE$_2$ at 3 representative stimuli-response enhancers. Genomic location of presented window and enhancer nearest gene are indicated at the bottom of the panel. For all analysis a randomly sampled set of background enhancers (486) was used.

associated with enhancers[40-42]. Once incorporated, these non-canonical histones make for destabilized 'fragile' nucleosomes[43]. We assessed histone variant abundance at regulatory regions in the presence and absence of dmPGE$_2$ stimulation through ChIP-Seq analysis for H2A.Z and H3.3. We observed a strong correlation between the histone variant deposition and the H3K2ac levels that defined enhancer activity (Supplemental Figure 7B, 7C). The histone variants were found to occupy slightly different sites within enhancers. H2A.Z localized more central to enhancers the region where TFs bind, whereas H3.3-variant nucleosomes followed a more dispersed pattern and localized to the flanks of enhancers in a profile similar to H3K27ac (Supplemental Figure 7B, 7C). We noted minimal changes in H2A.Z enrichment following dmPGE$_2$ stimulation. We did observe incorporation of the histone variant H3.3 in the nucleosomes that enhancers (Supplemental Figure 7B, 7C). When specifically assessing histone variants at pCREB$^+$ stimuli-responsive enhancers, we found that pCREB binding directly overlaps with H2A.Z-variant nucleosomes (Figure 6A, Supplemental Figure 7A). To confirm interaction of pCREB with H2A.Z-variant nucleosomes, we performed complex immunoprecipitation (Co-IP) for pCREB and found association between the TF and H2A.Z histones *in vitro* (Figure 6B). These results indicated that DNA binding of pCREB is not prohibited by H2A.Z-variant nucleosomes.

H2A.Z is associated with both gene repression and activation[44]. This dual function is attributed to post-translational regulation of histone variant. Different H2A.Z histone tail modifications recruit distinct interactors that mediate varying transcriptional outputs. H2A.Z acetylation is associated with active transcription and dynamically regulated in response to environmental signals[45,46]. Acetylation of H2A.Z occurs at active regulatory regions where it promotes nucleosomes destabilization and an open chromatin conformation. The histone acetyltransferases (HATs) p300 and Tip60 that acetylate H2A.Z are known interactors of pCREB[47,48]. Whereas Tip60 alone is not sufficient to acetylate H2A.Z, p300 can rapidly and effectively acetylate H2A.Z-containing nucleosomes on its own[49]. Hence, we assessed if acetylation of H2A.Z-variant nucleosomes at stimuli-responsive enhancers underlies increased MNase sensitivity and nucleosome accessibility after dmPGE$_2$ treatment. We found specific H2A.Zac enrichment at stimuli-inducible enhancers whereas this was not observed in background enhancers (Figure 6C, 6D Supplemental Figure 7C). We furthermore noted that dmPGE$_2$ increased the abundance of p300 at enhancers (Figure 6C, 6D, Supplemental Figure 7C). Binding patterns of p300 at enhancers were highly similar to pCREB after dmPGE$_2$ treatment (Figure 2B, Supplemental Figure 7C). Together, our data revealed that pCREB binds H2A.Z-variant nucleosomes within enhancers following dmPGE$_2$ stimulation. pCREB binding at stimuli-responsive enhancers is accompanied by nucleosome remodeling through H2A.Z acetylation, likely mediated via interaction of CREB with p300. H2A.Z acetylation at enhancers may underlie increased enhancer accessibility, allowing additional chromatin factors to engage and drive acute gene expression changes.

Discussion

Prostaglandin E2 is an important regulator of HSPC homeostasis. The distinct molecular mechanisms through which PGE_2 and its stable derivative $dmPGE_2$ affect HSC function are critical to understand yet remain elusive. Here, we find that the TF CREB is a key player in the acute transcriptional response to $dmPGE_2$ by binding to, and activating, distal regulatory elements. Specifically, we find that pCREB binds to, and acetylates, H2A.Z-variant nucleosomes that are retained within stimuli-induced, active enhancers. H2A.Z acetylation of enhancer nucleosomes increases local chromatin accessibility, which in turn may help to recruit and/or stabilize other HSPC specific TFs such as GATA2 to stimulate gene transcription.

CREB is a ubiquitously expressed nuclear, basic leucine zipper (bZIP) TF that regulates over 5,000 genes in the mammalian genome. This includes genes controlling proliferation, differentiation and cell survival [50]. The TF is activated by a wide variety of environmental stimuli and an important regulator of stress responses. Most studies focused on promoter-proximal effects of CREB binding. We showed that pCREB localizes to promoters of ubiquitously expressed genes, while inducible gene expression is characterized by binding of pCREB at TSS-distal enhancers regions. Regulation of transcriptional responses through binding of CREB at enhancers has also been observed in pancreatic beta cells[51]. Our data suggests that CREB employs distinct mechanism to regulate ubiquitous versus inducible gene expression.

Genome wide assessment of the epigenetic landscape revealed that $dmPGE_2$ works within the predetermined enhancer repertoire of HSPCs. $dmPGE_2$ stimulation activates a set of pre-existing $H3K4me1^+$ enhancers through chromatin reorganization. $dmPGE_2$-responsive enhancers rapidly gain accessibility and TF binding. Our work complements the studies that showed that STFs localize to binding sites adjacent to master regulators[28]. Although we did not identify the surfacing of latent enhancers, *i.e.* genomic regulatory elements devoid of TFs and enhancer marks in unstimulated cells[25], we do not exclude that a 2-hour pulse of $dmPGE_2$ may be too short to allow for partial reprogramming of the available cis-regulatory landscape. Latent enhancer activation may also be more associated with differentiated cells rather than stem and progenitor populations[25].

We found stimuli-driven enrichment of the HSPC specific MTFs GATA2 and PU.1 at inducible enhancers, especially those that gained pCREB binding. Cofactor driven binding is common among non-pioneer TFs. It was only recently implied that interaction between TFs can enhance pioneer factor binding at previously sampled target sites[52-54]. Many studies have described the vital role of GATA2 in establishing the regulatory landscape for STFs in HSPCs[28,55]. GATA2 facilitates enhancer-promoter loop formation[56]. Yet few have proposed signaling factors to play a role in GATA2 binding and recruitment to chromatin. We suggest that GATA2 occupancy may be directed and stabilized through cooperativity with STFs at stimuli-responsive enhancers. Our work supports a

model where occupancy of a pioneer factor at specific subsets of enhancers is partially determined by engagement with signaling specific cofactors.

An open chromatin structure surrounding TF binding sites constitutes a prerequisite for transcriptional regulation. Our results show that inducible enhancers rapidly gain accessibility and affinity for STFs and MTFs. While open chromatin regions are presumed nucleosome depleted regions, we find that dmPGE$_2$-responsive enhancers exhibit a high nucleosome occupancy both prior to and after stimulation of HSPCs. Moreover, retained enhancer nucleosomes are not prohibitive of STF binding. pCREB occupancy at enhancers overlaps with well-positioned nucleosomes, suggesting that pCREB can initiate chromatin engagement and access binding sites organized within a positioned nucleosome. Our work indicates that accessible nucleosomes at enhancers may facilitate cooperativity between STFs and MTFs to ensure rapid transcriptional induction. Retention of MNase accessible nucleosomes at regulatory elements was proposed to play a crucial roles in hormone signaling and tissue specific gene activation[31,38]. Retained nucleosomes likely stabilize the interaction of TFs with the DNA by facilitating interactions between TF-associated factors, such chromatin remodeling complexes, and histone tails [38].

Except for pioneer factors, most TFs are thought to be unable to bind nucleosomal DNA. Although CREB is not traditionally described as a pioneer factor[50,57,58], novel studies revealed the ability of CREB to open chromatin[59]. In an effort to determine the effect of nucleosomes on TF-DNA interactions, CREB was identified as a TFs that displays an orientated, asymmetric, binding preference near the dyad axis of the nucleosome when engaging nucleosomal DNA[60,61]. We hypothesize that MNase accessible nucleosomes within stimuli-responsive enhancers enable cooperative TF binding for rapid gene activation.

Nucleosomes retained within stimuli-inducible enhancers were epigenetically pre-marked by the histone variant H2A.Z. Despite the precise function of H2A.Z at enhancers at remaining unclear, H2A.Z is an important regulator of enhancer activity in response to stimuli. H2A.Z-rich enhancers display higher chromatin accessibility and gene induction by promoting RNA polymerase II recruitment[40,44,62]. In contrast to hormone stimulation, which was found to increase H2A.Z incorporation at enhancer nucleosomes[63], we found little change in H2A.Z distribution after dmPGE$_2$ treatment. Instead, we noted that H2A.Z-variant nucleosomes undergo histone tail acetylation following dmPGE$_2$ stimulation. Acetylated forms of H2A.Z are associated with an open chromatin conformation and directly regulate transcription of enhancer RNAs[64-66]. Our work implies that dmPGE$_2$-inducible H2A.Z acetylation underlies increased low MNase sensitivity and enhanced nucleosomes accessibility at stimuli-responsive enhancers following dmPGE$_2$ treatment. We found that changes in post-translational acetylation of H2A.Z at enhancers correlate directly with gene expression changes, indicating that H2A.Zac is a prerequisite for transcription induction during dmPGE$_2$ stimulation.

Labile, H2A.Z-marked nucleosomes do not present an obstacle for pCREB binding,

but may facilitate TF binding and enhancer activity. We suggest that a critical feature of pCREB is the recruitment of remodelers that opens the local nucleosome structure through H2A.Z acetylation in enhancers. pCREB interacts with a variety of chromatin remodelers. This includes the HATs p300 and Tip60, both known to catalyze H2A.Z acetylation [47,49]. The recruitment of p300 to chromatin in a stimulus-dependent manner observed here is consistent with interactions between chromatin remodelers and other STFs[67-69]. Localization of p300 is concomitant to pCREB binding, suggesting pCREB-p300 complex formation at enhancers upon dmPGE_2 stimulation. Subsequent acetylation of H2A.Z-variant enhancer nucleosomes may create a chromatin environment permissive of enhancer activity and transcription.

In conclusion, this study reveals how specific genomic reorganization at a stimuli-responsive group of enhancers is directly translated into regulatory element activation and transcriptional induction. Our findings support a model where STFs and MTFs cooperate with nucleosomes to regulate activity of cis-regulatory elements that mediate adequate responses to environment signals. While the combination of cooperative lineage-specific MTF and inducible STF binding provides context and responsiveness to external stimuli, histone variant nucleosomes retained within inducible enhancers may facilitate TF binding by stabilizing chromatin complexes. Subsequent acetylation of histone variant nucleosomes by TF-associated nucleosome remodelers creates the accessible nucleosome landscape required at active transcriptional enhancers to ensure strong gene activation.

Methods

Expansion of CD34⁺ cells

Human CD34⁺ hematopoietic stem and progenitor cells (HSPCs), isolated from peripheral blood of granulocyte colony-stimulating factor (G-CSF) mobilized healthy volunteers, were purchased from the Fred Hutchinson Cancer Research Center. The cells were maintained in suspension culture as previously described by Trompouki et al., 2011. Briefly the cells were expanded in StemSpan medium (Stem Cell Technologies Inc.) supplemented with StemSpan CC100 cytokine mix (Stem Cell Technologies Inc.) and 2% Penicillin-Streptomycin for a total of 6 days.

Cell culture

U937 cells were maintained in suspension culture in RPMI-1640 supplemented with 10% (v/v) heat-inactivated fetal bovine serum, 1X Glutamax and 1% Penicillin-Streptomycin at 37° in a humidified atmosphere of 5% CO_2.

dmPGE$_2$ treatment

16,16-dimethyl prostaglandin E2 was purchased reconstituted in DMSO from Cayman Chemicals (cat. #14750), aliquoted and stored in -80°C until use. Cells were counted, collected and resuspended in StemSpan medium with 2% PenStrep (CD34+ Cells) or RPMI with 1% Penicillin-Streptomycin, but in the absence of additional cytokines or growth factors. Cells were treated with dmPGE$_2$ (Cayman chemicals) or DMSO (vehicle control) for 2 hours.

qPCR analysis

RNA was extracted from using the RNeasy plus mini kit (Qiagen). cDNA synthesis was performed using the Superscript VILO (Invitrogen) and using equal amounts of starting RNA. The cDNA was analyzed with the Light Cycler 480 II SYBR green master mix (Applied Biosystems), and the QuantStudio 12K Flex (Applied Biosystems). All samples were prepared in triplicate. The PCR cycle conditions used are: (a) 95°C for 5 min, (b) [95°C for 10 sec, 54°C for 10 sec, 72°C for 15 sec] X 40 cycles. The analysis of Ct values were performed using 2^-ΔΔT method. The PCR primer-pairs used are:

CXCL2	Fwd: AAACCGAAGTCATAGCCACACTC
	Rev: AGCCACCAATAAGCTTCCTCCTTC
CXC4	Fwd: CCTATGCAAGGCAGTCCATGT
	Rev: GGTAGCGGTCCAGACTGATGA
EGR1	Fwd: GCGAGCAGCCCTACGAGCAC
	Rev: TGCAGGCTCCAGGGAAAAGC

FOS	Fwd: TGCCTCTCCTCAATGACCCTGA
	Rev: ATAGGTCCATGTCTGGCACGGA
FOSB	Fwd: TTCTGACTGTCCCTGCCAAT
	Rev: CGGGGTCAGATGCAAAATAC
FOSL2	Fwd: GCAGTTGGGTTTCTGGCTTGAG
	Rev: TCCTGCTACTCCTGGCTCATTC
GAPDH	Fwd: GAAGGTCGGAGTCAACGGATTT
	Rev: GAATTTGCCATGGGTGGAAT
ICER	Fwd: CACCATGGCTGTAACTGGAGATGAC
	Rev: AGGTCCAAGTCAAAGACAGTTACTC
NR4A1	Fwd: GAGTGCACAGAAGAACT
	Rev: CACAGGAGGAGGAAGA
NR4A2	Fwd: CACAGGTTGCAATGCGTTCG
	Rev: TCAATTATTGCTGGCGGTGG
NR4A3	Fwd: CGTCGAAACCGATGTCAGTA
	Rev: GACGACCTCTCCTCCCTTTC
PTGS2	Fwd: GAATCATTCACCAGGCAAATTG
	Rev: TCTGTACTGCGGGTGGAACA

Western blotting

Cells were treated for 2 hours, washed in 1X PBS, and collected in RIPA buffer with protease and phosphatase inhibitors. Samples were run on acrylamide gel and transferred onto a nitrocellulose membrane. Membrane was blocked for one hour in 5% of milk or TBS-T and incubated overnight at 4°C with anti-vinculin (Abcam 73412), anti-CREB (Santa Cruz SC-186), anti-pCREB (Ser133, Cell Signalling #9198), anti-H2A.Z (Abcam 4174), or anti-IgG (Cell Signalling #2729). The next day, membranes were washed, incubated with HRP-conjugated secondary antibodies for 1 hour at room temperature, and developed with SuperSignal West Pico Plus Chemiluminescent substrate.

Co-immunoprecipitation (Co-IP)

Cells were washed twice with 1X PBS. Nuclei were isolated with 0.05% Triton in PBS and lysed in nuclei lysis buffer (20mM Hepes-KOH pH7.9, 25% glycerol, 420mM NaCl, 1.5mM MgCl2, 0.2mM EDTA, 0.5mM DTT). 500ug of protein extracts was used and the salt concentration was diluted from 420mM to 150mM NaCl using 20mM Hepes-KOH pH7.9, 20% glycerol, 0.25mM EDTA, 0.05% NP-40. Cell lysates were pre-cleared with non-antibody bounds beads for 1 hour at 4°C. Antibody bounds protein G Dynabeads were added to pre-cleared lysate and samples incubated overnight at 4°C. Antibodies used: IgG (Cell Signalling #2729), phospho-CREB (Ser133, Cell Signalling #9198). Protein-bead complexes were then washed 5 times with wash buffer (20mM Hepes-KOH pH7.9, 10% Glycerol, 150mM NaCl, 1.5 MgCl2, 0.2mM EDTA, 0.5mM DTT) and beads were boiled

in 50µL Laemmli Buffer for 15min at 95°C to elute proteins. Subsequently, samples were subjected to western blot.

RNA-Seq

RNA from one million cells was isolated using the RNeasy plus mini kit (Qiagen #74134). 5µg of RNA was subjected to ribosomal and mitochondrial RNA depletion using the RiboZero Gold kit (Human/Mouse/Rat, Epicentre #MRZG12324) according to manufacturer's instructions. The ribo-zero treated RNA was used to create multiplexed RNA-Seq libraries using the NEBNext Ultra RNA Library Prep Kit (Illumina E7530) according to the manufacturer's instructions. Briefly 500pg of ribozero treated RNA was fragmented and used to produce cDNA libraries using the NEBnext Ultra RNA library prep kit (NEB, E7530S) according to the manufacturer's protocol. Purified double-stranded cDNA underwent end-repair and dA-tailing reactions following manufacturer's reagents and reaction conditions. The obtained DNAs were used for Adaptor Ligation using adaptors and enzymes provided in NEBNext Multiplex Oligos for Illumina (NEB #E7335) and following recommended reaction conditions. Eluted DNA was enriched with PCR reaction using Fusion High-Fidelity PCR Master Mix kit (NEB, M0531S) and specific index primers supplied in NEBNext Multiplex Oligo Kit for Illumina (Index Primer Set 1, NEB, E7335L). Conditions for PCR used are as follows: 98°C, 30 sec; [98°C, 10 sec; 65°C, 30 sec; 72°C, 30 sec] X 15 cycles; 72°C, 5 min; hold at 4°C. PCR reaction mix was purified using Agencourt AMPure XP beads (1X of reaction volume). Libraries were eluted in 20µl elution buffer. All the libraries went through quality control analysis using an Agilent Bioanalyzer and subjected to next-generation sequencing using Illumina Hiseq 2500 platform. Quality control of RNA-Seq datasets was performed by FastQC and Cutadapt to remove adaptor sequences and low-quality regions. The high-quality reads were aligned to UCSC hg19 for human using Tophat 2.0.11 without novel splicing form calls. Transcript abundance and differential expression were calculated with Cufflinks 2.2.1. FPKM values were used to normalize and quantify each transcript.

ChIP-Seq

For ChIP-Seq experiments the following antibodies were used: H3K27ac (Abcam ab4729), H3K4me1 (Abcam ab8895), H3K27me3, (Abcam ab195477), pCREB (Ser133, Cell Signalling #9198), GATA2 (Santa Cruz sc9008X), H2A.Z (Abcam ab4174), H2A.Zac (Abcam ab18262), H2B (Abcam ab1790), H3.3 (Milipore #09-838), H4 (Abcam ab7311), p300 (Millipore #05-257). ChIP experiments were performed as previously described by Trompouki et al., 2011. Briefly, 20 million cells were crosslinked by the addition of 1/10 volume 11% fresh formaldehyde for 10 min at room temperature. The crosslinking was quenched by the addition of 1/20 volume 2.5M glycine for 5 minutes. Cells were washed twice with ice-cold PBS. Cells were lysed in 10mL of Lysis buffer 1 (50mM HEPES-KOH, pH 7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, plus

protease and phosphatase inhibitors) for 10 min at 4°C. After centrifugation, cells were resuspended in 10 mL of Lysis buffer 2 (10mM Tris-HCl, pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, plus protease and phosphatase inhibitors) for 10 min at room temperature. Cells were pelleted and resuspended in 3mL of sonication buffer (10mM Tris-HCl, pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-Deoxycholate, 0.05% N-lauroylsarcosine, plus protease and phosphatase Inhibitors) and sonicated in a Bioruptor sonicator for 36 cycles of 30 sec each followed by a 1min resting interval. Samples were centrifuged for 10min at 18,000g and 1% Triton-X was added to the supernatant. Prior to the immunoprecipitation, 50mL of protein G beads (Invitrogen 100-04D) for each reaction were washed twice with PBS, 0.5% BSA twice. Finally, the beads were resuspended in 250µL of PBS, 0.5% BSA and 5µg of each antibody. Beads were rotated for at least 6 hours at 4°C and then washed twice with PBS with 0.5% BSA. Cell lysates were added to the beads and incubated at 4°C overnight. Beads were washed 1x with (20mM Tris-HCl. pH 8, 150mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100), 1x with (20mM Tris-HCl, pH 8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100), 1x with (10mM Tris-HCl, pH 8, 250nM LiCl, 2mM EDTA, 1% NP4-0) and 1x with TE and finally resuspended in 200µL elution buffer (50mM Tris-Hcl, pH 8.0, 10mM EDTA, 0.5%–1% SDS). 50µL of cell lysates prior to addition to the beads was kept as input. Crosslinking was reversed by incubating samples at 65°C for at least 6 hours. Afterwards the cells were treated with RNase and proteinase K and the DNA was extracted by Phenol/Chloroform extraction.

ChIP-Seq libraries were prepared using the following protocol. End repair of immunoprecipitated DNA was performed using the End-It End-Repair kit (Epicentre, ER81050) and incubating the samples at 25°C for 45 min. End-repaired DNA was purified using AMPure XP Beads (1.8X of the reaction volume) (Agencourt AMPure XP – PCR purification Beads, BeckmanCoulter, A63881) and separating beads using DynaMag-96 Side Skirted Magnet (Life Technologies, 12027). A-tails were added to the end-repaired DNA using NEB Klenow Fragment Enzyme (3'-5' exo, M0212L), 1X NEB buffer 2 and 0.2mM dATP (Invitrogen, 18252-015) and by incubating the reaction mix at 37µC for 30 min. A-tailed DNA was cleaned up using AMPure beads (1.8X of reaction volume). Subsequently, cleaned up A-tailed DNA went through Adaptor ligation reaction using Quick Ligation Kit (NEB, M2200L) following manufacturer's protocol. Adaptor-ligated DNA was first cleaned up using AMPure beads (1.8X of reaction volume), eluted in 100µl and then size-selected using AMPure beads (0.9X of the final supernatant volume, 90µl). Adaptor-ligated DNA fragments of proper size were enriched with PCR reaction using Fusion High-Fidelity PCR Master Mix kit (NEB, M0531S) and specific index primers supplied in NEBNext Multiplex Oligo Kit for Illumina (Index Primer Set 1, NEB, E7335L). Conditions for PCR used are as follows: 98°C, 30 sec; [98°C, 10 sec; 65°C, 30 sec; 72°C, 30 sec] X 15 to 18 cycles; 72°C, 5 min; hold at 4°C. PCR enriched fragments were cleaned up using AMPure beads (1X of reaction volume). Libraries were

eluted in 20µl elution buffer. All the libraries went through quality control analysis using an Agilent Bioanalyzer and subjected to next-generation sequencing using Illumina Hiseq 2500 platform. All the libraries went through quality control analysis using an Agilent Bioanalyzer and subjected to next-generation sequencing using Illumina Hiseq 2500 platform. All ChIP-Seq datasets were aligned to UCSC build version hg19 of the human genome using Bowtie2 (version 2.2.1; Langmead et al., 2012) with the following parameters: -end-to-end, -N0, -L20. We used the MACS2 version 2.1.0 (Zhang et al., 2008) peak-finding algorithm to identify regions of ChIP-Seq peaks, with a q-value threshold of enrichment of 0.05 and false discovery rate of < 0.01% for all datasets. The genome-wide occupancy profile figures were generated by deeptools (Ramirez et al., 2016) using the reference-point mode and the scale-regions mode. The genomic distribution of peaks was plotted using the ChIPSeeker R package, annotatePeak to assign peaks to a genomic annotation, which includes whether a peak is in the TSS, Exon, 5' UTR, 3' UTR, Intronic or Intergenic. The genome annotation is from the R-bioconductor annotation packages. Heat maps of the ChIP-Seq binding were generated using the input normalized results of the MACS peak calling output. The outputted bedGraph files were converted to BigWig files. Those files were then processed using computeMatrix and plotHeatmap tools within the deeptools 3.0 package. Enhancers were assigned to genes using GREAT to the nearest genes within 15 kb of peaks.

Defining Enhancer Categories

We used the SPP package to call clusters of H3K27ac enrichment, normalized to input, from ChIP-Seq data (Kharchenko et al., 2008). Regions within 500bp of each other were merged and only regions reproduced between two independent biological H3K27ac ChIP-Seq replicates (in either sample $dmPGE_2$ or DMSO) were included for further analysis. Enhancers were defined as TSS distal H3K27ac regions that are 1000bp in length and located 2kb away from TSS. This yielded a total of 25,998 H3K27ac enriched ChIP-Seq regions, here named enhancers. P-value was computed using paired t-test between $dmPGE_2$ and DMSO on enrichment values for every region.

Called enhancers were classified based on the three following criteria and using cut-off described in the table below: (1) H3K27ac enrichment in each replicate in each condition, (2) delta H3K27ac enrichment upon stimulation, (3) p-value of delta H3K27ac enrichment. All regions not classified as not meeting above mentioned criteria were classified as 'Background' enhancers.

	H3K27ac Enrichment		Δ Enrichment	Significance
	DMSO	$dmPGE_2$	[$dmPGE_2$ - DMSO]	p-value
De Novo	≤ 1	≥ 3	≥ 1	≤ 0.05
Enhanced	≥ 1	≥ 2	≥ 2	≤ 0.05

ATAC-Seq

50.000 cells per condition were harvested by spinning at 500g for 5 min, 4°C. Cells were washed once with 50µl of cold 1X PBS and spun down at 500g for 5 min, 4°C. After discarding supernatant, cells were lysed using 50µl cold lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL) and spun down immediately at 500g for 10 min at 4°C. Then the cells were precipitated and kept on ice and subsequently resuspended in 25µl 2X Tagment DNA Buffer (Illumina Nextera kit), 2.5µl Transposase enzyme (Illumina Nextera kit, 15028252) and 22.5µl Nuclease-free water in a total of 50uL reaction for 1 hour at 37°C. DNA was then purified using Qiagen MinElute PCR purification kit (28004) in a final volume of 10µl. Libraries were constructed according to Illumina protocol using the DNA treated with transposase, NEB PCR master mix, Sybr green, universal and library-specific Nextera index primers. The first round of PCR was performed under the following conditions: 72°C, 5 min; 98°C, 30 sec; [98°C, 10 sec; 63°C, 30 sec; 72°C, 1 min] X 5 cycles; hold at 4°C. Reactions were kept on ice and using a 5µL reaction aliquot, the appropriate number of additional cycles required for further amplification was determined in a side qPCR reaction: 98°C , 30 sec; [98°C, 10 sec; 63°C, 30 sec; 72°C, 1 min] X 20 cycles; hold at 4°C. Upon determining the additional number of PCR cycles required further for each sample, library amplification was conducted using the following conditions: 98°C, 30 sec; [98°C, 10 sec; 63°C, 30 sec; 72°C, 1 min] X appropriate number of cycles; hold at 4°C. Libraries prepared went through quality control analysis using an Agilent Bioanalyzer and then subjected to next generation sequencing using Illumina Hiseq 2500 platform. We used the MACS2 version 2.1.0 (Zhang et al., 2008) peak-finding algorithm to identify regions of ATAC-Seq peaks, with the following parameter -- nomodel --shift -100 --extsize 200. A q-value threshold of enrichment of 0.05 was used for all datasets.

MNase-Seq

$CD34^+$ HSPC cells were crosslinked by the addition of 1/10 volume 11% fresh formaldehyde for 10 min at room temperature. The crosslinking was quenched by the addition of 1/20 volume 2.5M glycine for 5 minutes. Cells were washed twice with ice-cold PBS. For MNase digestion, the nuclei pellet was resuspended in MNase digestion buffer (50mM Tris, pH 7.4, 25mM KCl, 4mM MgCl2, 1mM CaCl2, 12.5% Glycerol and COMPLETE protease inhibitors (Roche)). Digestion took place with 10^6 cells per titration point in a volume of 500µl MNase digestion buffer. Either 1, 2, 4, 8, 16, 32, 64, 128 or 256 units of MNase (Worthington Biochemical) were added to pre-warmed nuclei and incubated at 25 °C for 15 min. Digestion was halted with 25mM EDTA/EGTA and 0.5% SDS and 125mM NaCl was added to the samples. Digestions were incubated with RNase (Roche) for 1 hour at 37 °C, with proteinase K (Roche) for 1 hour at 55 °C, and cross-link reversal was performed at 65 °C for 16 hours. DNA was purified by Phenol/Chloroform extraction and ethanol precipitation. MNase digestion was evaluated on a 2% agarose gel

and fragments from four MNase concentrations representing 10%, 25%, 50% and 75% mono-nucleosomal fragments were individually prepared for next generation sequencing. Ampure SPRI beads (Beckman Coulter) were used in a double size selection with ratios of 0.7X and 1.7X to obtain a range of fragment sizes from 100bp to 1,000bp. DNA was eluted from the beads and used as input into the library preparation protocol. DNA libraries were prepared for each individual titration point using the NEBNext Ultra II DNA Library Prep Kit for Illumina (E7370, New England Biolabs) and barcoded using NEBNext Multiplex Oligos for Illumina (Index Primers Set 1 & 2; New England Biolabs). Number of PCR cycles was calculated using a real-time qPCR-based approach (Lion et al., 2020). Libraries prepared went through quality control analysis using an Agilent Bioanalyzer. Four barcoded titration libraries were pooled in one sample, and paired-end 50-cycle sequencing in an Illumina Hiseq 2500 instrument was performed. Three biological replicated were sequenced. The sequenced paired-end reads were mapped to hg19 using Bowtie aligner v. 0.12.9. Only uniquely mapped reads with no more than two mismatches were retained. The reads with the insert sizes <50 bp or >500 bp were filtered out. Genomic positions with the numbers of mapped tags above the significance threshold of Z-score=7 were identified as anomalous, and the tags mapped to such positions were discarded. Read frequencies were computed in 300bp non-overlapping bins for each titration point independently. The read frequencies were normalized by the corresponding library sizes to represent values per one million of mapped reads. Nucleosome occupancy analysis was carried out as previously described by Mieczkowski et al., 2016.

Acknowledgments

We thank A. Choudhuri and J.M. Ordovas-Montanes for critical reading of our manuscript; A. Choudhuri, R. E. Kingston and the members of the Zon laboratory for discussions on the project. A.S. was supported bya Boehringer Ingelheim PhD Fellowship. A.S. was supported by a Boehringer Ingelheim PhD Fellowship. This work was supported by the following grants from L.I.Z.: R01 HL04880, P015PO1HL32262-32, 5P30 DK49216, 5R01 DK53298, 5U01 HL10001-05, R24 DK092760, 1R24OD017870-01.

Author contributions

A.S., E.M.F, and M.E.M., performed experiments. M.P. and B.M. performed bioinformatics analysis of next generation sequencing data. A.C., R.E.K., and S.Y. provided insights on data analysis and interpretation. L.I.Z conceived and managed the study. M.Y.T. managed computational analysis of the study. A.S. and L.I.Z. wrote the manuscript.

Conflict of interests

L.I.Z. is founder and stockholder of Fate, Inc., Scholar Rock, Camp4 therapeutics and a scientific advisor for Stemgent. The other authors declare no competing interests.

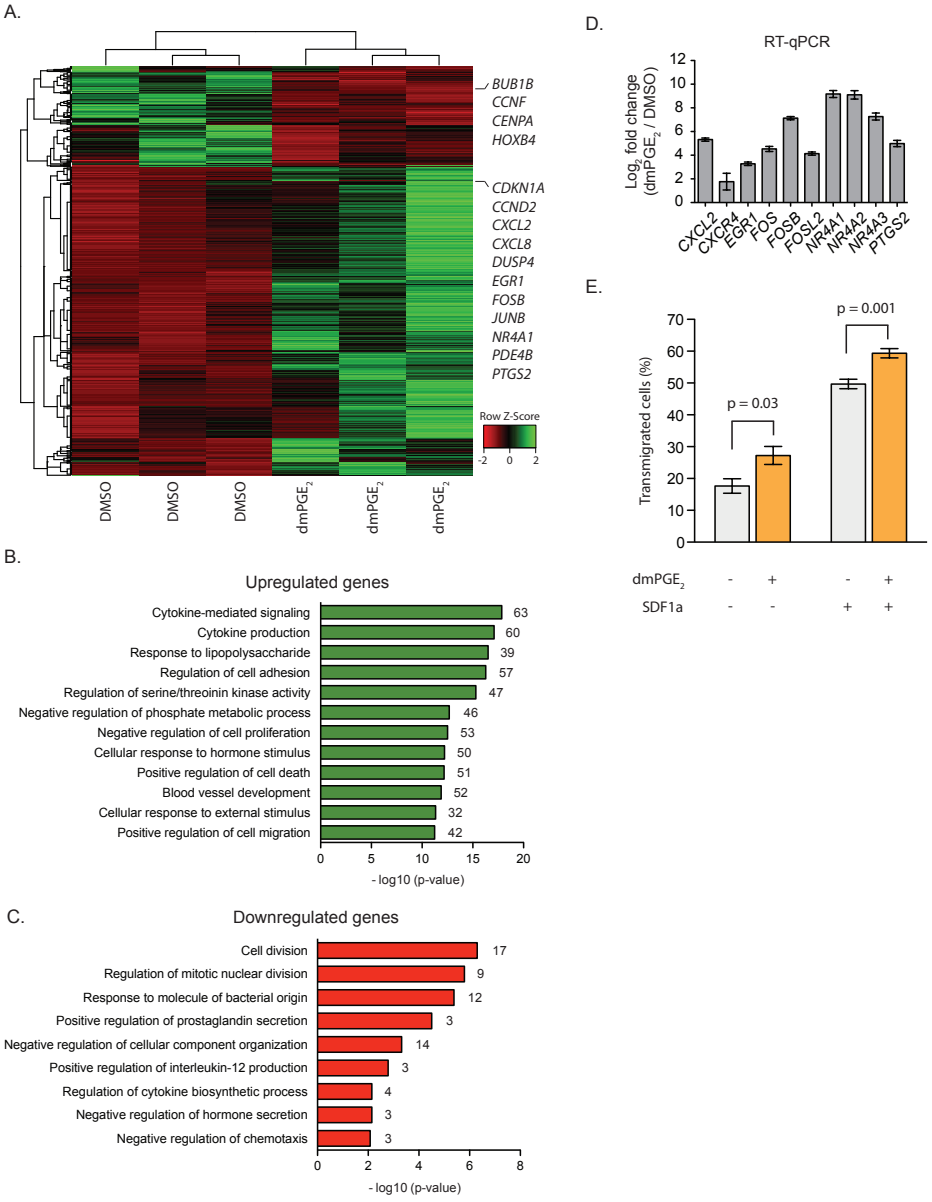
References

1. North, T. E. et al. Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. Nature 447, 1007–1011 (2007).
2. Cutler, C. et al. Prostaglandin-modulated umbilical cord blood hematopoietic stem cell transplantation. Blood 122, 3074–3081 (2013).
3. Morrison, S. J. & Scadden, D. T. The bone marrow niche for haematopoietic stem cells. Nature 505, 327–334 (2014).
4. Vihervaara, A., Duarte, F. M. & Lis, J. T. Molecular mechanisms driving transcriptional stress responses. Nat. Rev. Genet. 19, 385–397 (2018).
5. de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. Nat. Rev. Genet. 12, 833–845 (2011).
6. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 13, 613–626 (2012).
7. Natoli, G. Maintaining cell identity through global control of genomic organization. Immunity 33, 12–24 (2010).
8. Kim, H. D. & O'Shea, E. K. A quantitative model of transcription factor-activated gene expression. Nature Structural & Molecular Biology 15, 1192–1198 (2008).
9. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. 20, 207–220 (2019).
10. Bell, O., Tiwari, V. K., Thomä, N. H. & Schübeler, D. Determinants and dynamics of genome accessibility. Nat. Rev. Genet. 12, 554–564 (2011).
11. Miller, S. B. Prostaglandins in health and disease: an overview. Semin Arthritis Rheum 36, 37–49 (2006).
12. Goessling, W. et al. Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. Cell 136, 1136–1147 (2009).
13. Hoggatt, J., Singh, P., Sampath, J. & Pelus, L. M. Prostaglandin E2 enhances hematopoietic stem cell homing, survival, and proliferation. Blood 113, 5444–5455 (2009).
14. Goessling, W. et al. Prostaglandin E2 enhances human cord blood stem cell xenotransplants and shows long-term safety in preclinical nonhuman primate transplant models. Cell Stem Cell 8, 445–458 (2011).
15. Klein, T., Shephard, P., Kleinert, H. & Kömhoff, M. Regulation of cyclooxygenase-2 expression by cyclic AMP. Biochim Biophys Acta 1773, 1605–1618 (2007).
16. Sinclair, A. et al. CXCR2 and CXCL4 regulate survival and self-renewal of hematopoietic stem/progenitor cells. Blood 128, 371–383 (2016).
17. Freire, P. R. & Conneely, O. M. NR4A1 and NR4A3 restrict HSC proliferation via reciprocal regulation of C/EBPα and inflammatory signaling. Blood 131, 1081–1093 (2018).
18. Santaguida, M. et al. JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. Cancer Cell 15, 341–352 (2009).
19. Antonchuk, J., Sauvageau, G. & Humphries, R. K. HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. Cell 109, 39–45 (2002).
20. Luan, B. et al. CREB pathway links PGE2 signaling with macrophage polarization. Proc. Natl. Acad. Sci. U.S.A. 112, 15642–15647 (2015).
21. Parker, D. et al. Phosphorylation of CREB at Ser-133 induces complex formation with CREB-binding protein via a direct mechanism. Mol. Cell. Biol. 16, 694–703 (1996).
22. Gonzalez, G. A. & Montminy, M. R. Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell 59, 675–680 (1989).
23. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences 107, 21931–21936 (2010).
24. Malik, A. N. et al. Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. Nat Neurosci 17, 1330–1339 (2014).
25. Ostuni, R. et al. Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell 152, 157–171 (2013).

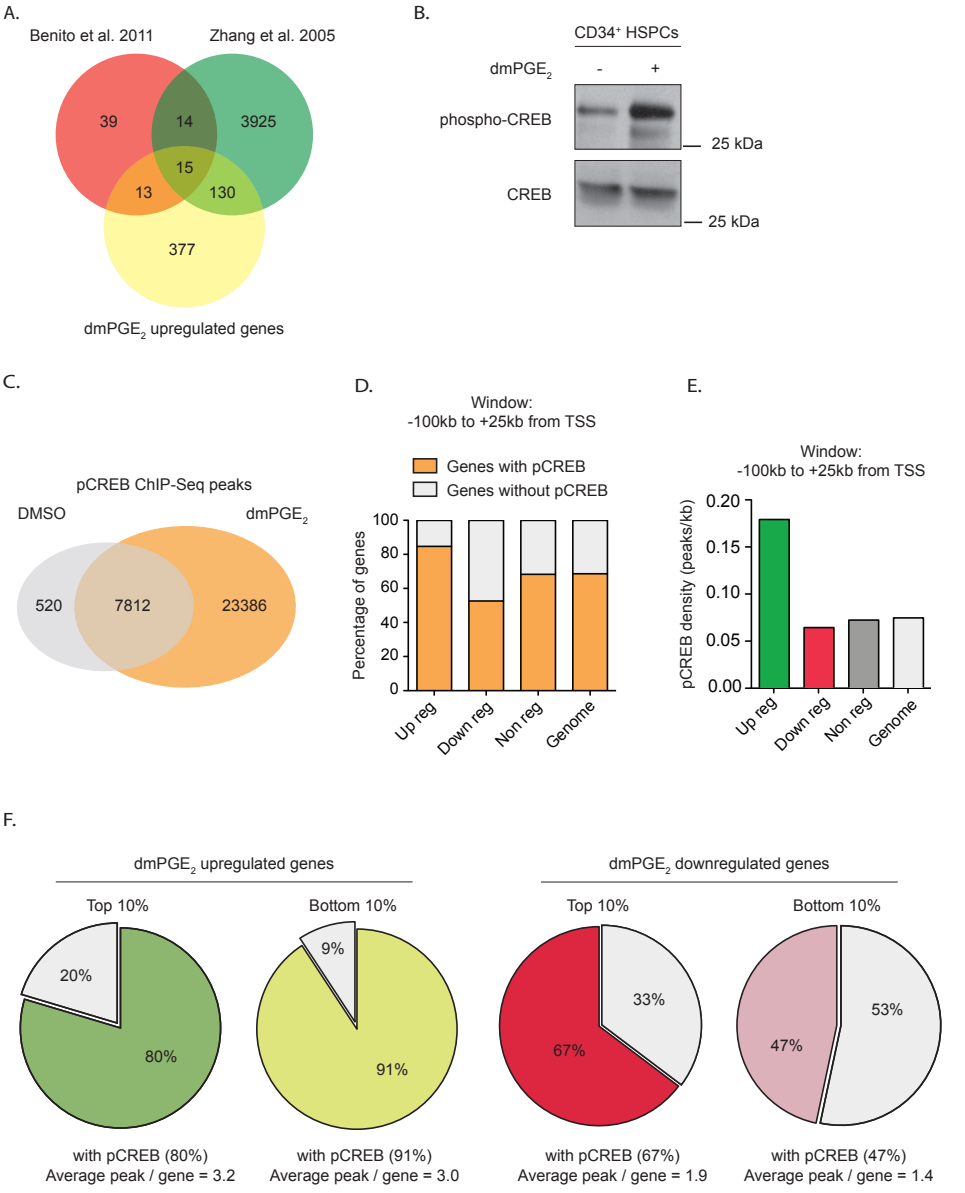
26. Zhang, B. et al. A dynamic H3K27ac signature identifies VEGFA-stimulated endothelial enhancers and requires EP300 activity. Genome Research 23, 917–927 (2013).
27. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. Nature 489, 75–82 (2012).
28. Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell 147, 577–589 (2011).
29. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat. Genet. 43, 264–268 (2011).
30. Mieczkowski, J. et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nature Communications 7, 11485 (2016).
31. Iwafuchi-Doi, M. et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol. Cell 62, 79–91 (2016).
32. Mueller, B. et al. Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes & Development 31, 451–462 (2017).
33. Petesch, S. J. & Lis, J. T. Overcoming the nucleosome barrier during transcript elongation. Trends Genet. 28, 285–294 (2012).
34. Zhang, Y. et al. Canonical nucleosome organization at promoters forms during genome activation. Genome Research 24, 260–266 (2014).
35. Schones, D. E. et al. Dynamic regulation of nucleosome positioning in the human genome. Cell 132, 887–898 (2008).
36. Valouev, A. et al. Determinants of nucleosome organization in primary human cells. Nature 474, 516–520 (2011).
37. Ramani, V., Qiu, R. & Shendure, J. High Sensitivity Profiling of Chromatin Structure by MNase-SSP. Cell Rep 26, 2465–2476.e4 (2019).
38. Ballaré, C. et al. Nucleosome-driven transcription factor binding and gene regulation. Mol. Cell 49, 67–79 (2013).
39. Talbert, P. B. & Henikoff, S. Histone variants — ancient wrap artists of the epigenome. Nature Reviews Molecular Cell Biology 11, 264–275 (2010).
40. Brunelle, M. et al. The histone variant H2A.Z is an important regulator of enhancer activity. Nucleic Acids Res. 43, 9742–9756 (2015).
41. Chen, P. et al. H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. Genes & Development 27, 2109–2124 (2013).
42. Hu, G. et al. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell Stem Cell 12, 180–192 (2013).
43. Jin, C. & Felsenfeld, G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & Development 21, 1519–1529 (2007).
44. Giaimo, B. D., Ferrante, F., Herchenröther, A., Hake, S. B. & Borggrefe, T. The histone variant H2A.Z in gene regulation. Epigenetics Chromatin 12, 37–22 (2019).
45. Adam, M., Robert, F., Larochelle, M. & Gaudreau, L. H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol. Cell. Biol. 21, 6270–6279 (2001).
46. Gévry, N. et al. Histone H2A.Z is essential for estrogen receptor signaling. Genes & Development 23, 1522–1533 (2009).
47. Gavaravarapu, S. & Kamine, J. Tip60 inhibits activation of CREB protein by protein kinase A. Biochem. Biophys. Res. Commun. 269, 758–766 (2000).
48. Vo, N. & Goodman, R. H. CREB-binding protein and p300 in transcriptional regulation. J. Biol. Chem. 276, 13505–13508 (2001).
49. Colino-Sanguino, Y. et al. A Read/Write Mechanism Connects p300 Bromodomain Function to H2A.Z Acetylation. iScience 21, 773–788 (2019).
50. Zhang, H., Roberts, D. N. & Cairns, B. R. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. Cell 123, 219–231 (2005).
51. Van de Velde, S. et al. CREB Promotes Beta Cell Gene Expression by Targeting Its Coactivators

to Tissue-Specific Enhancers. Mol. Cell. Biol. 39, 141 (2019).
52. Swinstead, E. E. et al. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell 165, 593–605 (2016).
53. Liu, Z. & Kraus, W. L. Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. Mol. Cell 65, 589–603.e9 (2017).
54. Donaghey, J. et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. Nat. Genet. 50, 250–258 (2018).
55. de Pater, E. et al. Gata2 is required for HSC generation and survival. J. Exp. Med. 210, 2843–2850 (2013).
56. Wu, D. et al. Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer. Nucleic Acids Res. 42, 3607–3622 (2014).
57. Carroll, J. S. et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122, 33–43 (2005).
58. Li, G. et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148, 84–98 (2012).
59. Sherwood, R. I. et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat Biotechnol 32, 171–178 (2014).
60. Zhu, F. et al. The interaction landscape between transcription factors and the nucleosome. Nature 562, 76–81 (2018).
61. Zaret, K. S. Pioneer Transcription Factors Initiating Gene Network Changes. Annu. Rev. Genet. 54, annurev–genet–030220–015007 (2020).
62. Dalvai, M., Fleury, L., Bellucci, L., Kocanova, S. & Bystricky, K. TIP48/Reptin and H2A.Z requirement for initiating chromatin remodeling in estrogen-activated transcription. PLoS Genet. 9, e1003387 (2013).
63. Johnson, T. A. et al. Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin in vivo. Nucleic Acids Res. 46, 203–214 (2018).
64. Valdés-Mora, F. et al. Acetylated histone variant H2A.Z is involved in the activation of neo-enhancers in prostate cancer. Nature Communications 8, 1346–17 (2017).
65. Valdés-Mora, F. et al. Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. Genome Research 22, 307–321 (2012).
66. Ishibashi, T. et al. Acetylation of vertebrate H2A.Z and its effect on the structure of the nucleosome. Biochemistry 48, 5007–5017 (2009).
67. Zhong, H., Voll, R. E. & Ghosh, S. Phosphorylation of NF-kappa B p65 by PKA stimulates transcriptional activity by promoting a novel bivalent interaction with the coactivator CBP/p300. Mol. Cell 1, 661–671 (1998).
68. Merika, M., Williams, A. J., Chen, G., Collins, T. & Thanos, D. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Mol. Cell 1, 277–287 (1998).
69. Hottiger, M. O. & Nabel, G. J. Interaction of human immunodeficiency virus type 1 Tat with the transcriptional coactivators p300 and CREB binding protein. J Virol 72, 8252–8256 (1998).

Supplemental Figures



Supplemental Figure 1. dmPGE2 induces acute transcriptional responses in HSPCs. (A) Hierarchical clustering heatmap of FPKM values from differentially expressed genes (DEGs; 687) 2 hours post-treatment. DEG criteria: FPKM ≥ 1 after treatment; fold change ≥ 1.5 or ≤ 0.67. (B, C) GO term enrichment analysis of genes upregulated (C, in green) and downregulated (D, in red) 2 hours post dmPGE$_2$ treatment. The number of genes associated with each GO term are shown at the end of the bars. P-values were calculates using hypergeometric test and Benjamini-Hochberg correction. (D) RT-qPCR in CD34$^+$ HSPCs of DEGs (mean values ± SEM). (E) Transwell migration assay of CD34$^+$ HSPCs exposed to dmPGE2 or DMSO. After 2h stimuli were washed out and cells were then placed in the top chamber of the transwell system, with or without recombinant human SDF1α in the bottom chamber. After 24 hours, cells migration to the bottom chamber was quantified as percentage of total cells seeded (mean values ± SEM).



Supplemental Figure 2. pCREB binds near differentially expressed genes. (A) Venn diagram showing overlap between upregulated genes (535) and previously identified CREB target genes. (B) Western blot analysis for phospho-CREB in CD34+ HSPCs stimulated with vehicle control (DMSO) or $dmPGE_2$ for 2 hours. Total CREB protein was used as loading control. (C) Venn diagram showing overlap between pCREB peaks present in DMSO after $dmPGE_2$ stimulation, as identified by ChIP-Seq. (D) Number of genes containing at least one pCREB peak in the proximity after $dmPGE_2$ stimulation within -100kb upstream to +25kb downstream of the TTS. (E) Correlation between pCREB binding and gene expression in response to $dmPGE_2$. pCREB density was calculated by dividing the total number of pCREB peaks associated to each gene category (up-, down-, and nonregulated genes) by the total amount of base pairs that this category occupies in the genome. Peak density in the genome was calculated by considering random distribution of pCREB sites in the whole genome. (F) pCREB in $dmPGE_2$-response genes. Top and bottom 10% correspond to the 10% most upregulated and downregulated genes, respectively. pCREB peaks were assigned to a gene when located within a window from -5kb upstream to +5kb downstream of the TTS.

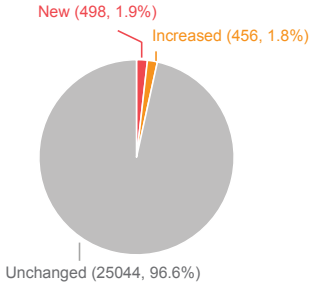
A.

CD34⁺ HSPCs → 2 hour treatment with DMSO or dmPGE$_2$ → ChIP-Seq for H3K27ac (2 biological replicates)

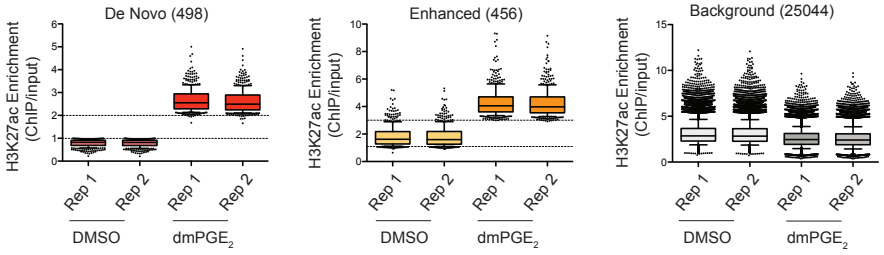
	Enhancer classification criteria			
	H3K27ac Enrichment DMSO	H3K27ac Enrichment dmPGE$_2$	Δ Enrichment (dmPGE$_2$ - DMSO)	P-value
De Novo	≤ 1	≥ 2	≥ 1	≤ 0.05
Enhanced	≥ 1	≥ 3	≥ 2	≤ 0.05
Background	All other regions			

B.

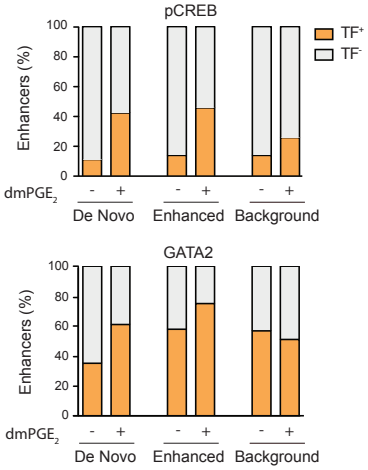
Total	25998 peaks (100%)
Background	25044 peaks (96.3%)
Stimuli-Inducible	954 peaks (3.7%)
De Novo	498 peaks (1.9%)
Enhanced	456 peaks (1.8%)



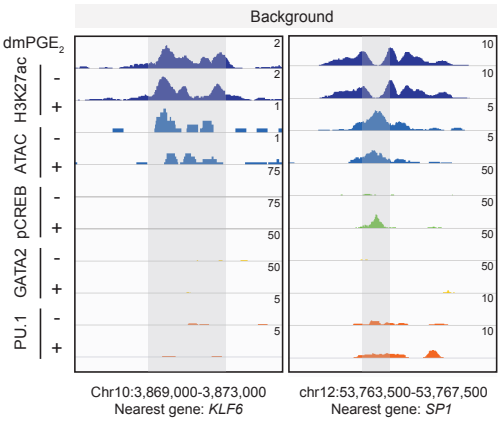
C.



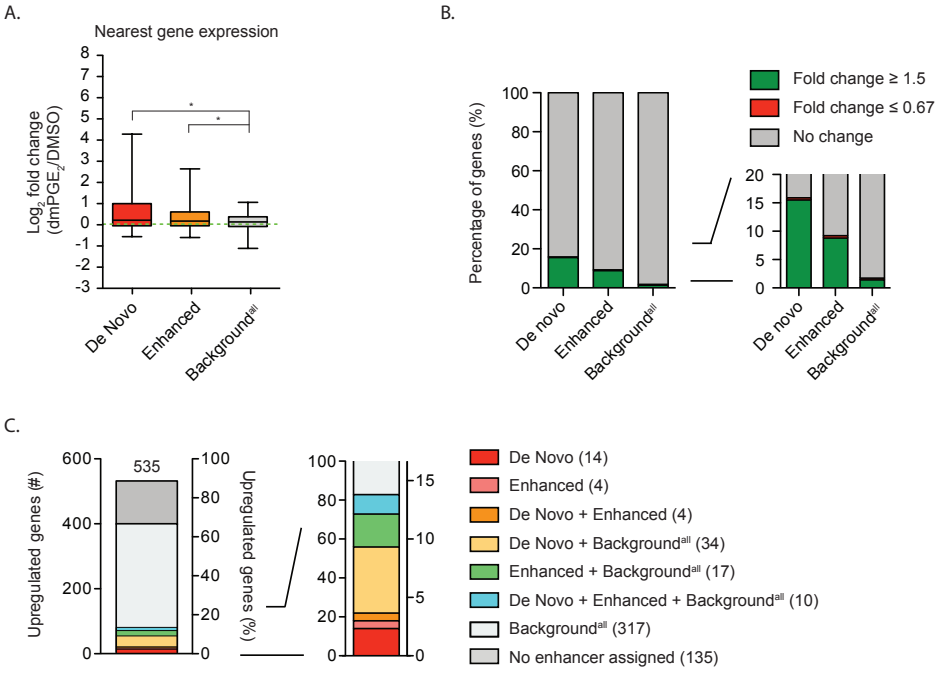
D.



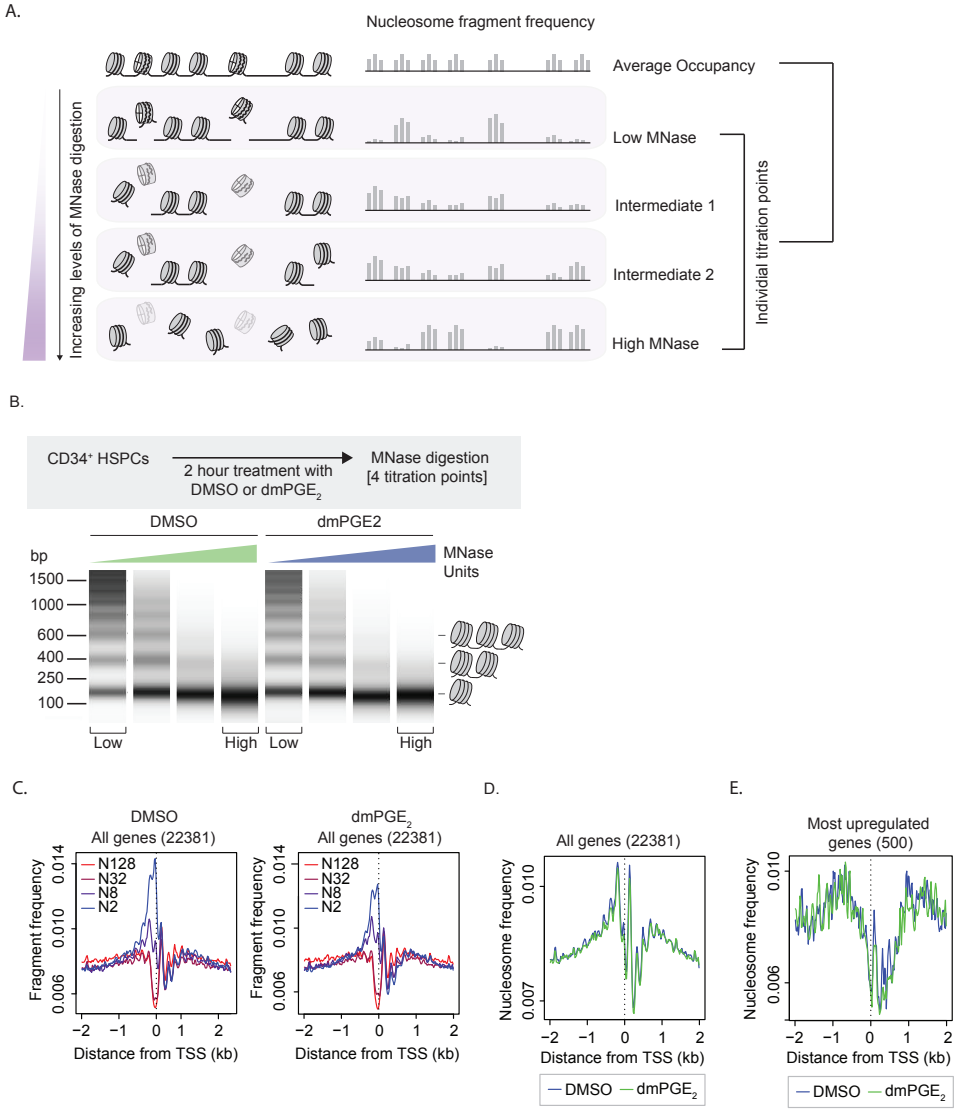
E.



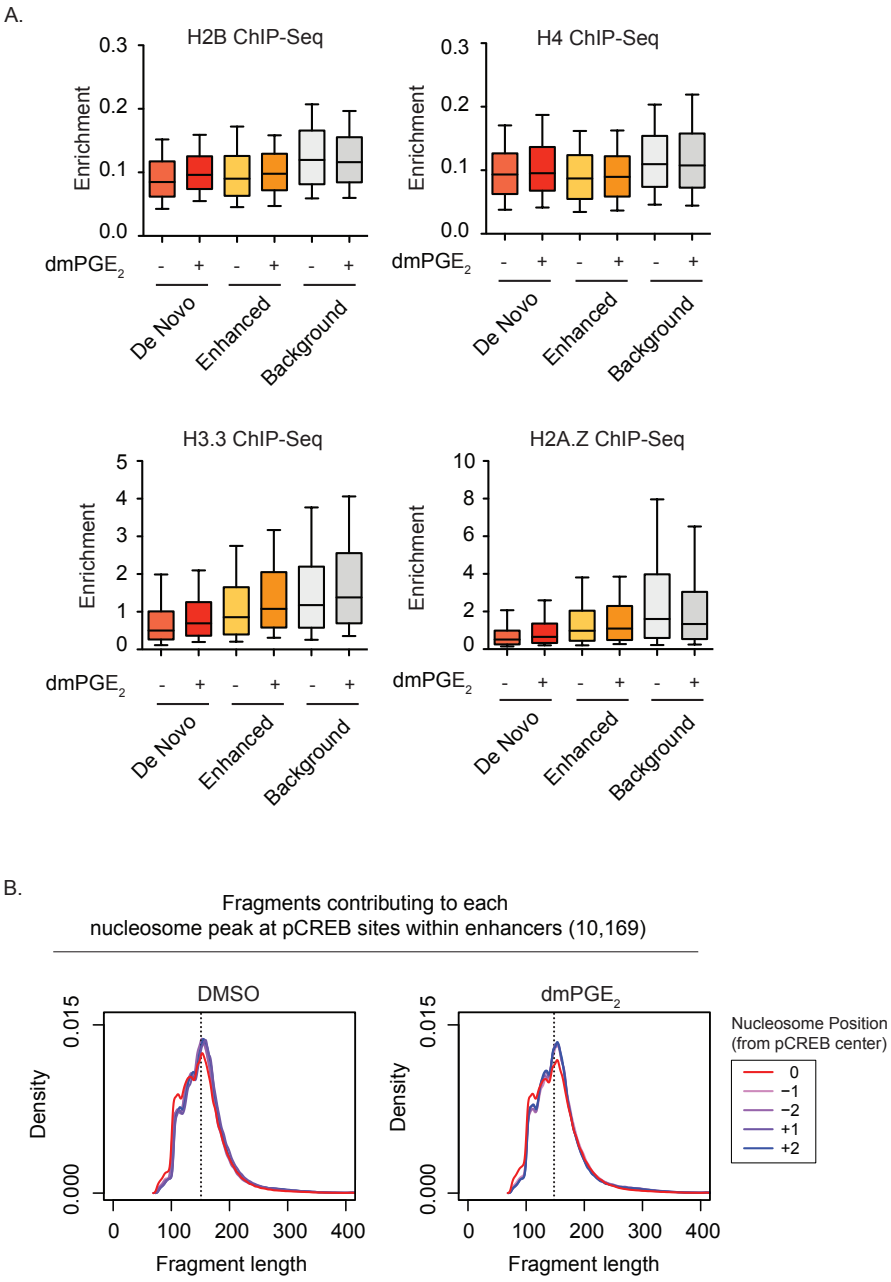
Supplemental Figure 3. Identification of stimuli-inducible enhancers in HSPCs. (A) Experimental set up and enhancer classification criteria. (B) Number of identified enhancer regions and their distribution into the different categories based on two independent replicate experiments. (C) H3K27ac enrichment levels at enhancers within each indicated category for 2 biologically independent replicate (rep) ChIP-Seq experiments. Dotted lines indicate cutoff values used for enhancer classifications. (D) Number of enhancers as percentage of total within each category containing enrichment for pCREB (upper panel) and GATA2 (lower panel). (E) Enrichment of histone mark, ATAC accessibility and transcription factor binding in response to dmPGE$_2$ at representative 2 background enhancers. Genomic location of presented window and nearest gene are indicated at the bottom of the panel.



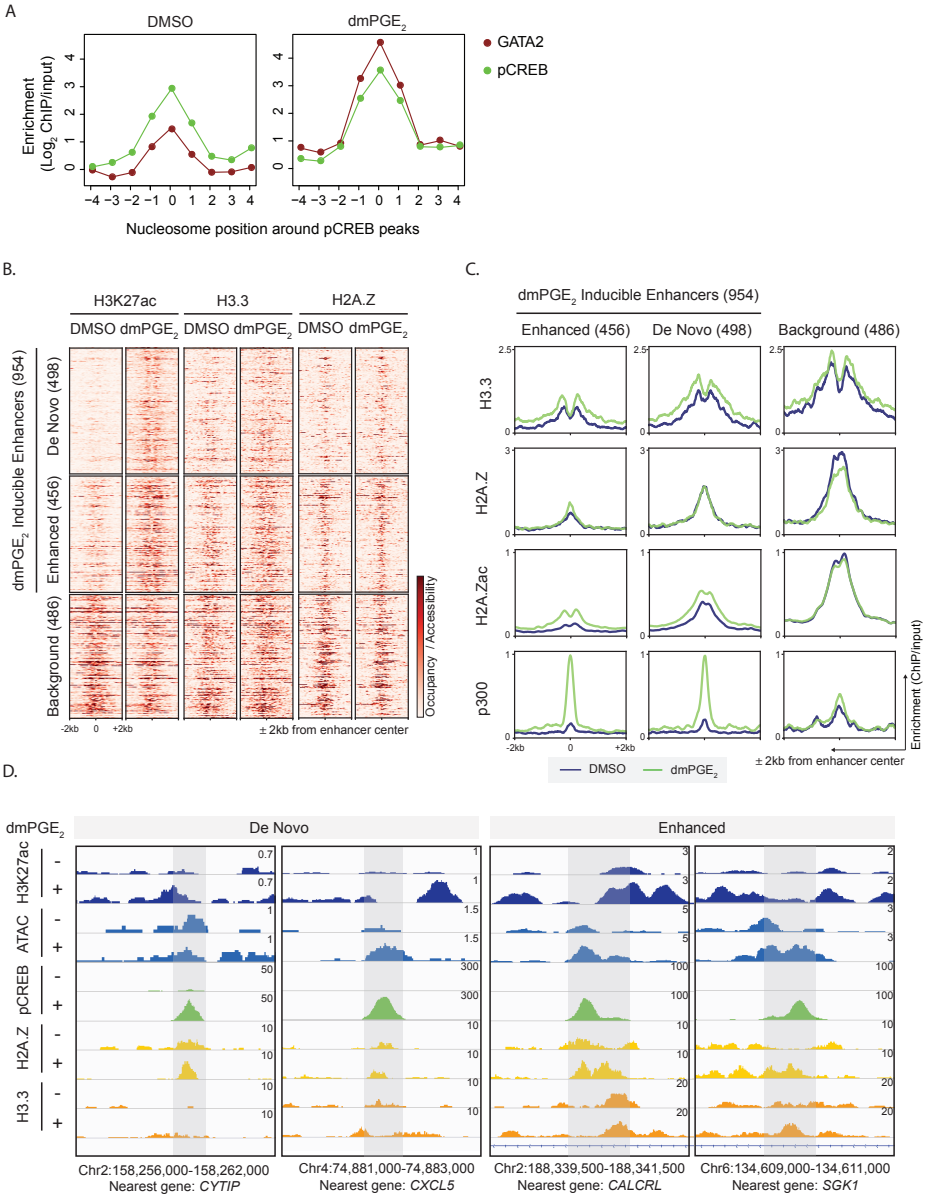
Supplemental Figure 4. Stimuli-responsive enhancers mediate gene expression changes. (A) Gene expression changes of genes associated with stimuli-responsive and background enhancers. Enhancers were assigned to an individual nearest gene. Only genes with a mapped TSS within 15kb of an enhancer were considered. Box plots shows median, 25th and 75th percentiles, whiskers are from 5th and 95th percentiles. (B) Percentages of enhancer nearest genes with fold changes in expression ≥ 1.5-fold or ≤ 0.67-fold for each enhancer category. (C) Upregulated genes with a fold change in expression ≥ 1.5 (535) and their associated enhancers. For all analysis presented in A, B, and C the entire set of background enhancers (25,044) was used. * = $p < 0.0001$



Supplemental Figure 5. MNase-sequencing in DMSO and dmPGE$_2$ treated HSPCs. (A) Schematic representation of MNase-Seq using 4 titration point. Average nucleosome occupancy is determined through pooled analysis of 4 individual MNase digestion levels, as indicated (n = 3 independent biological experiments). (B) Capillary electrophoresis of digestion products from a typical MNase titration experiment. Cells are stimulated for 2 hours with dmPGE$_2$ or vehicle control (DMSO) after which MNase digestion was performed. (C) MNase-Seq profiles around TSS (transcription start sites) of all genes. Colors indicates MNase concentration levels (2, 8, 32 and 128 Units of MNase), with blue corresponding to the lowest concentration and red corresponding to the highest. (D) The average nucleosome profile at the TSS of all genes in DMSO treated (blue) and dmPGE$_2$ (green) treated HSPCs, as determined from 4 individual MNase titration point per experimental condition. (E) The average nucleosome profile at the TSS of the 500 most upregulated in DMSO treated (blue) and dmPGE$_2$ (green) treated HSPCs, as determined from 4 individual MNase titration point per experimental condition.



Supplemental Figure 6. Histones enrichment and MNase-fragment size at enhancers. (A) Box and whisker plots show H2B, H3, H2A.Z and H3.3 enrichment (ChIP/input) at the central 500bp of stimuli-responsive and background enhancers, in DMSO and dmPGE$_2$ treated HSPCs. Box plots shows median, 25th and 75th percentiles, whiskers are from 10th and 90th percentiles. For analysis presented here, a randomly sampled comparable number of background enhancers (486) is used. (B) Size distribution of DNA fragment reads mapped to the corresponding nucleosome position displayed. Sequencing libraries prepared from MNase-generated fragments were subjected to paired-end sequencing, and the sizes of the fragments were inferred from the positions of the mapped ends. Nucleosome position 0 indicates the nucleosome overlapping with pCREB peak centers.



Supplemental Figure 7. Enhancers nucleosomes contain histone variants. (A) Heat maps of histone variant enrichment around enhancers before and after dmPGE$_2$ treatment. H3K27ac enriched regions identified using ChIP-Seq are classified as De Novo, Enhanced or Background enhancers according to the change in H3K27ac levels observed following dmPGE$_2$ stimulation (n = 2 biologically independent experiments). (B) Enrichment of histone variants at nucleosomes positions surrounding pCREB peaks within enhancers before and after dmPGE$_2$ stimulation. Position 0 indicates the nucleosome overlapping with pCREB peak centers. (C) Enrichment of ATAC accessibility, pCREB binding and histone variant deposition in response to dmPGE$_2$ at 4 representative stimuli-response enhancers. Genomic location of presented window and nearest gene are indicated at the bottom of the panel. (D) Average enrichment profiles of acetylated (ac) H2A.Z and p300 before and after dmPGE$_2$ treatment in De Novo, Enhanced or Background enhancers. For all analysis presented in B and D a randomly sampled comparable number of background enhancers (486) is shown.

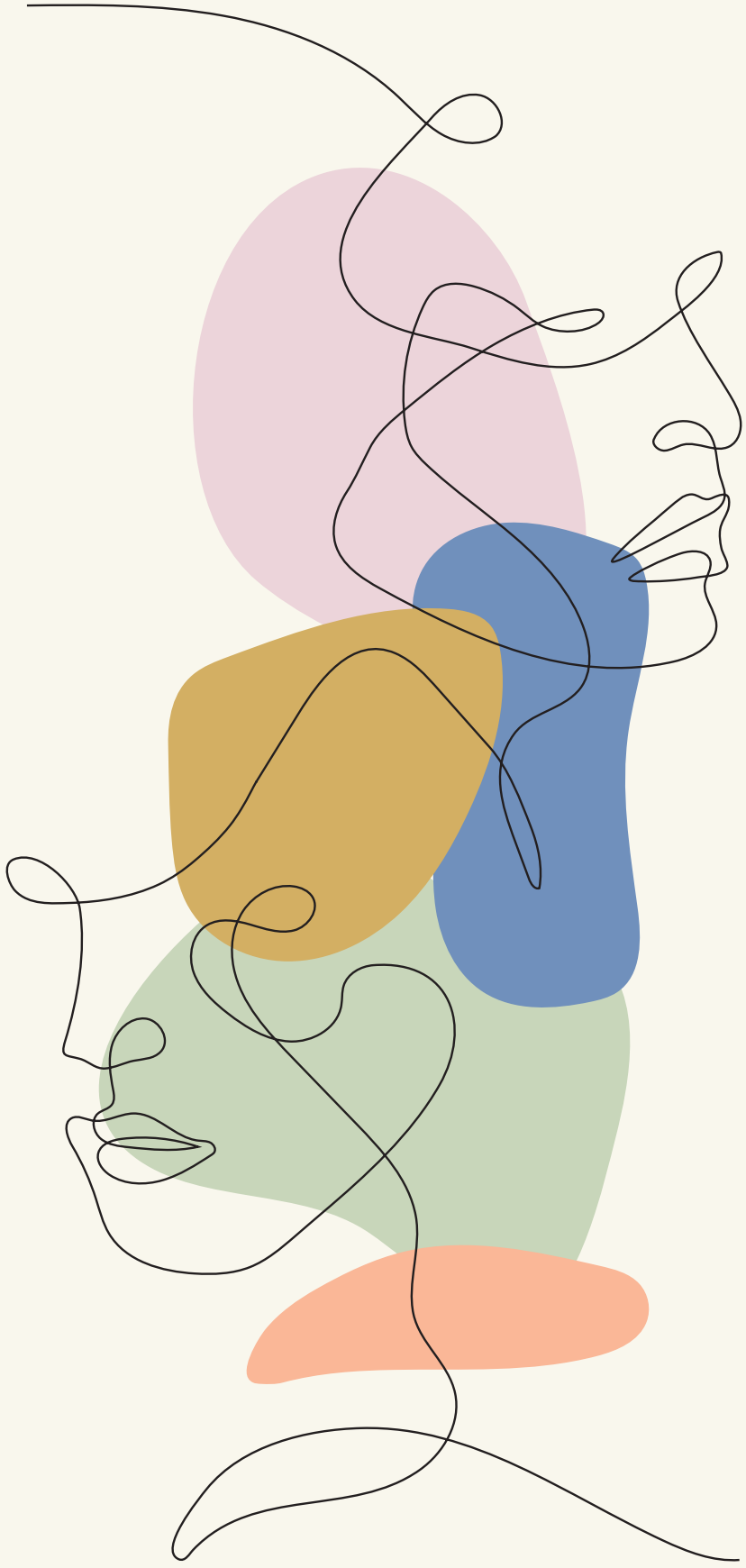
Supplementary Tables

Supplementary Table 1:
RNA-Seq analysis in DMSO and dmPGE$_2$ treated CD34$^+$ HSPCs

Supplementary Table 2:
pCREB ChIP-seq peaks in DMSO and dmPGE$_2$ treated CD34$^+$ HSPCs

Supplementary Table 3:
Active enhancers in DMSO and dmPGE$_2$ treated CD34$^+$ HSPCs

Supplementary tables are available upon request



6

The many faces of gene regulation: Extrinsic control of cell fate and function

Summarizing discussion

Chapter 6

The many faces of gene regulation: Extrinsic control of cell fate and function

Audrey Sporrij, Leonard I. Zon

Cell fate and function is determined by the complex interplay between extrinsic and intrinsic factors[1]. Environmental signals play important roles in the processes that underlie cell behavior. All cells have the ability to recognize extracellular molecules and can respond by altering the transcriptional programs that establish and maintain cell identity **(Chapter 1)**. While extrinsic stresses are known to affect gene regulation, exactly how these signals are integrated into highly specific transcriptional programs to drive certain cell fates remains poorly understood. The objective of the research presented here was to dissect the mechanisms through which environmental signals regulate gene expression and control cell fate.

Our work revealed that activation of intracellular signaling pathways by environmental molecules is directly received, interpreted, and answered to, by chromatin. This leads to transcriptional activation or repression. The chromatin landscape is highly dynamic and exceedingly reactive to external stress. We identified several mechanisms by which chromatin adapts to extrinsic signals and modifies gene expression (Figure 1).

Environmental conditions can be sensed by multifunctional proteins that masquerade as transcription factors (TFs) or transcriptional co-activators and convert the stress into a transcriptional response. We observed that nucleotide deficiency is interpreted by the RNA binding protein DDX21 that also functions as a chromatin factor to facilitate transcription elongation. Following stress sensing, DDX21 disengages from chromatin to enforce transcriptional pausing during suboptimal cellular conditions. Instead, DDX21 engages mRNAs to potentially stabilize transcripts that are required to restore homeostasis. A stress-related function for chromatin factors and RNA binding proteins could allow for the regulation of various cellular processes at once. We propose that the mechanism of chromatin factors functioning as stress sensors is evolutionarily conserved and suggest that other chromatin proteins with similar dual functions likely exist.

This research also found that the chromatin structure itself facilitates induction of stress response genes. The local nucleosome structure determines the responsiveness of individual regulatory elements to stress. We found that stress-inducible enhancers are pre-marked with histone variant H2A.Z-containing nucleosomes. These enhancer nucleosomes are not prohibitive of TF binding. Rather, the observed epigenetic characteristics enable rapid chromatin remodeling and quick, transient alterations in gene expression. The stimuli-responsive enhancer landscape is such that it can promptly receive, and respond to, signals from the extracellular environment. We hypothesize that the chromatin architecture of stimuli-responsive regulatory regions is organized at earlier stages of cell lineage specification. Precisely when and how the epigenetic pre-marking of inducible enhancers happens is currently unknown and requires further investigation.

Together, the work described here identified how extrinsic signals directly impact chromatin. Adaptive mechanisms such as multifunctional chromatin factors, epigenetic pre-marking of regulatory regions, and swift chromatin remodeling permit acute regulation of genes specific to cell fate. While our work specifically focused on transcriptional control

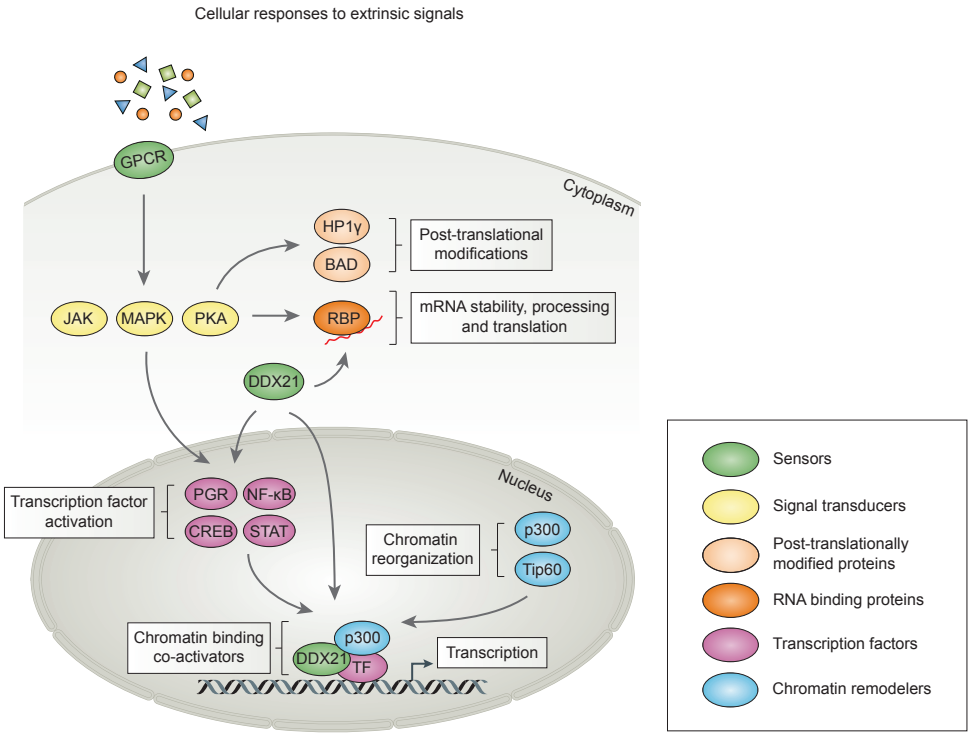


Figure 1. Gene regulation in response to environmental signals occurs at multiple levels. Different environmental stimuli, such as growth factors and inflammatory cytokines, are sensed by membrane bound or intracellular receptors (in green). These sensors can activate signaling transduction pathways or exert stress-related functions themselves. Intracellular sensors such as DDX21 (in green) can carry multiple functions and shift their role accordingly during stress. Membrane bound sensors typically activate protein kinases and other signal transducers (in yellow). Downstream targets of transducer proteins include transcription factors (in purple), such as CREB, NF-κB, and STATs, and chromatin regulators, such as p300 and Tip60 (in blue, not indicated with an arrow in this overview). Signal-transducing protein kinases are also involved in the control of gene expression at the post-transcriptional level by affecting mRNA processing and post-translational activity of proteins (in peach and orange).

of cells from the two most regenerative systems of the body, those being the skin and the blood, our findings may hold true for other cell types.

Extrinsic signals control neural crest cell fate

Neural crest cells (NCCs) are multipotent progenitor cells that give rise to a plethora of cell types. This includes the pigment producing cells of the skin known as melanocytes. Commitment to melanocyte fate is instructed by inductive, extrinsic signals that converge on transcription[2]. Dysregulation of the transcriptional programs that control melanocyte fate can transform normal melanocytes into malignant derivates[3]. Understanding precisely how environmental signals control gene expression provides opportunity for new cancer therapies.

Previous studies revealed that melanomas undergo transcriptional reprogramming and adopt a NCC-like fate during malignant transformation[4]. The DHODH inhibitor leflunomide leads to depletion of nucleotide pools. This impedes the expression of genes that govern NCC fate through suppression of transcription elongation[4]. As a result, both NCC development and melanoma formation is inhibited. The exact means through which leflunomide abrogates transcriptional elongation is currently unknown.

RNA helicase DDX21 controls cell fate during nucleotide stress

We studied the mechanism by which leflunomide-induced nucleotide depletion affects gene regulation. A reduction in nucleotide pools is sensed by the nucleolar RNA helicase DDX21 **(Chapter 2)**. DDX21 exerts functions in ribosomal RNA (rRNA) production and processing in the nucleolus[5,6]. DDX21 also acts as a regulator of Polymerase II (Pol II) mediated transcription in the nucleoplasm[7,8]. At promoters, DDX21 facilitates transcription via interactions with TFs and by facilitating P-TEFb release through binding with the 7SK small nuclear ribonucleoprotein (7SK snRNP) complex[8,9]. Thus, DDX21 has a dual primary function. The RNA helicase exerts roles in rRNA metabolism in the nucleolus, while it in parallel acts as a transcriptional regulator in the nucleoplasm[8-10]

We observed reduced promoter-bound DDX21 following DHODH-induced nucleotide depletion **(Chapter 2)**. Loss of DDX21 from chromatin may enforce transcriptional pausing when nucleotide pools are low to prevent high transcriptional activity during times of metabolic stress. Instead, DDX21 engages mRNAs in the nucleoplasm. Binding of DDX21 to mRNAs of genes involved in nucleotide metabolisms might indicates that DDX21 aims to minimize stress defects by stabilization of transcripts that are important for adaptation. While we suggest that DDX21 binding to mRNA affects transcript stability, functions of DDX21 in mRNA splicing or nuclear export remain also a possibility. Further research will have to resolve the exact role of DDX21 binding to mRNA and assess if this effect is specific to nucleotide depletion or part of a general response to stress.

We hypothesize that DDX21 senses changes in nucleotide levels. Low nucleotide pools direct DDX21 to its alternative, mRNA-related function in the nucleoplasm. This alternative function is concomitant with dissociation of DDX21 from rRNA along with its translocation from the nucleolus, as well as with disengagement from chromatin and subsequent defective transcription elongation. A role for DDX21 in both basic cellular functions and stress adaptation is in line with the dual functions hold by other stress-related proteins[11]. Stress mechanisms are shaped by positive natural selection to enhance an organism's ability to cope with, and adapt to, a changing environment. The increased repertoire of extrinsic signals that can be recognized by cells, and the diversification of intracellular response pathways, played a fundamental role in the evolutionary success of multicellular organisms[12]. A stress-related role of the already multifunctional DDX21 protein may allow cells to regulate various cellular processes at once. Affecting several intracellular systems

via one single protein could facilitate a quick and coordinated adaptive response. One remaining question is whether DDX21 exerts such function only in nucleotide stress conditions or whether other types of stress also evoke this function.

We were surprised to find that a reduction in DDX21 expression confers resistance to transcriptional defects resulting from nucleotide depletion **(Chapter 2)**. Decreased levels of the RNA helicase during nucleotide stress might redirect residual DDX21 towards its primary functions. The transcriptional role of DDX21 on chromatin and function in rRNA metabolisms in the nucleolus may be prioritized over mRNA binding when both nucleotide levels and DDX21 levels are low. This restores effective elongation and rescues cell fate *in vivo*.

As a sensor of nucleotide pools, RNA helicase DDX21 inhibits transcription when nucleotide levels are low and shifts its binding from chromatin and rRNA to mRNA. DDX21 is an enzyme with a multitude of functions that are likely highly interconnected and co-regulated. Exactly how DDX21 is directed towards certain functions in specific cellular contexts still remains to be elucidated.

Progesterone signaling affects DDX21 function

One of the factors that influences DDX21 function is progesterone and progesterone receptor (PGR) signaling. The PGR controls transcription by binding to progesterone response elements (PREs) near its response genes. Here, PGR facilitates transcription initiation and elongation through recruitment of the transcription machinery and p-TEFb, respectively[13-15]. PGR also affects the expression of genes that lack defined PREs via interactions with other TFs[16-18]. Furthermore, PGR can exert effects through nongenomic mechanisms by altering production of second messenger molecules and by influencing activity of transduction pathways, such as the MAPK pathway[19,20]. PGR was identified as a complex interaction partner of DDX21 **(Chapter 2)**. Although the implications of the interaction are currently unknown, DDX21 likely has a specific function when associated with PGR. Modification of progesterone signaling, and more specifically a decrease in PGR transcriptional activity, ameliorates the effects of nucleotide depletion and restores NCC fate. Downregulation of PGR could lead to reduced TF activity and hereby reduce the transcriptional demand in NCCs. This would alleviate the effects of low nucleotide abundance imposed by the DHODH inhibitor leflunomide.

We suggest that DDX21 acts as a mediator between nuclear hormone signaling and nucleotide stress signaling. In normal conditions, DDX21 exerts its primary functions in rRNA metabolism in the nucleolus and in Pol II-mediated transcription in the nucleoplasm[8-10]. During nucleotide stress, DDX21 enforces polymerase pausing and shifts is functions towards a mRNA regulatory role. However, alterations in progesterone signaling can drive DDX21 activity back towards its primary functions, even in conditions of low nucleotide levels. We hypothesize that changes in transcriptional demand,

due to reduced PGR activity, redirect DDX21 to chromatin. Loss of PGR activity could provide a second stress signal to DDX21 that overrules, or counteracts, stress resulting from nucleotide depletion. In an effort to compensate for the lost PGR activity, DDX21 shifts back towards its primary roles. As a result, effective transcription and cell fate are restored in NCCs. In conditions of competition between nucleotide stress signals and altered progesterone signaling, DDX21 activity is skewed toward regulation of transcription. This work revealed the intricate and highly complex regulation of gene programs during stress. It furthermore shows that activation of multiple signaling pathways by extrinsic molecules can provide the means to manipulate cell fate.

Gene regulation by environmental signals in hematopoietic cells

Adult humans require over a billion new blood cells every day. This task is carried out by the hematopoietic stem cells (HSCs) that resides within the bone marrow[21]. The self-renewal potential that characterizes HSCs and enables the production of blood cells throughout our lives, makes HSCs also uniquely susceptible to malignant transformation. Proper control of HSC fate therefore acts as a critical barrier to cancer. Because HSCs have the potential to reconstitute the entire hematopoietic system, they are clinically used for transplantation in patients with a variety of blood and immune disorders[22]. Hence, understanding regulators of HSC fate can not only drive development of novel cell-based therapies but also improve current treatment regimens.

Signal-responsive transcription factors govern gene regulation in HSCs

A variety of extrinsic niche factors affect HSC fate and function[23]. This includes signaling molecules, such as WNT and BMP, and inflammatory mediators, such as interferon (IFN) and prostaglandin E2 (PGE_2)[24-26]. The changes in HSC behavior mediated by a given environmental signal depend, for a significant part, on the signal-responsive TFs (STFs) and transcriptional programs that are affected by the stimuli **(Chapter 3)**.
The genomic landscape defined by lineage-specific master TFs (MTFs) primes chromatin for activation and directs STFs to modify gene expression in response to environmental factors[27]. We noted that activation and binding of the STF CREB to MTF-bound regulatory regions is concomitant with increased MTF occupancy at these loci **(Chapter 5)**. This demonstrates that STFs can modulate the binding pattern of MTFs. We hypothesize a model of bidirectional cooperativity between MTFs and STFs in hematopoietic stem and progenitor cells (HSPCs). MTFs direct STFs to specific locations in the genome, while STFs can stabilize MTF engagement at stimuli-inducible regions to reinforce expression of cell fate-specific genes. MTFs are known to perform as co-factors to other MTFs, but such properties are generally not attributed to non-pioneer factors[28]. We propose that co-factor functions are not confined to MTFs only but that STFs, such as CREB, can act as

co-factors for MTFs.
These findings support a model wherein interactions between TFs are highly dynamic. While MTFs define cell type-specific transcriptional states, binding of STFs links extracellular signals to cell fate-specific gene expression by altering MTF occupancy at regulatory loci. Although the effect was specifically observed in HSPCs following stimulation with 16,16-dimethyl PGE_2 ($dmPGE_2$), we suggest that this mechanism of cell fate reinforcement also holds for other extrinsic signals.

Chromatin dynamics in response to extrinsic signals

The epigenetic state of a cell determines receptiveness to certain signals. Structurally different regions of chromatin can be either permissive or prohibitive of transcription[29]. We noted increased enhancer chromatin accessibility after exposure of HSPCs to the inflammatory mediator $dmPGE_2$. Despite gene expression changes being predominantly transient following stimulation **(Chapter 3)**, alterations in chromatin accessibility may remain. Other extrinsic signals, such as the infectious agent LPS, are known to induce persistent changes in the epigenetic landscape of enhancers[30,31].
Removal of nucleosomes plays a key role in the activation of gene expression. Recent studies, however, also support a model where nucleosomes are not evicted but rather transiently destabilized in active regulatory regions[32,33]. This was particularly observed at enhancers that control the expression of genes that are responsive to environmental stresses[34,35]. We found retention of enhancer nucleosomes during acute transcriptional induction of HSPCs **(Chapter 5)**. Rapid changes in nucleosome accessibility, but not occupancy, at enhancers defines gene induction during the inflammatory response to $dmPGE_2$. Destabilization, rather than eviction, of nucleosomes might be a more generalizable model of the response to extrinsic signals. Thus, the local nucleosome structure is responsive to stress.
Mechanisms to alter the accessibility of DNA on the surface of nucleosomes include the incorporation of histone variants[36] and destabilization of DNA-nucleosome interactions through acetylation of lysine residues in histone tails[37]. We demonstrate that nucleosomes in inducible regulatory regions are pre-marked by histone variants and destabilized through acetylation following stimulation **(Chapter 5)**. Specifically, H2A.Z-variant nucleosomes are found, and retained, at stimuli-responsive enhancers in HSPCs. Histone variant H2A.Z plays a role in the activation of genes that maintain self-renewal and differentiation potential in stem cells[38,39]. Pre-marking of chromatin with labile H2A.Z-variant nucleosomes prior to stimulation may ensure signal receptiveness of inducible regulatory regions. Precisely when and how the epigenetic pre-marking of enhancers occurs requires further investigation.
The functions of H2A.Z vary based on its post-translational modifications. Histone acetylation is critical for the maintenance of HSCs[40]. H2A.Z acetylation confers an open

chromatin conformation and in concomitant with gene activation[41,42]. Acetylation of labile H2A.Z-variant nucleosomes at enhancers seen upon signaling might create a chromatin structure with sufficient accessibility for certain chromatin factors to bind **(Chapter 5)**. Immediate eviction of the accessible nucleosomes may not be required for enhancer activation. This mechanism makes gene regulation following stimulation swift and energetically favorable for cells, as nucleosome remodelers are ATP-dependent but histone modifiers are not.

H2A.Z can be acetylated by the histone acetyltransferases (HATs) Tip60 and p300[43]. These HATs are essential for HSC fate by controlling genes pivotal for balanced self-renewal, proliferation and differentiation[44,45]. Both Tip60 and p300 are complex interaction partners of the STF CREB[46,47]. Therefore, we suggest that remodeling of retained, H2A.Z-variant nucleosomes during stress signaling is a result of STF binding at enhancers. Accessible nucleosomes are not necessarily prohibitive to TF binding but rather might facilitate, and stabilize, the binding of trans-acting factors[33].

The pre-defined chromatin architecture of inducible regulatory regions directly determines signal responsiveness and facilitates induction of the transcriptional programs required for adaptation of cell fate.

Intrinsic HSC heterogeneity may predispose for specific signaling responses

Cells can respond heterogeneously to extracellular signals, but the complexity of the heterogeneity nor its contribution to different outcomes in cell fate are well understood[48]. We studied the heterogeneity of HSCs at steady state and after stimulation with niche signals. HSCs reside in continuous transcriptional states during normal homeostasis. These HSC states are continuous, rather than distinct, cellular subpopulations **(Chapter 4)**. Niche signals can induce transitions between transcriptional states of HSCs and allow for surfacing of novel transcriptional states that were previously not represented within the population.

We found that responses to extrinsic signals between HSCs and hematopoietic progenitor cells (HPCs) are highly similar. However, specific changes within transcriptional states of HSCs exist that are not observed in HPCs. The transcriptional state being prominently affected by various niche signals in HSCs, but not HPCs, relates to the cell cycle. Control of proliferation is crucial in HSCs to prevent premature exhaustion[49]. Regulation of the cell cycle related genes may ensure maintenance of HSC function in changing environmental conditions. Cell fate depends on a complex interplay between the transcriptional and epigenetic state, which are both dynamically altered in response to extrinsic factors. A new study linking gene expression to hematopoietic cell fate using lineage tracing methods identified that single cell transcriptional states do not solely define cell fate[50]. We demonstrated that HSCs not only reside in fluent transcriptional states but also present in heterogenous epigenetic states. Based on their chromatin organization, HSCs in different

transcriptional states have an equal opportunity to receive, and respond to, certain niche signals **(Chapter 4)**. Diversity in the functional outcome of a signal could be driven by epigenetic differences of regulatory regions involved in the secondary response to the signal. The heterogenic effects on cell fate might thus be due to epigenetic variation in accessibility of transcription factor motifs that serve later in the response to niche signals. The combinations of a specific transcriptional and epigenetic state may underlie distinct functional outcomes and can be important determinants of the fate and function of HSCs and likely also other cell types.

Towards a better understanding of gene regulation through signaling

Regardless of heterogeneity in the response of cells to any given signal, many different extrinsic signals have a similar overall effect on the behavior of cells from a certain population[51]. Since distinct signals can lead to an identical functional outcome, one might hypothesize that the different signals converge on the same set of genomic regions. These regulatory elements are inducible by a variety of factors and drive the expression of core cell fate genes. We tentatively named genomic regions where STFs engage with MTFs 'Transcriptional Signaling Centers' (TSCs) (Figure 2A). These TSCs could be activated or repressed by extrinsic molecules to regulate gene expression and control cell behavior. Comparing the mechanisms through which different signals control similar transcriptional programs could provide a unifying model to promote and maintain cell fate.

Extrinsic signals converge on cell fate-specific regulatory elements

To test whether indeed different signals convergence on cell fate-specific regulatory regions, one needs to perform extensive profiling of the transcriptional and epigenetic landscape after stimulation with individual signals and determine common versus signal-specific changes. Overlapping the genome-wide occupancy of STFs activated by distinct environmental molecules will reveal if, and where, different signals converge in the genome. This could identify individual and shared roles of STFs in the control of cell fate.
The signaling molecules WNT, BMP, TGF-β, and dmPGE$_2$ act as regulators of HSPCs. Their downstream STFs bind near sites occupied by the MTFs GATA2 and PU.1 following stimulation[24] **(Chapter 5)**. Preliminary analysis showed that regions bound by CREB after exposure to dmPGE$_2$ can also gain binding of other STFs upon activation by the corresponding signaling molecules (Figure 2B). Such TSCs are localized near genes critical for HSPC fate[52].
Combining these comparative genomic and epigenomic studies with machine learning approaches will enable us to study the logic of TSCs. Assessing whether there is conservation of specific TF motifs, how motifs are positioned within TSCs, and where TSCs are localized in the genome in relation to non-responsive regulatory elements will

advance our understanding of their role in cell fate control[53]. In-depth analysis of such enhancers could provide a mechanistic model for cell fate control. The distinct regulatory regions where signals converge may be one of the defining elements of HSPC fate and function. Deciphering the mechanisms that regulate the activity of TSCs and the corresponding genes might provide novel insights on cell fate.

Unique cell fate-specific chromatin complexes at signaling centers

It is great interest to compare the composition of chromatin complexes found at TSCs prior to and following stimulation with extrinsic signals. This assessment could identify novel protein interactions that are only observed upon activation of specific signaling pathways and their respective STFs. Furthermore, comparing enhancer-associated protein complexes between different signals may identify common factors that play crucial roles in the reinforcement of HSPC fate (Figure 3). Finally, it could be important to dissect the differences in protein complexes associated with TSC versus proteins observed at non-responsive, ubiquitous regulatory regions. Given our observations that proteins can masquerade as TFs and transcriptional activators, the potential importance of factors not transitionally associated with gene regulation, such as RNA binding proteins, should not be overlooked.

To identify the proteins associated with TSCs specifically, endogenous complex immunoprecipitation studies should be performed in a locus-specific manner. A novel pulldown technique might provide an elegant way to approach this question. This method, called CRISPR affinity purification in situ of regulatory elements (CAPTURE), allows for the analysis of chromatin complexes at specific, endogenous loci. It uses a biotinylated

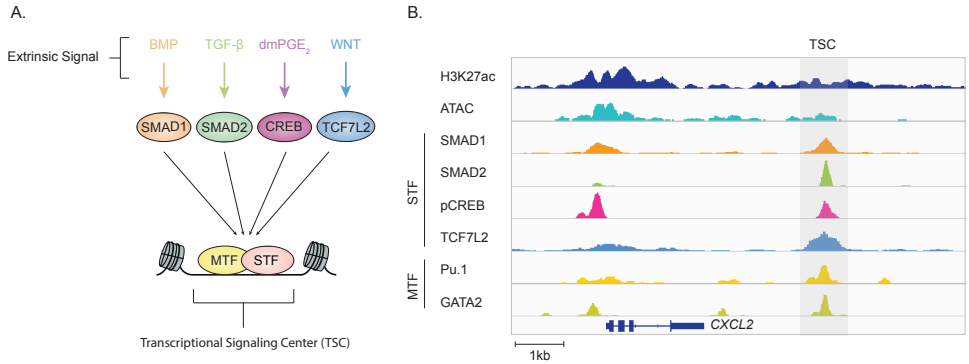


Figure 2. Transcriptional signaling centers control expression of cell fate-specific genes. (A) Signaling transcription factors (STFs) converge on regulatory regions occupied by master transcription factors (MTFs) known as 'Transcriptional Signaling Centers' (TSCs). (B) Overlap of ATAC-Seq and ChIP-Seq data for the STFs CREB (cAMP signaling), SMAD1 (BMP signaling), SMAD2 (TGF-β signaling) and TCF7L2 (Wnt signaling) following activation with their respective signals. STF binding overlaps with other STFs and with MTFs of the hematopoietic stem and progenitor cell lineage GATA2 and PU.1 near the cell fate-specific gene *CXCL2*. The region identified as a TSC is marked in light grey.

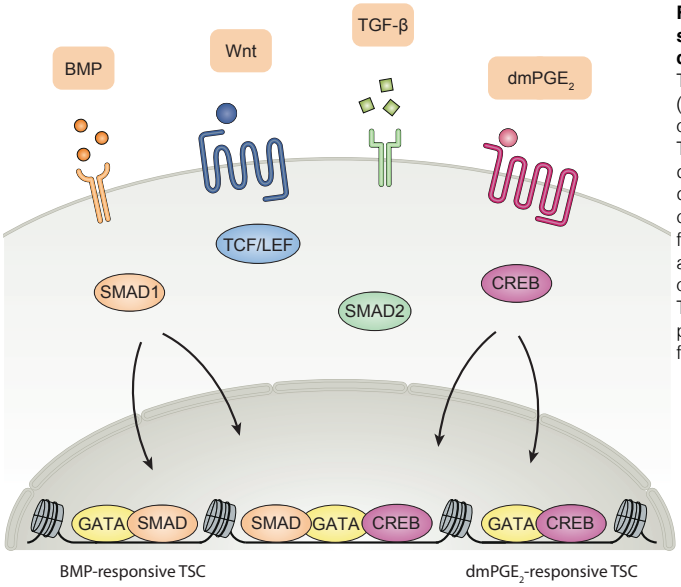


Figure 3. Distinct transcriptional signaling centers may hold different functional properties. Transcriptional signaling centers (TSCs) receptive to a wide variety of signals (here indicated as TSC responsive to both BMP or $dmPGE_2$ signaling) could hold different functional properties compared to those receptive to fewer signals (here indicated as a TSC responsive to either BMP or $dmPGE_2$ signaling). Different TSCs might associate with different protein complexes and chromatin factors

nuclease-deficient Cas9 (dCas9) protein that is guided to a genomic region of interest by a sequence-specific single guide RNA (sgRNA)[54]. Streptavidin-mediated pulldown will then isolate the dCas9-tethered locus together with the associated protein complexes. Comparison of the protein complex composition at TSCs could highlight factors that are required for regulation of cell fate-specific transcriptional programs.

Cell fate control via transcriptional signaling centers

Upon identification of TSCs, their contribution and requirement for cell fate should be carefully assessed. Different TSCs may hold distinct functional properties. This could be assessed through perturbation studies. The importance of lineage-specific MTFs at regulatory elements is widely appreciated. However, the contribution of STFs to cell fate is less well understood. To understand the functional consequence of STF TSCs, one can genetically perturb STF binding sites in such genomic loci using CRISPR-Cas9. One can then determine the effect of TSCs as well as distinct STFs on gene expression and the functional properties of cells. In addition, mutation of protein-interacting domains of STFs and other identified chromatin factors will alter the composition of TSC-associated complexes. This will help understand the contribution of a given factor within a TSC. Together, these studies could provide insights on the mechanism through which extrinsic signals control gene expression and cellular behavior.

The clinical importance of insights on extrinsic control of cell fate

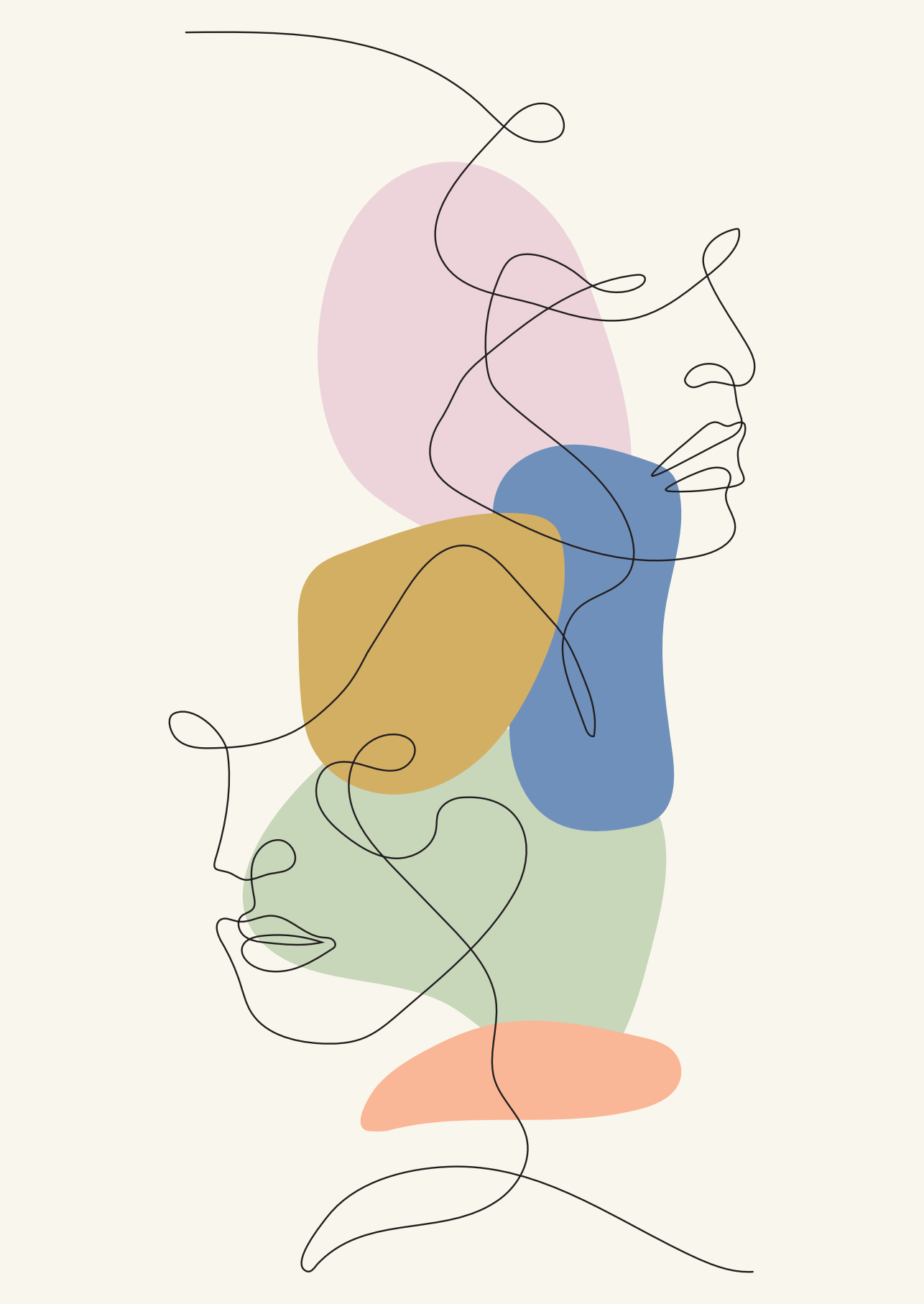
The work presented here strived to elucidate the fundamental principles by which extrinsic signals regulate transcriptional changes. While many questions remain, the research presented here revealed the complex dynamics and the multitude of layers at which gene expression and cell function is regulated in different environmental contexts.
A better understanding of these biological processes has the potential to directly impact treatment regimens[55]. Our current strategies to manipulate stem and progenitor cells for therapeutic applications are limited by our inability to precisely control cell fate. A proper understanding of the mechanisms that regulate gene expression in response to environmental signals could allow for beneficial modulation of cell fate and function. Controlling cell fate becomes especially important as (stem) cells increasingly form the foundation of novel interventions, such as cell-based therapies. Regulation and maintenance of a specific cell fate is immediately tied to success of these therapies. Moreover, one could exploit this understanding to confer, or reinforce, certain cellular functions using *ex vivo* manipulation methods.
These studies highlight the clinical potential evoked by proper understanding of cell fate control by extrinsic signals. While tremendous clinical progress has been made in using extrinsic signals to manipulate cell fate, conditions that do not compromise long-term cell function remain unknown. This clearly illustrates the continued need to uncover the many faces of gene regulation by extrinsic signals to control cell fate.

References

1. Rizo, A., Vellenga, E., de Haan, G. & Schuringa, J. J. Signaling pathways in self-renewing hematopoietic and leukemic stem cells: do all stem cells need a niche? Hum. Mol. Genet. 15 Spec No 2, R210–9 (2006).
2. Takeda, K. et al. Induction of melanocyte-specific microphthalmia-associated transcription factor by Wnt-3a. J. Biol. Chem. 275, 14013–14016 (2000).
3. Shain, A. H. & Bastian, B. C. From melanocytes to melanomas. Nat Rev Cancer 16, 345–358 (2016).
4. White, R. M. et al. DHODH modulates transcriptional elongation in the neural crest and melanoma. Nature 471, 518–522 (2011).
5. Henning, D., So, R. B., Jin, R., Lau, L. F. & Valdez, B. C. Silencing of RNA helicase II/Gualpha inhibits mammalian ribosomal RNA production. J. Biol. Chem. 278, 52307–52314 (2003).
6. Yang, H. et al. Down-regulation of RNA helicase II/Gu results in the depletion of 18 and 28 S rRNAs in Xenopus oocyte. J. Biol. Chem. 278, 38847–38859 (2003).
7. Westermarck, J. et al. The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. EMBO J. 21, 451–460 (2002).
8. Calo, E. et al. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature 518, 249–253 (2015).
9. Calo, E. et al. Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. Nature 554, 112–117 (2018).
10. Xing, Y.-H. et al. SLERT Regulates DDX21 Rings Associated with Pol I Transcription. Cell 169, 664–678.e16 (2017).
11. Kültz, D. Evolution of the cellular stress proteome: from monophyletic origin to ubiquitous function. J Exp Biol 206, 3119–3124 (2003).
12. Weake, V. M. & Workman, J. L. Inducible gene expression: diverse regulatory mechanisms. Nat. Rev. Genet. 11, 426–437 (2010).
13. Jacobsen, B. M. & Horwitz, K. B. Progesterone receptors, their isoforms and progesterone regulated transcription. Mol Cell Endocrinol 357, 18–29 (2012).
14. Bertucci, P. Y. et al. Progesterone receptor induces bcl-x expression through intragenic binding sites favoring RNA polymerase II elongation. Nucleic Acids Res. 41, 6072–6086 (2013).
15. Kininis, M., Isaacs, G. D., Core, L. J., Hah, N. & Kraus, W. L. Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes. Mol. Cell. Biol. 29, 1123–1133 (2009).
16. Clarke, C. L. & Graham, J. D. Non-overlapping progesterone receptor cistromes contribute to cell-specific transcriptional outcomes. PLoS ONE 7, e35859 (2012).
17. Mohammed, H. et al. Progesterone receptor modulates ERα action in breast cancer. Nature 523, 313–317 (2015).
18. Bamberger, A. M., Bamberger, C. M., Gellersen, B. & Schulte, H. M. Modulation of AP-1 activity by the human progesterone receptor in endometrial adenocarcinoma cells. Proceedings of the National Academy of Sciences 93, 6169–6174 (1996).
19. Bayaa, M., Booth, R. A., Sheng, Y. & Liu, X. J. The classical progesterone receptor mediates Xenopus oocyte maturation through a nongenomic mechanism. Proceedings of the National Academy of Sciences 97, 12607–12612 (2000).
20. Migliaccio, A. et al. Activation of the Src/p21ras/Erk pathway by progesterone receptor via cross-talk with estrogen receptor. EMBO J. 17, 2008–2018 (1998).
21. Pinho, S. & Frenette, P. S. Haematopoietic stem cell activity and interactions with the niche. Nature Reviews Molecular Cell Biology 20, 303–320 (2019).
22. Jenq, R. R. & van den Brink, M. R. M. Allogeneic haematopoietic stem cell transplantation: individualized stem cell and immune therapy of cancer. Nat Rev Cancer 10, 213–221 (2010).
23. Morrison, S. J. & Scadden, D. T. The bone marrow niche for haematopoietic stem cells. Nature 505, 327–334 (2014).
24. Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell 147, 577–589 (2011).
25. North, T. E. et al. Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis.

Nature 447, 1007–1011 (2007).
26. Li, P. et al. Epoxyeicosatrienoic acids enhance embryonic haematopoiesis and adult marrow engraftment. Nature 523, 468–471 (2015).
27. Mullen, A. C. et al. Master transcription factors determine cell-type-specific responses to TGF-β signaling. Cell 147, 565–576 (2011).
28. Donaghey, J. et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. Nat. Genet. 50, 250–258 (2018).
29. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. 20, 207–220 (2019).
30. de Laval, B. et al. C/EBPβ-Dependent Epigenetic Memory Induces Trained Immunity in Hematopoietic Stem Cells. Cell Stem Cell 26, 793 (2020).
31. Ostuni, R. et al. Latent Enhancers Activated by Stimulation in Differentiated Cells. Cell 152, 157–171 (2013).
32. Iwafuchi-Doi, M. et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol. Cell 62, 79–91 (2016).
33. Ballaré, C. et al. Nucleosome-driven transcription factor binding and gene regulation. Mol. Cell 49, 67–79 (2013).
34. Sexton, B. S., Druliner, B. R., Avey, D., Zhu, F. & Dennis, J. H. Changes in nucleosome occupancy occur in a chromosome specific manner. Genom Data 2, 114–116 (2014).
35. Knight, B. et al. Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. Genes & Development 28, 1695–1709 (2014).
36. Jin, C. et al. H3.3/H2A.Z double variant–containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat. Genet. 41, 941–945 (2009).
37. Garcia-Ramirez, M., Rocchini, C. & Ausio, J. Modulation of chromatin folding by histone acetylation. J. Biol. Chem. 270, 17923–17928 (1995).
38. Li, Z. et al. Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell 151, 1608–1616 (2012).
39. Hu, G. et al. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell Stem Cell 12, 180–192 (2013).
40. Sun, X.-J., Man, N., Tan, Y., Nimer, S. D. & Wang, L. The Role of Histone Acetyltransferases in Normal and Malignant Hematopoiesis. Front Oncol 5, 108 (2015).
41. Adam, M., Robert, F., Larochelle, M. & Gaudreau, L. H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol. Cell. Biol. 21, 6270–6279 (2001).
42. Gévry, N. et al. Histone H2A.Z is essential for estrogen receptor signaling. Genes & Development 23, 1522–1533 (2009).
43. Colino-Sanguino, Y. et al. A Read/Write Mechanism Connects p300 Bromodomain Function to H2A.Z Acetylation. iScience 21, 773–788 (2019).
44. Numata, A. et al. Lysine acetyltransferase Tip60 is required for hematopoietic stem cell maintenance. Blood 136, 1735–1747 (2020).
45. Rebel, V. I. et al. Distinct roles for CREB-binding protein and p300 in hematopoietic stem cell self-renewal. Proceedings of the National Academy of Sciences 99, 14789–14794 (2002).
46. Gavaravarapu, S. & Kamine, J. Tip60 inhibits activation of CREB protein by protein kinase A. Biochem. Biophys. Res. Commun. 269, 758–766 (2000).
47. Vo, N. & Goodman, R. H. CREB-binding protein and p300 in transcriptional regulation. J. Biol. Chem. 276, 13505–13508 (2001).
48. Altschuler, S. J. & Wu, L. F. Cellular heterogeneity: do differences make a difference? Cell 141, 559–563 (2010).
49. Trumpp, A., Essers, M. & Wilson, A. Awakening dormant haematopoietic stem cells. Nat Rev Immunol 10, 201–209 (2010).
50. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science 367, eaaw3381 (2020).
51. Orkin, S. H. & Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell 132,

631–644 (2008).
52. Sinclair, A. et al. CXCR2 and CXCL4 regulate survival and self-renewal of hematopoietic stem/progenitor cells. Blood 128, 371–383 (2016).
53. Minnoye, L. et al. Cross-species analysis of enhancer logic using deep learning. Genome Research gr.260844.120 (2020). doi:10.1101/gr.260844.120
54. Liu, X. et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell 170, 1028–1043.e19 (2017).
55. Wilkinson, A. C., Igarashi, K. J. & Nakauchi, H. Haematopoietic stem cell self-renewal in vivo and ex vivo. Nat. Rev. Genet. 21, 541–554 (2020).



A

Addenda

Nederlandse samenvatting

Acknowledgements

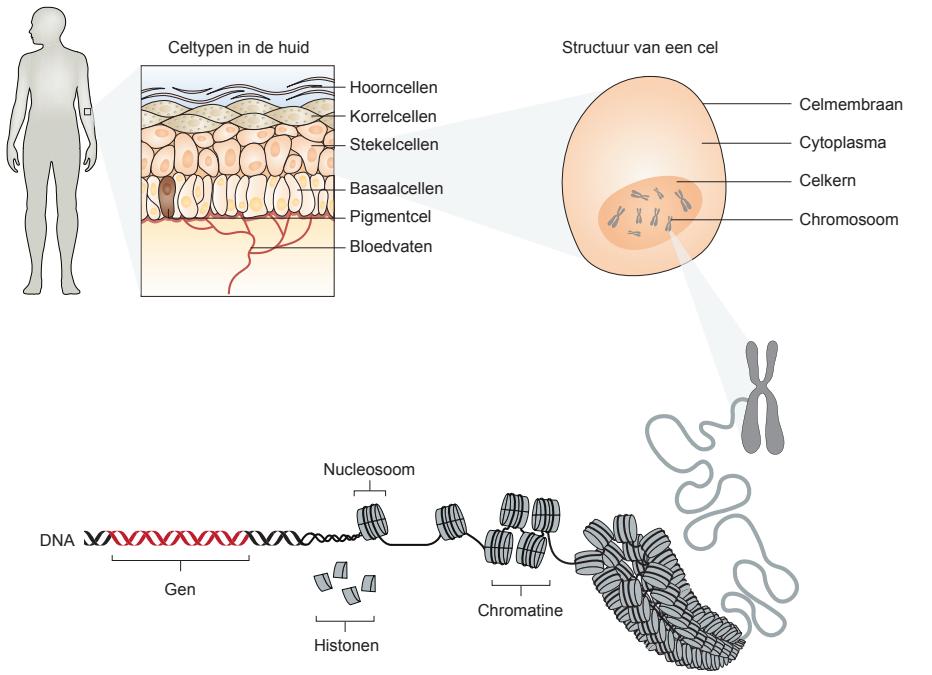
List of publications

About the author

Nederlandse samenvatting

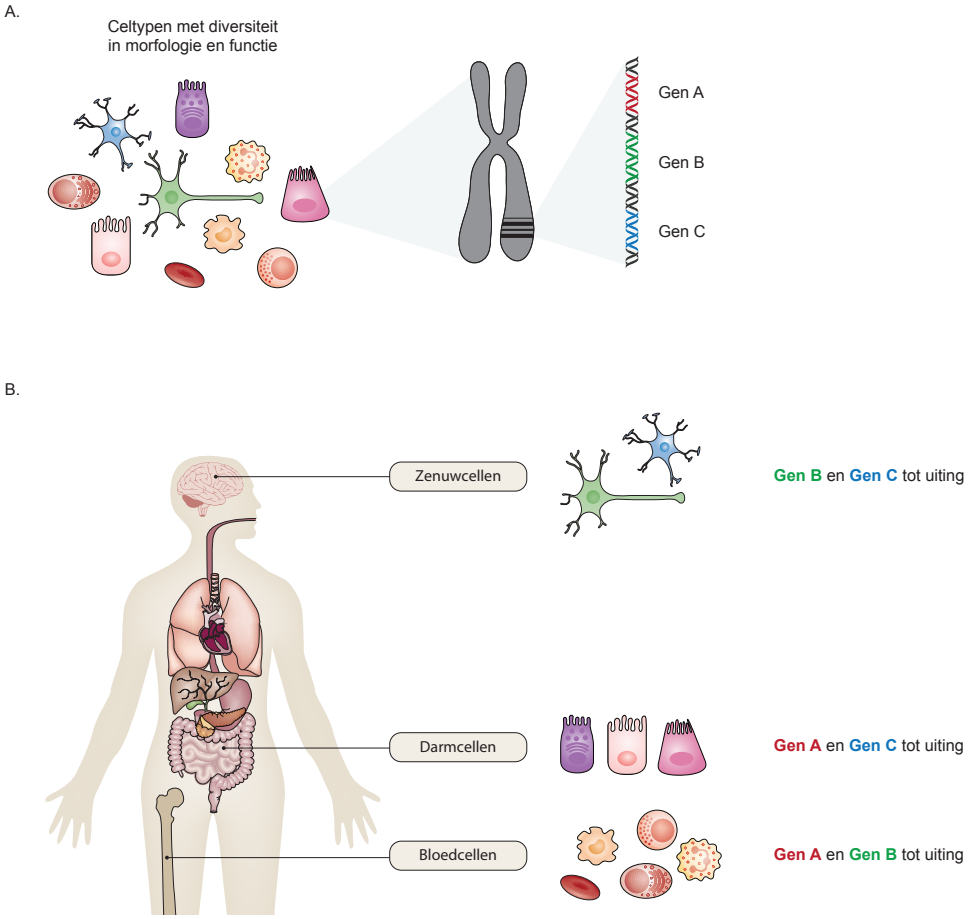
Introductie

Elke cel in ons lichaam bevat, onder normale omstandigheden, 46 chromosomen. Deze chromosomen bestaan ieder uit één heel lang dubbelstrengs DNA-molecuul. Deze 46 chromosomen bevatten samen wel 2 meter aan DNA, welke allemaal in de kern van een cel moet passen. De celkern heeft echter maar een diameter van rond de 10µm (micrometer, een miljoenste deel van een meter). Om ervoor te zorgen dat al het erfelijke materiaal in onze cellen past, worden de DNA-strengen om chromosoom-eiwitten (histonen) heen gevouwen. Acht van deze histonen vormen samen een eiwitbolletje waaromheen DNA wordt gewonden. Het complex van histonen met daaromheen gewonden dubbelstrengs DNA vormt een nucleosoom (Figuur 1). Een reeks van nucleosomen leidt tot een kralenstructuur genaamd chromatine. De nucleosomen worden vervolgens dichter bij elkaar gebonden tot een vezelachtige structuur. Dit zorgt voor een verdere compactheid. Uiteindelijk leidt een hoger orde opbinding van deze vezels tot gecondenseerde chromosomen, de meeste compacte vorm van chromatine.



Figuur 1. Compactheid van DNA door een complex met eiwitten in de celkern van cellen. DNA wordt om histoneiwitten heen gevouwen en vormen samen een nucleosoom. Een reeks van nucleosomen leidt tot een kralenstructuur genaamd chromatine. De nucleosomen worden vervolgens dichter bij elkaar gebonden tot een vezelachtige structuur. Een hogere orde opbinding van deze vezels leidt tot gecondenseerde chromosomen, de meeste compacte vorm van chromatine.

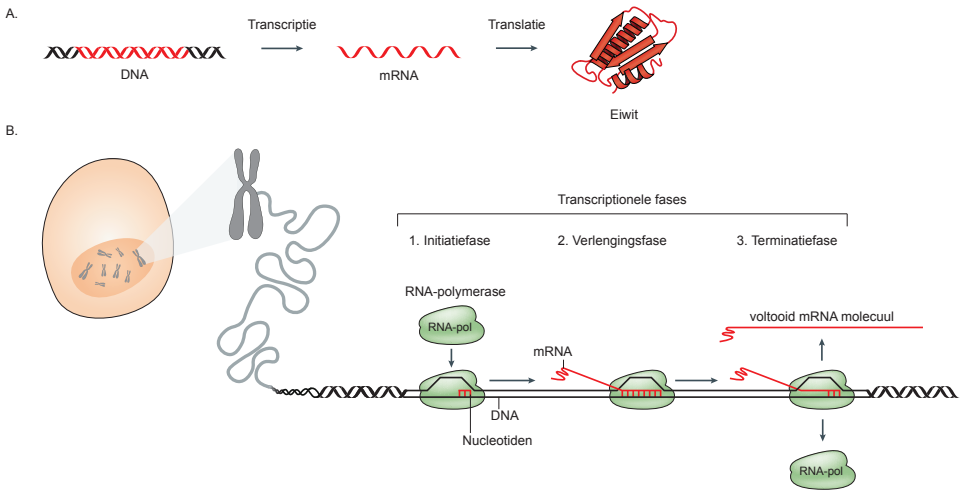
Het DNA fungeert als de drager van ons erfelijk materiaal, oftewel onze genen. Ondanks dat in elke cel hetzelfde genetische materiaal aanwezig is, zijn er veel verschillen in de karakteristieken en functies van cellen (Figuur 2A). Dit is het gevolg van verschillen in de uiting, of expressie, van genen. Zo komen andere genen tot expressie in cellen van de darmen dan de genen die zich uiten in zenuwcellen (Figuur 2B). Een gen komt tot uiting door het uitlezen, overschrijven en vertalen van DNA. Het uitlezen en overschrijven van de DNA-code wordt uitgevoerd door een speciaal enzym genaamd RNA-polymerase. RNA-polymerase leest het DNA af in de juist volgorde en vertaalt de boodschap naar mRNA (messenger-RNA, oftewel boodschapper-RNA). Het biologisch proces van het overschrijven van DNA naar mRNA heet transcriptie. Het mRNA kan vervolgens vertaald worden naar een functioneel eiwit door een proces genaamd translatie (Figuur 3).



Figuur 2. De uiting van genen bepaalt de functie en karakteristieken van cellen. (A) Er is een grote diversiteit aan celtypen. Ondanks de zichtbare verschillen tussen cellen, bevatten zij allemaal hetzelfde genetische materiaal. (B) Verschillen in de uiting van genen leidt tot cellen met unieke eigenschappen. De genen die tot expressie komen in de cellen van de darmen zijn anders dan de genen die zich uiten in zenuwcellen.

De regulatie van genexpressie in een veranderde omgeving

Het reguleren van de genexpressie maakt het mogelijk voor cellen om verschillende eiwitten te produceren en hiermee de eigenschappen en de functie van een cel te bepalen. Dit gebeurt niet alleen tijdens de ontwikkeling van een organisme waarbij bepaalde celtypen gevormd moeten worden, maar ook op andere momenten wanneer dit nodig is gedurende de levensduur van een cel. Welke genen tot expressie komen en in welke mate kan op verschillende manieren beïnvloed worden. Over het algemeen wordt de expressie van genen zeer nauwkeurig gereguleerd. Dit is om te voorkomen dat cellen de verkeerde eigenschappen of functies aannemen, zoals vaak kenmerkend is voor ziekten zoals kanker. Hierbij is de expressie van onder andere de genen die de celdeling reguleren ernstig verstoord. Als gevolg van een defecte regulatie van deze genen, blijven deze cellen zich ongelimiteerd delen.
Signalen uit de omgeving (de niche) van cellen zijn belangrijke indicatoren voor de regulatie van genexpressie. Zo is het bijvoorbeeld belangrijk dat cellen uit de alvleesklier reageren of signaleren gerelateerd aan de bloedsuikerwaardes, zodat zij op het juiste moment de genen aanzetten die betrokken zijn bij de insulineproductie. Ook bloedcellen moeten snel kunnen reageren op signalen uit de omgeving. Denk hierbij bijvoorbeeld aan de immuuncellen die hun gedrag moeten aanpassen wanneer men een infectie oploopt om deze snel onder controle te krijgen. Cellen hebben elegante mechanismen ontwikkeld om omgevingsfactoren waar te nemen, door te geven aan de celkern, en

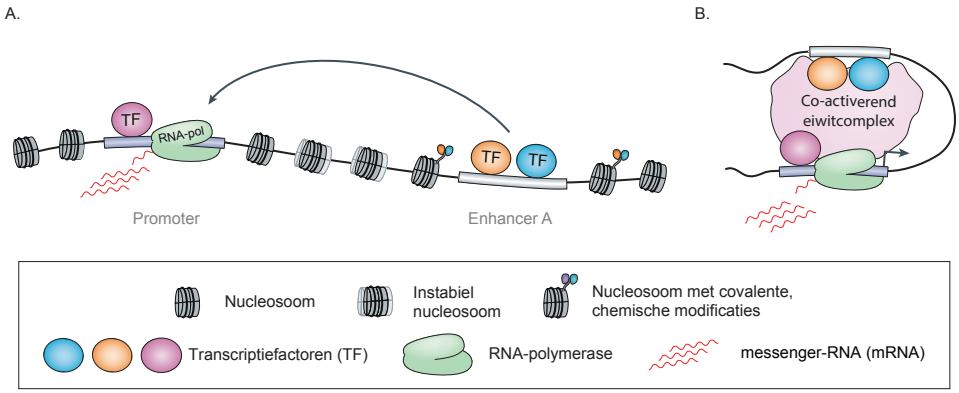


Figuur 3. Genen komen tot uiting door transcriptie- en translatieprocessen. Een gen komt tot uiting door het uitlezen, overschrijven en vertalen van DNA. (A) Tijdens transcriptie wordt de DNA-code uitgelezen en omgeschreven naar mRNA (messenger-RNA, oftewel boodschapper-RNA). Dit mRNA wordt vervolgens vertaald naar een eiwit tijdens de translatie. (B) Transcriptie bestaat uit verschillende fases. Tijdens de initiatiefase bindt RNA-polymerase aan het DNA en begint de transcriptie. Na de aanvang van transcriptie volgt er een verlengingsfase waarbij het mRNA-molecuul groeit naarmate RNA-polymerase meer DNA afleest. Wanneer het gen volledig is afgelezen en omgeschreven naar mRNA, volgt de terminatiefase waarbij RNA-polymerase loslaat van het DNA en de mRNA molecuul vrijkomt voor translatie.

om de genexpressie aan te passen. Deze veranderingen in genexpressie kunnen erg snel zijn en soms al worden waargenomen binnen enkele minuten na blootstelling aan bepaalde omgevingsstressen. De acute aanpassing is cruciaal voor maximale overleving van cellen in veranderende omgevingsomstandigheden. De eigenschap van cellen om zich te kunnen aanpassen aan signalen uit het milieu vormt een van de belangrijkste drijfkrachten van ons evolutionair succes. Desalniettemin, draagt het ook bij aan vatbaarheid voor ziekten. Defectieve reacties op signalen uit de omgeving leidt tot ontstemming van het gedrag van cellen en draagt bij aan bijvoorbeeld veroudering en de ontwikkeling van ziekten zoals kanker. Het blootleggen van de exacte mechanismen die de veranderingen in genexpressie reguleren ten gevolge van blootstelling aan bepaalde omgevingssignalen, biedt daarom nieuwe mogelijkheden voor klinische therapieën.

De aanpassing van genexpressie hangt niet alleen af van het type en de hoogte van het signaal, maar ook van het type cel dat wordt getroffen. Veranderingen in genexpressie worden op verschillende niveaus gereguleerd. Zo kunnen cellen de structuur van chromatine, oftewel de 'opwinding' van het DNA, veranderen. Door deze strakker samen te winden of juist losser te maken, worden bepaalde genen minder makkelijk of juist makkelijker toegankelijk voor factoren, zoals RNA-polymerase, die betrokken zijn bij transcriptie. Dit kan onder andere worden bereikt door het verwijderen van histonen. Bepaalde eiwitcomplexen kunnen histonen verwijderen uit een nucleosoom. Hierdoor vervalt deze structuur en komt het DNA dat om het histonencomplex heen gewonden zat bloot te liggen. Dit DNA is nu vrij toegankelijk voor factoren die betrokken zijn bij de transcriptie van genen en dus onze genen tot uiting brengen. Ook kunnen er structurele veranderingen worden aangebracht aan de histonen waaruit een nucleosoom bestaat. Deze covalente, chemische modificaties kunnen de elektrostatische lading van de histonen in een nucleosoom veranderen. Als gevolg hiervan verandert de aantrekking tussen de histoneiwitten en het negatief geladen DNA. Een minder grote aantrekking, of zelfs afstoting, leidt tot toegankelijkheid van het DNA.

Ons DNA codeert niet enkel genen, maar bevat ook elementen die de uiting van een gen reguleren. Naast de genetische informatie voor een gen, moet er namelijk ook gezorgd worden dat de uiting van het gen kan worden bestuurd. Deze elementen, welke de expressie van een gen controleren, worden regulatie-elementen genoemd. Er zijn verschillende typen elementen. Zo kan een 'promotor' de transcriptie van een gen aan- of uitzetten, kunnen 'silencers' de transcriptie van een gen afremmen en 'enhancers' de transcriptie juist bevorderen (Figuur 4). Signalen uit de omgeving kunnen bepaalde eiwitten genaamd transcriptiefactoren activeren die gericht binden aan specifieke cis-regulatie-elementen in het DNA en daarbij hun functie bevorderen. Op die manier kan er direct invloed worden uitgeoefend op de expressie van een gen. De stimulatie van factoren die binden aan een 'promotor' kan een gen dat voorheen niet tot uiting kwam tot expressie laten komen (Figuur 4). De activatie van 'enhancers' leidt tot een verhoogde expressie van de genen die door dit element gereguleerd worden. Externe signalen



Figuur 4. Gen-regulerende elementen controleren de expressie van genen. Ons DNA codeert niet enkel genen, maar bevat ook elementen die de uiting van een gen reguleren. (A) Enhancers kunnen de transcriptie van een gen bevorderen. Deze elementen bevinden zich voor of achter de promotor, welke de transcriptie aan of uitzet, van een gen. (B) Gen-regulerende elementen liggen soms op enige afstand van de genen die zij reguleren. Ondanks de afstand kunnen enhancers direct de genexpressie beïnvloeden door middel van vouwingen (looping) van het DNA.

kunnen ook leiden tot de activatie van eiwitten die binden aan 'repressor' elementen, en hierbij de uiting van een gen verminderen.

De uiting van een gen wordt niet alleen gereguleerd door wel of geen transcriptie te laten plaatsvinden, maar ook door het controleren van de transcriptie als deze eenmaal is begonnen. Na de aanvang van transcriptie volgt er een verlengingsfase waarbij het mRNA-molecuul groeit naarmate RNA-polymerase meer DNA afleest (Figuur 3). De transcriptionele verlenging kan worden gepauzeerd of vroegtijdig worden gestopt. Hierdoor ontstaat er een onvolledige omschrijving naar de DNA-code naar mRNA, welke niet kan worden vertaald naar een functioneel eiwit. Omgevingssignalen kunnen de pauzering of vroegtijdige afbraak van transcriptie stimuleren wanneer dit nodig is.

De veranderingen in genexpressie ten gevolge van signalen uit de omgeving zijn het resultaat van een gecoördineerde regulering van al deze verschillende processen. Aanpassingen in de toegankelijkheid van het DNA gaan samen met controle over de activiteit en efficiëntie van de productie van een mRNA-molecuul. Het vermogen om verschillende stadia van de genexpressie te reguleren, biedt de mogelijkheid om de uiting van genen aan te passen aan de omgeving. De snelle en specifieke aanpassing aan signalen uit het milieu is cruciaal om de levensvatbaarheid en functie van onze cellen te behouden. Daarom bepaalt de genregulatie in respons op een omgevingssignaal uiteindelijk het lot van een cel.

De openstaande vragen over genregulatie door omgevingssignalen

Hoewel onze kennis over de regulering van genexpressie door omgevingssignalen enorm is gevorderd door jaren van onderzoek, blijven veel vragen onbeantwoord. De inzichten in de verschillende functionele eigenschappen van eiwitten die reageren op signalen en de

veranderingen in genexpressie die ze teweegbrengen zijn nog altijd onvolledig. Ook blijft het onduidelijk hoe de chromatinestuctuur deze acute veranderingen toelaat en faciliteert. Het ophelderen van de mechanismen die ten grondslag liggen aan de gecoördineerde processen die plaatsvinden wanneer een cel veranderingen in de omgeving waarneemt, is belangrijk voor een nauwkeurig begrip van het gedrag van cellen in zowel gezondheid als ziekte.

Bevindingen beschreven in dit proefschrift

Dit proefschrift heeft tot doel om de moleculaire mechanismen die acute veranderingen in genexpressie teweegbrengen in kanker en stamcellen ten gevolge van omgevingsstress te ontcijferen. Ons werk richt zich hierbij op de twee meest regeneratieve systemen van het lichaam, namelijk de huid en het bloed.

Extrinsieke signalen bepalen het lot van huidkankercellen

De pigmentcellen van de huid (melanocyten) liggen ten grondslag aan een zeer agressieve en kwaadaardige vorm van huidkanker, beter bekend als melanoom. Melanoomcellen vertonen een afwijkende genexpressie ten opzichte van melanocyten. Zo komen genen die normaal gesproken tot expressie komen in de stam- en voorlopercellen (neurale kamcellen) waaruit melanocyten vormen tijdens de ontwikkeling ongepast tot uiting in melanoomcellen. Dit hint op een proces waarbij melanoomcellen 'stamcelachtig' gedrag aannemen tijdens de transformatie van goedaardige melanocyt naar kwaadaardige melanoomcel door het induceren van een genexpressie patroon van een neurale kamcel.
Uit eerder onderzoek is gebleken dat het geneesmiddel Leflunomide specifiek de expressie van neurale kamcelgenen kan inhiberen in melanoomcellen. Dit gebeurt doordat Leflunomide leidt tot een verstoorde transcriptionele verlengingsfase en dus defectieve productie van mRNA-moleculen. Inhibitie van de uiting van neurale kamcelgenen in de melanoomcellen vermindert de groei en uitzaaiing van melanoom. In **hoofdstuk 2** onderzochten wij het onderliggende moleculaire mechanisme waardoor Leflunomide de transcriptionele verlenging remt. Leflunomide blokkeert de productie van de bouwstenen van ons DNA als RNA, genaamd nucleotiden. Wij identificeerden dat het eiwit DDX21 fungeert als een sensor van de nucleotidenniveaus in melanoomcellen. Dit is echter niet de enige functie van DDX21. DDX21 is ook betrokken in de regulatie van genexpressie. DDX21 wekt transcriptie op door te binden aan de 'promotors' van genen en stimuleert de transcriptionele verlengingsfase. Lage niveaus van nucleotiden, zoals geïnduceerd door Leflunomide, worden waargenomen door DDX21 en resulteert in het verlies van DDX21 van het transcriptiecomplex. Op deze manier wordt de transcriptionele verlenging, en dus de efficiënte mRNA productie, voorkomen in omstandigheden waarbij

er niet genoeg bouwstenen aanwezig zijn voor dit proces in de cel. Wanneer losgekomen van chromatine, dan bindt DDX21 aan eerder geproduceerde mRNA-moleculen. Hier speelt DDX21 waarschijnlijk een rol in het stabiliseren van reeds aanwezige mRNA-moleculen. Mogelijk omdat deze belangrijk zijn voor de overleving van de cel en het herstel van de stressvolle omstandigheden. Dit onderzoek beschreef een nieuwe functie van DDX21 als sensor van nucleotidenniveaus en laat zien dat DDX21 een belangrijke factor is ten tijde van nucleotidenstress. Patiënten die chemotherapie ondergaan die gericht is op blokkeren van het nucleotidemetabolisme, ontwikkelen vaak resistentie tegen deze behandeling. Onze studie suggereert dat het moduleren van DDX21 niveaus een aantrekkelijke therapeutische strategie zou kunnen zijn om ziekteprogressie te vertragen.

Omgevingssignalen controleren de genexpressie en functie van bloedstamcellen

Naast de cellen van de huid, bestudeerden wij ook de meest regeneratieve cellen van het bloed. Volwassen mensen hebben elke dag meer dan een miljard nieuwe bloedcellen nodig. Deze taak wordt uitgevoerd door de hematopoëtische stamcellen (HSC) die zich in het beenmerg bevinden. HSC hebben het potentieel om alle typen bloedcellen te maken en worden klinisch gebruikt voor transplantatie bij patiënten met een verscheidenheid aan bloed- en immuunstoornissen. Hierbij worden HSC van een gezonde stamceldonor toegediend aan een patiënt. De stamcellen migreren vervolgens naar het beenmerg van de patiënt waar zij zich inplanten en de bloedproductie overnemen. Het zelfvernieuwingspotentieel dat HSC kenmerkt en de productie van bloedcellen gedurende ons hele leven mogelijk maakt, maakt deze stamcellen ook uniek vatbaar voor kwaadaardige transformatie. Een goede controle over het lot van HSC fungeert daarom als een kritische barrière tegen kanker.
Prostaglandinen zijn hormoonachtige stoffen die een rol spelen op vele fysiologische processen, zoals ontstekingen (inflammatie) en pijn. Prostaglandine E2 (PGE_2) en het stabiele derivaat 16,16-dimethyl-prostaglandine E2 ($dmPGE_2$) zijn eerder beschreven als signalen welke het inplantingspotentieel van hematopoëtische stamcellen (HSC) sterk verbeteren. Een korte stimulatie van donor HSC met $dmPGE_2$ voorafgaand aan een stamceltransplantatie leidt tot een betere en versnelde inplanting van de donor HSC in het beenmerg van de patiënten die deze cellen ontvangen. Om de respons die $dmPGE_2$ teweegbrengt in HSC in detail te begrijpen, bestudeerden wij in **hoofdstuk 3** de veranderingen in de uiting van genen, eiwitten en stofwisseling die $dmPGE_2$ induceert in menselijke HSC. Wij vonden dat een stimulatie van HSC met $dmPGE_2$ leidt tot acute, tijdelijke verhoging van de expressie van genen die betrokken zijn bij de migratie en deling van HSC. Ook beschreven we een aanhoudende inductie van genen die coderen voor factoren die de beenmergniche voorbereiden op de stamcelkolonisatie. Daarnaast identificeerden we veranderingen in eiwitten en metabolieten die betrokken zijn bij de chromatinestructuur en de overleving van HSC. Dit wetenschappelijk werk biedt de

eerste uitgebreide ontleding van de acute veranderingen die dmPGE$_2$ teweegbrengt in hematopoëtische stamcellen. Het onderzoek toonde dat het gedrag van HSC in respons op dmPGE$_2$ een gevolg is van cumulatieve aanpassingen op gen, eiwit en metabool niveau.

In **hoofdstuk 4**, breidden wij onze studies uit naar onderzoek van hematopoëtische stamcellen en hun respons op omgevingssignalen wanneer zij zich in hun natuurlijke omgeving bevinden, namelijk de beenmergniche. Wij stelden een zoogdierlijk modelorganisme (de huismuis) bloot aan verschillende inflammatie-gerelateerde signalen. Na blootstelling isoleerden we de bloedstamcellen uit het beenmerg van het organisme en profileerden wij de cellen. Wij onderzochten de veranderingen in genexpressie en chromatinetoegankelijkheid van HSC na stimulatie met dmPGE$_2$, Poly I:C, of G-CSF. Stimulatie met deze stresssignalen leidde tot snelle veranderingen in de genexpressie van HSC wanneer zij zich in hun natuurlijke niche bevinden. Daarnaast beschreven wij dat HSC onder normale omstandigheden een heterogene populatie vormen waarin de stamcellen zich in verschillende genexpressie en chromatinestaten bevinden. Deze condities zijn zeer dynamisch en HSC kunnen zich in korte tijd tussen deze staten verplaatsen, zoals werd waargenomen in respons op de omgevingssignalen. De geobserveerde heterogeniteit in het genexpressie en chromatine landschap van HSC kan ten grondslag liggen aan de functionele diversiteit in de stressresponsen van HSC.

Om het moleculaire mechanisme van stress geïnduceerde geninductie in te begrijpen, onderzochten we in **hoofdstuk 5** de veranderingen in chromatine structuur en in menselijke HSC. Wij stimuleerden HSC met dmPGE$_2$ en profileerden vervolgens het chromatinelandschap vóór en na blootstelling aan deze stress. We ontdekten dat dmPGE$_2$ de transcriptiefactor CREB activeert, welke vervolgens specifieke enhancers bindt en deze aanzet om de expressie van de bijhorende genen aan te passen. De enhancers die geactiveerd worden door dmPGE$_2$ hebben een speciale chromatineorganisatie. Zij bevatten namelijk instabiele nucleosomen die zijn opgebouwd met het variante histoneiwit H2A.Z. Deze H2A.Z-bevattende nucleosomen worden structureel gemodificeerd in respons op dmPGE$_2$. De H2A.Z histoneiwitten ontvangen namelijk een acetylgroep ten gevolge van stress. Deze covalente, chemische modificatie neutraliseert de positieve elektrostatische lading van histoneiwit H2A.Z. Hierdoor vermindert de aantrekking tussen het nucleosoom en het negatief geladen DNA. Dit leidt tot een grotere toegankelijkheid van het DNA in stress-induceerbare enhancers. Er was een directe correlatie tussen stress-gemedieerde acetylering van H2A.Z in enhancers en de inductie van genen die door deze enhancers gereguleerd worden. Binding van CREB in deze enhancers kan de acetylering van H2A.Z faciliteren door directe interactie van de transcriptiefactor met het enzym p300, welke acetylgroepen overdraagt en vastmaakt aan H2A.Z. Deze studie onthult een mechanisme waarbij transcriptiefactoren gerichte veranderingen in het chromatinelandschap van stress-gevoelige enhancers in gang zetten. De daaropvolgende activatie van deze enhancers zorgt voor de waargenomen geninductie.

Concluderende opmerkingen

Tot slot worden de bevindingen van dit proefschrift in **hoofdstuk 6** in het bredere licht besproken. Dit onderzoek streefde ernaar de fundamentele principes te identificeren waarmee extrinsieke signalen invloed uitoefenen op genexpressie en dus het lot en de functie van cellen. Adaptieve mechanismen zoals multifunctionele DNA-bindende factoren, activatie van gen-regulerende elementen en veranderingen in het chromatinelandschap geeft cellen de mogelijkheid om de uiting van genen direct aan te passen aan signalen uit de omgeving. Hoewel ons werk zich specifiek richtte op cellen uit de twee meest regeneratieve systemen van het lichaam, namelijk de huid en het bloed, gelden onze bevindingen wellicht ook voor andere typen cellen. Ondanks dat er veel vragen onbeantwoord zijn gebleven, onthult het hier beschreven werk de complexiteit, dynamiek en interactie van de vele verschillende niveaus waarop de genexpressie wordt gereguleerd in een veranderde omgeving.
Een beter begrip van deze biologische processen kan rechtstreeks invloed hebben op behandelingsregime voor verschillende ziekten. Onze huidige strategieën om stamcellen te manipuleren voor therapeutische toepassingen wordt beperkt door ons onvermogen om het lot van een cel uiterst gecontroleerd te beheersen. Nauwkeurig begrip van de mechanismen die genexpressie reguleren kan de gecontroleerde modulatie van cellen mogelijk maken en dus basis vormen voor nieuwe medische interventies.

Acknowledgements

The secret is to surround yourself with people who make your heart smile.

None of this would have been possible without the support of the brightest, kindest and most loving people I am fortunate enough to have in my life. While there simply are not enough, or even the right, words to express the gratitude I feel towards those who have supported me over the last 5 years, allow me to try to the best of my abilities.

To my promotor, **Prof. dr. Leonard Zon**. Dear **Len**, most of my gratitude goes to you. Thank you for creating an environment where I was able to fully develop myself and grow in all sorts of directions, both professionally and personally. You have been incredibly supportive from the very first moment I walked into your office during my research internship. Although a little overwhelmed at first by your suggestion to stay for my PhD, it was one of the best decisions I have ever made. I remain impressed by the way you lead the group, by your ever-continuing passion for science, and your enthusiasm to come up with new ideas at all times. No experiment or idea is out of reach for you. At most, there are just some logistical obstacles that can easily be overcome by reaching out to the expert in a field, without any hesitation. You are an extraordinary advisor, caring not only about the science but equally much about the people. I am and will remain thankful for your mentorship and the experience of being part of your lab. I hope we will keep finding ourselves at the same concerts in the future. Because while one can check out, you can never leave the Zon lab.

Mijn promotor, **Prof. dr. Hans Clevers**. Beste **Hans**, ook zonder jou was dit boekje er simpelweg niet geweest. Bedankt voor het vertrouwen om op te treden als mijn promotor en daarmee dit alles mogelijk te maken. Ik heb altijd onwijs genoten van onze eerlijke, nuchtere gesprekken over de voortgang van mijn projecten en de volgende stappen in mijn carrière. Jouw advies, objectieve kijk, en mening is me erg veel waard geweest gedurende dit hele traject. Bedankt dat je deze zo openlijk met mij hebt gedeeld.

To the members of my dissertation committee; **Prof. dr. Frank Holstege, Prof. dr. Wouter de Laat, Prof. dr. Alexander van Oudenaarden, Dr. Catherine Robin, Prof. dr. Eva van Rooij** and **Prof. dr. Michiel Vermeulen**. Thank you for the willingness to serve on my thesis committee and as opponents during my defense. I deeply appreciate the time all of you gave to me and your efforts spent on my dissertation.

I owe immense gratitude towards every individual in **the Zon lab**. Each and every member has contributed to my development as a scientist and a person over the last 5 years. I would not be here without any of you. The Zon lab is a place where talented, kind, and enthusiastic scientists come together. Many have generously shared their time to teach and help me during this PhD. I appreciate all the input and feedback I was provided on my project, and thoroughly enjoyed the conferences, retreats, and socials we attended together. I will hold dear the friendships that were made.

While the entire lab played an instrumental role, a number of past and present members have been especially important to me during this journey and deserve a few extra words.

To my lab spouse, **Eva**. You were not only incredibly important during my time in the Zon lab but also beyond. From research internship mentor, to Prostaglandin partner in crime, but foremost a close friend. I was very fortunate to work with you. Knowing that I had you around for both the research and social aspects of my life gave me the confidence to do my PhD in Boston. I continue to love your honest, critical, and realistic view on science and life. Thank you for keeping me grounded. I am so grateful we remain to be close friends.

Margot, where would I be without you. You walked in as my new technician, and out as one of my very best friends. Sometimes you meet people and you know you will want to keep in your life for an infinite amount of time. You are not only incredibly bright, but also an amazing person. Thank you for all the moments we continue to share together: from talking about books, to walks around the pond and weekly phone calls. You will go on to accomplish so many great things and I can't wait to see it all happen. I'm so proud of you.

Anne, I cannot imagine a life in Boston without you being part of it. It was instantly clear to both of us that we both wanted to keep each other around, with or without the lab to connect us. You are truly a wonderful friend and I love every moment we spend together. I feel blessed to be a part of your life and am constantly amazed by all that you do.

Dear **Cristina**, my experience wouldn't be what it is if it weren't for you. The lab was incredibly lucky to have you. You are one of the kindest people I know. Your hugs and encouragement have meant a lot to me. Thank you for being there during my PhD and after. Us working closely together during the revisions of the paper remains one of the most joyful moments of my PhD. I treasure this time and will remember it fondly. Thank you for trusting me. I would not be here without your generosity and confidence in me.

Avik, from teaching me all the different sequencing techniques, to pushing my scientific thinking, and rooting for me from the first row whenever I gave a talk. You were one of my strongest cheerleaders. You have always valued my opinion, hereby giving me the confidence to speak up from the very beginning. Simultaneously, it was you who taught me when to hold back a little and how to politically rephrase my thoughts since my Dutch directness is not always appreciated. You were the calmness in my PhD and always there to talk to about whatever was on my mind. I am so grateful for you.

To my PhD buddy, **Maurizio**. We did it. How proud I am of us. It has been so wonderful to take this journey on with you. From starting together in the Zon lab to graduating less than 2 months apart. I wouldn't have made it without you. Thank you for letting me share all my strong opinions without qualms and for giving yours in similar vein. Our conversations have kept me sane. I look forward to seeing you take on the next steps. I have no doubt you are going to be fantastic at all that you will do.

After years of being the only graduate student in the Cambridge side of the Zon lab, there was **Georgia**. Someone who shares my deep love for incredibly bad movies. I am so happy I got to know you these past years. You have a beautiful heart. I admire how true you have stayed to yourself, your goals, hopes, and dreams. We shared the happiest and the saddest days of our PhD together. I don't know what I would have done without your loving support. Thank you for being you.

Matt, in a place where a lot of people come from a background very different than ours, we quickly understood each other's view on life and way of thinking. You therefore immediately became a good friend who is there for me at all times. I never have to hold back when talking to you. I admire your strength and resilience. You are incredibly smart and I have no doubt you will achieve all that you set your mind to.

Dear **John**, you made not only work, but also the world, a better place. I loved every minute that I was able to spend with you. You were the one who always looked out for me. I am grateful for all the things we got to do together. But mostly, I am thankful for all the conversations we had about all aspects of life. I will cherish our friendship forever. Boston was a better place with you in it and it has never been quite the same without you around. I miss your hugs. Thank you for all that you gave me.

Naast Boston, zijn er natuurlijk ook enkele hele speciale mensen in Nederland. Zij hebben mij gemaakt tot wie ik ben en me aangemoedigd om te gaan en staan waar ik maar wenste. Ik kan hen dan ook niet genoeg bedanken.

Mijn nichtje, **Dena**. Ik kan nog altijd niet geloven dat jij intussen al op de helft van je Bachelor studie bent. Wat heerlijk dat ik je vanaf de eerste dag jezelf heb mogen zien ontwikkelen tot de persoon die je nu bent. Ik ben onwijs trots op hoe jij je eigen weg aan het uitcarven bent. Ik hoop dat ik je zo af en toe heb kunnen laten zien dat je vooral volledig moet gaan voor de dingen die jij belangrijk vindt. Daarnaast heb jij ook een grote rol gespeeld bij dit boekje, want zonder jou zouden er toch wel heel wat meer spelfouten in de Nederlandse samenvatting zitten.

Lieve **Tim**, daar staan we dan meer dan 5 jaar later. Ik had nooit durven hopen dat de studentenkamer in het Stratenum mij zo een dierbare vriend zou opleveren. Iemand met wie ik zowel in Utrecht en Boston gezellig kon borrelen. Ik ben onwijs blij dat onze vriendschap ondanks de afstand nog altijd sterk is en wij ook tijdens onze PhD door zijn blijven kletsen. Bedankt dat je er altijd voor me bent. Ik kijk er naar uit om te zien wat de toekomst je allemaal gaat brengen en twijfel er niet aan dat jij veel mooie dingen zal doen.

De liefste meisjes, **Tess en Charlotte**. Vriendinnen maken op een nieuwe school valt niet mee. Maar als je jezelf dan toevallig alle drie in deze situatie bevindt en bij elkaar in de klas wordt gezet, dan gaat dat toch wat makkelijker. Na enkele weken tekenles was het dan ook duidelijk dat dit een mooie vriendschap zou worden. Hoewel we door de jaren heen ieder onze eigen paden bewandelden, was onze vriendschap altijd daar. Ik ben zo onwijs dankbaar voor jullie en alles wat wij bereikt hebben in de afgelopen 15 jaar. Lieve Tess, wat fijn dat jij daarnaast ook nog eens mijn paranimf bent op deze bijzondere dag.

Het allerbeste vriendinnetje, **Dani**. Eigenlijk kan ik er nog altijd niet over uit hoe bijzonder onze vriendschap is. Ik heb weinig woorden nodig om een gevoel, gedachte of idee aan jou uit te leggen. Het blindelings vertrouwen dat wij al 10 jaar in elkaar hebben, is me dan ook ontzettend dierbaar. Maastricht, Finland, Utrecht, Boston, en de Grand Canyon, met jou ga ik overal naartoe. Welke turbulente tijd het ook was, wij konden elkaar altijd een knuffel geven wanneer dat het meest nodig was. Ik denk dan ook vaak terug aan de momenten waarop wij samen waren en de gesprekken die we voerden. Ik ben onwijs trots op wat jij allemaal bereikt hebt, niet alleen professioneel maar vooral persoonlijk. Bedankt dat jij mijn paranimf wilt zijn. Ik kan me niet voorstellen dat er iemand anders naast me zou staan op dit moment dan jij.

Lieve, kleine **Leyna**. Wat mooi ben je toch. Jij brengt enkel liefde en vrolijkheid in de familie. Zo klein als je bijna 5 jaar geleden was, zo dapper en wijs ben je nu. Van je eerste woordjes naar hele verhalen die je graag vertelt. Heerlijk eigenwijs en met een sterke mening, misschien lijk je dan ook wel een klein beetje op mij. Ik kijk er naar uit om te zien welke avonturen jij zal opzoeken. Ik hoop dat je weet dat er niets is dat buiten je bereik ligt. Ik zal je er altijd in steunen om je doelen te bereiken. Je bent prachtig op elke manier.

Mijn zus, **Claudia**. Ik kan me niemand indenken die zo dicht bij me staat en tegelijkertijd zo verschillend is van mij. Hoewel we allebei totaal andere dingen willen uit het leven, hoop ik dat je weet dat ik altijd naast je sta. Ik zal er altijd zijn om je hand vast te houden en je lief te hebben. Welke keuzes je ook maakt en wat je ook wilt doen. Ik ben trots op wat jij allemaal doet en hoe jij altijd voor ons klaar staat. Ik kan me geen betere zus voorstellen. Jij verrijkt me. Ik houd van jou.

Papa, dit is allemaal voor jou. Wat had ik jou graag hier bij me gehad. Er zijn zo veel gesprekken die ik nog met jou zou willen voeren en ontelbare momenten die ik met je zou willen delen. Want hoewel het gemis door de jaren went, wordt het nooit kleiner. Ik hoop dat je trots op ons bent. Ik heb je lief.

Mijn **Mama**, ik weet niet eens waar ik moet beginnen. Wat moest ik toch zonder jou. Ik zou hele pagina's vol kunnen schrijven over hoe dankbaar ik voor jou en over hoe bijzonder jij bent. Je bent de sterkste vrouw die er is. De kracht die jij toont en de opofferingen die jij voor ons gemaakt hebt, zal ik nooit als vanzelfsprekend beschouwen. Jouw onvoorwaardelijke liefde en grenzeloos vertrouwen heeft me gemaakt tot wie ik ben. Jij bent dan ook de absolute en enige reden dat dit ik hier sta. Jij zorgde ervoor dat de wereld aan mijn voeten lag. Wat ik ook voor ogen had, door jou was niets buiten bereik. Op elke dag, op ieder moment, heb jij er alles aan gedaan om het allemaal mogelijk te maken. Ik zou nergens zijn zonder jou. Er zijn simpelweg niet genoeg woorden om uit te leggen hoe dierbaar jij me bent en hoeveel ik van je houd. Jij betekent alles voor mij.

The very best thing I will take with me from this entire adventure is you, **Jose**. There are many days I wonder about the serendipity of finding you and ask myself whether I really deserve you. You are every hope I have ever had in human form. Thank you for not giving up on getting to know me. I remain constantly impressed by you. Not only do you amaze me by all the things that you do but perhaps even more so by the way you approach it all. Your considerations and care have a lasting impact on people. You inspire me to be a better person every day. I look forward to all that the future holds for us. I cannot wait to take on the adventures that life will bring us together. Thank you for your support and patience at all times. But mostly, thank you for choosing to love me. Loving you is the easiest thing.

With Love,
Veel Liefs,
Audrey

List of publication

Common variants in signaling transcription-factor-binding sites drive phenotypic variability in red blood cell traits. Choudhuri A*, Trompouki E*, Abraham BJ*, Colli LM, Kock KH, Mallard W, Yang ML, Vinjamur DS, Ghamari A, **Sporrij A**, Hoi K, Hummel B, Boatman S, Chan V, Tseng S, Nandakumar SK, Yang S, Lichtig A, Superdock M, Grimes SN, Bowman TV, Zhou Y, Takahashi S, Joehanes R, Cantor AB, Bauer DE, Ganesh SK, Rinn J, Albert PS, Bulyk ML, Chanock SJ, Young RA, Zon LI. *Nature Genetics.* 2020 Dec;52(12):1333-1345.

RNA helicase DDX21 mediates nucleotide stress responses in neural crest and melanoma cells. Santoriello C*, **Sporrij A***, Yang S, Flynn RA, Henriques T, Dorjsuren B, Custo Greig E, McCall W, Stanhope ME, Fazio M, Superdock M, Lichtig A, Adatto I, Abraham BJ, Kalocsay M, Jurynec M, Zhou Y, Adelman K, Calo E, Zon LI. *Nature Cell Biology.* 2020 Apr;22(4):372-379.

Blood on the tracks: hematopoietic stem cell-endothelial cell interactions in homing and engraftment. Perlin JR*, **Sporrij A***, Zon LI. *Journal of Molecular Medicine.* 2017 Aug;95(8):809-819.

Niche signals regulate continuous transcriptional states in hematopoietic stem cells. Fast EM, **Sporrij A**, Manning ME, Lummertz da Rocha E, Yang S, Zhou Y, Guo J, Baryawno N, Barkas N, Zon LI. *Submitted.*

*equal author contribution

About the author

Audrey Sporrij was born on September 4, 1992 in Eindhoven, The Netherlands. She graduated from pre-university education in 2010 and started a bachelor's degree program in Biomedical Sciences at Maastricht University that same year. During her bachelor's, Audrey performed research in the lab of Dr. U. Von Rango in the Department of Anatomy and Embryology and studied placental development during complicated pregnancies. Audrey graduated from university with a Bachelor of Science (BSc) in 2013; majoring in Molecular Life Sciences and minoring in Public Health from the University of Eastern Finland (Kuopio), which she attended for 4 months during the third year of the program.

Audrey continued her academic training at Utrecht University, where she pursued a master's degree with a focus on Cancer, Stem Cells and Developmental Biology. During her master's, Audrey performed research in the lab of Dr. P. Derksen in the Department of Pathology at University Medical Center Utrecht, and the lab of Prof. dr. L. Zon in the Department of Stem Cell and Regenerative Biology at Harvard University in Boston, USA. Here, she studied mechanisms of breast cancer metastasis and blood stem cell development, respectively. Audrey obtained her Master of Science (MSc) with *cum laude* distinction in 2015.

After graduating with her MSc, Audrey returned to the lab of Prof. dr. L. Zon at Harvard University for her PhD studies under co-supervision with Prof. dr. J. Clevers at the Hubrecht Institute in The Netherlands. She was awarded the Boehringer Ingelheim Fonds PhD fellowship for doctoral students who wish to pursue an ambitious PhD project in basic biomedical research in an internationally renowned laboratory. Her doctoral research focused on the transcriptional regulation of stress responses in blood stem cells and skin cancer. The results of her work are described in this thesis.

During her PhD, Audrey worked as a business development fellow within Harvard's Office of Technology Development to assess the commercial potential of Harvard innovations and develop commercialization strategies. She also played an active leadership role for the Harvard Healthcare Innovation and Commercialization course. Additionally, Audrey was seated on the board of the Harvard Graduate Consulting Club from 2018 to 2021, of which she served two years as club co-president. Upon graduation from her PhD, Audrey will combine her passion for life sciences and drive to make impactful changes in healthcare by starting a new role as strategy consultant at a leading healthcare consulting firm.