Cell

Endothelial Cells Are Central Orchestrators of Cytokine Amplification during Influenza Virus Infection

John R. Teijaro,[1,5] Kevin B. Walsh,[1,5] Stuart Cahalan,[2] Daniel M. Fremgen,[1] Edward Roberts,[3] Fiona Scott,[4] Esther Martinborough,[4] Robert Peach,[4] Michael B.A. Oldstone,[1,*] and Hugh Rosen[2,*]
[1]Department of Immunology and Microbial Science
[2]Department of Chemical Physiology
[3]Department of Chemistry
The Scripps Research Institute, La Jolla, CA 92037, USA
[4]Receptos, Inc., La Jolla, CA 92037, USA
[5]These authors contributed equally to this work
*Correspondence: mbaobo@scripps.edu (M.B.A.O.), hrosen@scripps.edu (H.R.)
DOI 10.1016/j.cell.2011.08.015

SUMMARY

Cytokine storm during viral infection is a prospective predictor of morbidity and mortality, yet the cellular sources remain undefined. Here, using genetic and chemical tools to probe functions of the $S1P_1$ receptor, we elucidate cellular and signaling mechanisms that are important in initiating cytokine storm. Whereas $S1P_1$ receptor is expressed on endothelial cells and lymphocytes within lung tissue, $S1P_1$ agonism suppresses cytokines and innate immune cell recruitment in wild-type and lymphocyte-deficient mice, identifying endothelial cells as central regulators of cytokine storm. Furthermore, our data reveal immune cell infiltration and cytokine production as distinct events that are both orchestrated by endothelial cells. Moreover, we demonstrate that suppression of early innate immune responses through $S1P_1$ signaling results in reduced mortality during infection with a human pathogenic strain of influenza virus. Modulation of endothelium with a specific agonist suggests that diseases in which amplification of cytokine storm is a significant pathological component could be chemically tractable.

INTRODUCTION

Morbidity and mortality caused by severe influenza infections reflect properties that are intrinsic to the virus strain, including replication potential, receptor usage, and cytopathic effects on pulmonary epithelial cells (García-Sastre, 2010; Tscherne and García-Sastre, 2011). In addition to viral-intrinsic factors, host-specific traits such as divergent susceptibilities to infection as well as differences in host immune responses may ameliorate or exacerbate both infection and clinical outcome. An overly aggressive innate response, with early recruitment of inflammatory leukocytes to the lung, was a key contributor to the morbidity of the 1918 influenza infection (Kobasa et al., 2007). More recent clinical literature on avian H5N1 infection documented a significant association between excessive early cytokine responses, immune cell recruitment, and poor outcome (de Jong et al., 2006). Public health approaches to influenza pandemics have relied primarily on preventative vaccine strategies and supportive measures, including the use of antiviral therapies. Nevertheless, the speed at which the 2009 H1N1 influenza virus swine pandemic spread during the lag of vaccine availability highlighted the need to identify additional mechanisms for amelioration of influenza virus infection (Openshaw and Dunning, 2010). Antiviral drugs inhibiting virus replication may select for mutational escape, rendering the therapy ineffective. Modulation of the host immune response has the potential advantage of exerting less-selective pressure on viral populations. Though the prospect of blunting cytokine storm is enticing, a major limitation to treating diseases in which cytokine storm contributes to pathogenesis is the limited understanding of the cellular triggers of this process. Thus, we sought to define cellular signaling pathways that are chemically tractable to test the hypothesis that modulating cytokine storm would provide insight into influenza pathogenesis, with potential therapeutic implications.

Chemical modulators of the sphingosine-1-phosphate (S1P) signaling system have provided insight into immune cell trafficking and immune responses (Rivera et al., 2008; Rosen et al., 2009). Due to the immunomodulatory properties of S1P receptor signaling, chemical agonists have been used successfully for the treatment of relapsing and remitting multiple sclerosis (Brinkmann et al., 2010). Though the systems biology of the S1P receptors is complex, with five receptor subtypes, differential expression, coupling, attenuation, and catabolism (Rosen and Liao, 2003), the availability of selective chemical probes together with mouse genetic models allows detailed insights into immunopathogenesis. We previously showed that a non-selective S1P receptor agonist inhibited cytokine storm, dendritic cell migration, and subsequent antigen-specific T cell

proliferation, protecting the lung tissue from host-mediated injury. Alleviation of immunopathology by the nonselective S1P receptor agonist did not affect virus clearance nor the production of neutralizing antibodies, demonstrating that, although cytokine storm was diminished, retention of a sufficient cellular and humoral immune response, as well as long-term protective immunity, still occurs (Marsolais et al., 2008; Marsolais et al., 2009).

Here, through the use of an $S1P_1$ receptor subtype-selective agonist as well as genetic and biochemical tools, we define a crucial endothelial signaling loop that is important for the initiation of cytokine storm. We reveal that cytokine secretion and immune cell infiltration are separable events that are both regulated by the pulmonary endothelium. Further, we demonstrate that suppression of early innate immune responses through $S1P_1$ signaling results in reduced mortality during human pathogenic influenza virus challenge. Thus, $S1P_1$ receptor signaling in endothelium provides a mechanism for attenuation of influenza virus-induced morbidity and reveals an unexpected role for endothelial cells as regulators of cytokine storm.

RESULTS

Administration of an $S1P_1$ Agonist Blunts Cytokine Storm

Previously, we reported that treatment of influenza virus-infected mice with AAL-R, a promiscuous S1P receptor agonist for $S1P_1$ and $S1P_{3-5}$ receptors, inhibits early proinflammatory cytokine expression and innate immune cell accumulation within the lung (Marsolais et al., 2008; Marsolais et al., 2009). To assess the contribution of $S1P_1$ receptor signaling in inhibition of influenza virus-induced early inflammation, infected mice were treated with the $S1P_1$ receptor-specific agonist CYM-5442 (Gonzalez-Cabrera et al., 2008). Treatment with 2 mg/kg of CYM-5442 twice daily significantly inhibited secretion of cytokines and chemokines associated with influenza virus-induced pathology, including IFN-α, CCL2, IL-6, TNF-α, and IFN-γ (Figure 1A) in addition to CCL3, CCL5, CXCL2, and IL-1α (Figure S1A), compared to vehicle-treated mice 48 hr postinfection. CYM-5442 reduction of IFN-α, CCL2, CCL3, CCL5, IL-1α, and IL-6 expression was as complete as treatment with the promiscuous agonist; however, it was not as effective as AAL-R in suppressing CXCL2, TNF-α, and IFN-γ (Figure 1A and Figure S1A), suggesting a role for other S1P receptors in modulating those cytokines. In addition to inhibiting cytokine/chemokine production, AAL-R and CYM-5442 administration to influenza virus-infected mice blunted the accumulation of innate inflammatory infiltrate characterized as macrophages/monocytes ($CD11b^+$, $F480^+$, $Ly6G^-$), neutrophils ($CD11b^+$, $LyG6^+$, $F480^-$), and NK cells ($NK1.1^+$, $CD3^-$), although AAL-R was more effective at inhibiting macrophage/monocyte and NK cell accumulation in the lung (Figure 1B). We also observed significantly reduced CD69 expression on macrophage/monocytes and NK cells following treatment with either CYM-5442 or AAL-R 48 hr postinfection, demonstrating diminished cell activation (Figures 1C and 1D). Despite significant blunting of innate immune cell recruitment and cytokine/chemokine responses, we observed no differences in viral titers following AAL-R or CYM-5442 treatment compared

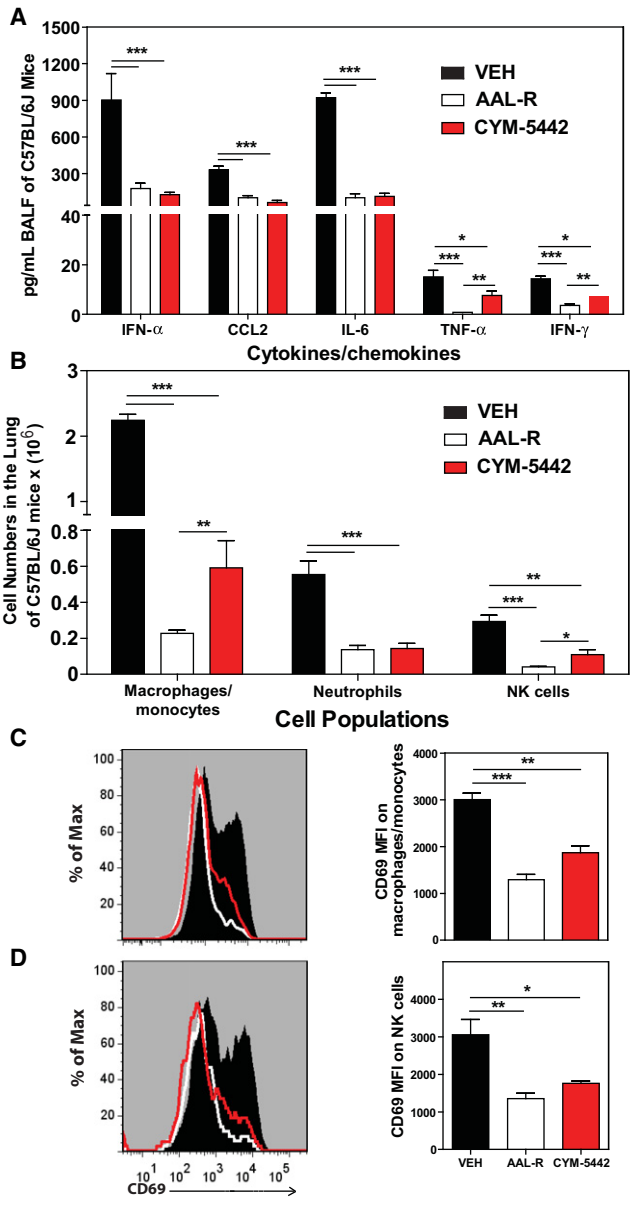


Figure 1. $S1P_1$ Receptor Agonism Suppresses Early Proinflammatory Cytokine and Chemokine Production and Innate Immune Cell Recruitment during Influenza Virus Infection

Mice were infected with 1×10^4 PFU WSN influenza virus, and vehicle (water), AAL-R (0.2 mg/kg) (1 hr postinfection), or CYM-5442 (2 mg/kg) (1,13, 25, and 37 hr postinfection) were administered i.t. to mice.

(A) Proinflammatory cytokines and chemokines were measured 48 hr postinfection in BALF by ELISA.

(B) Total numbers of innate immune cells were quantified from collagenase-digested lungs by flow cytometry at 48 hr postinfluenza virus infection.

(C) Histograms (left) and mean fluorescent intensity (MFI) (right) of CD69 expression on macrophages was quantified on vehicle, AAL-R-, or CYM-5442-treated mice 48 hr postinfluenza virus infection by flow cytometry staining.

(D) Histograms (left) and MFI (right) of NK cell CD69 expression quantified as in (C).

Data represent average ± SEM from four to mice per group. *$p < 0.05$; **$p < 0.005$; ***$p < 0.0005$. Results are representative of greater than six independent experiments. See also Figure S1 and Figure S2.

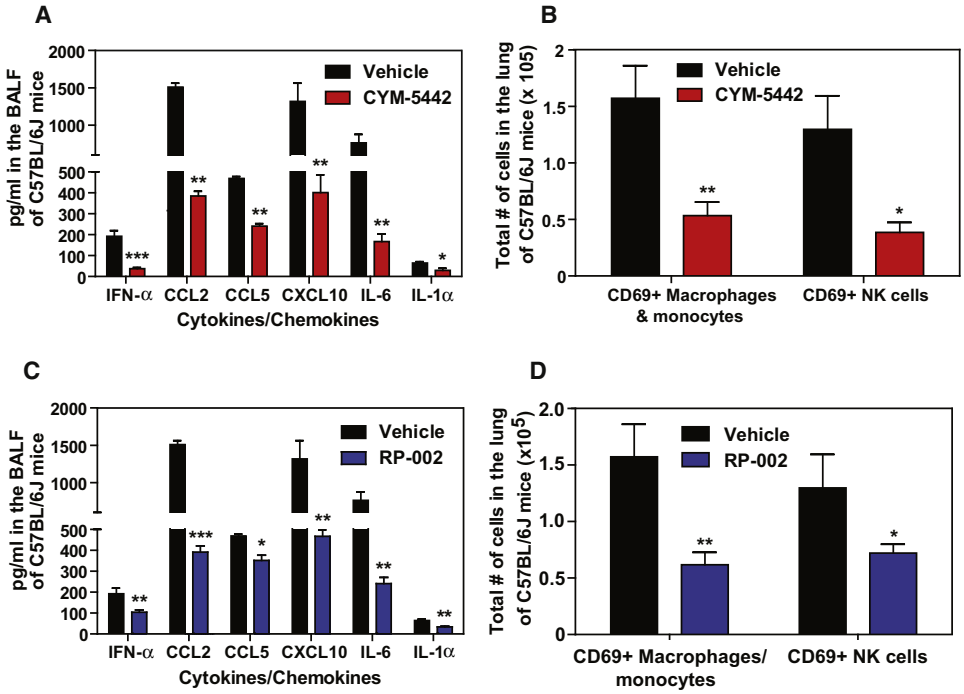


Figure 2. $S1P_1$ Receptor Agonism Suppresses Early Proinflammatory Cytokine and Chemokine Production and Recruitment of Activated Innate Immune Cells during Human Pathogenic H1N1:2009 Swine Influenza Virus Infection

(A–D) Mice were infected with 1 × 10^5 PFU A/Wisconsin/WSLH34939/09 influenza virus, and either vehicle (water), CYM-5442 (2 mg/kg) (1, 13, 25, and 37 hr postinfection), or RP-002 (2 mg/kg on 1 and 25 hr postinfection) were administered i.t. to mice. Proinflammatory cytokines and chemokines were measured 48 hr postinfection in BALF by ELISA in either CYM-5442- (A) or RP-002-treated mice (C). Total numbers of innate immune cells were quantified from collagenase-digested lungs by flow cytometry at 48 hr postinfluenza virus infection in mice treated with either CYM-5442 (B) or RP-002 (D). Data represent average ± SEM from four to five mice per group. *$p < 0.05$; **$p < 0.005$; ***$p < 0.0005$. Results are representative of two independent experiments. See also Figure S3.

to vehicle-treated mice (Figure S2), demonstrating that $S1P_1$ agonism neither inhibits nor enhances viral replication. Collectively, these results reveal that activation of a single sphingosine-1-phosphate receptor, $S1P_1$, is sufficient to blunt global innate inflammation following mouse-adapted influenza virus infection in mice.

We next asked whether $S1P_1$ receptor agonism could suppress early innate cytokine and chemokine responses following infection with a human pathogenic isolate of influenza virus. To test this, we infected mice with a virulent isolate of the 2009 H1N1 pandemic influenza virus that was isolated from a hospitalized patient and has never been passaged through mice (A/Wisconsin/WSLH34939/09) (Itoh et al., 2009). Infection of C57BL/6J mice with this strain causes severe disease, resulting in rapid mortality beginning between days 4 and 6 postinfection (data not shown). Similar to infection with mouse-adapted WSN influenza virus, treatment with CYM-5442 following infection with H1N1:2009 significantly suppressed cytokine and chemokine responses and the accumulation of activated innate immune cells 48 hr postinfection (Figures 2A and 2B). We also tested an additional $S1P_1$ receptor-selective agonist (RP-002) to provide support for $S1P_1$ receptor specificity and to show directly that S1P1-selective agonists share this modulation of innate immunopathology. We infected mice with H1N1:2009 pandemic influenza virus as above and treated mice with RP-002 at 1 and 25 hr postinfection. Similar to what we observed with CYM-5442, RP-002 treatment significantly inhibited the production of multiple proinflammatory cytokine and chemokines (Figure 2C) and suppressed the accumulation of activated ($CD69^+$) macrophages/monocytes and NK cells in the infected lung 48 hr postinfection (Figure 2D). Moreover, RP-002-mediated suppression of innate immune cell recruitment and cytokine/chemokine production occurred without altering lung viral titers (Figure S3), demonstrating that $S1P_1$ agonist-mediated suppression of cytokines and chemokines is not due to direct effects on influenza virus replication.

$S1P_1$ Agonist-Mediated Suppression of Early Innate Immune Responses Results in Protection to Human Pathogenic Influenza Virus Challenge in Mice

Early dysregulated innate immune responses in the lung have been associated with morbidity and mortality during infection with highly pathogenic strains of influenza virus (Cillóniz et al., 2009; Kobasa et al., 2007). Therefore, we asked whether blunting early innate cytokine/chemokine responses using an $S1P_1$ agonist could protect mice from lethal infection with a virulent human isolate of pandemic 2009 H1N1 that had not been passaged in mice (A/Wisconsin/WSLH/34939/09). In order not to impose additional stress on the infected lung and because oral delivery is a popular route for drug administration, we orally administered the $S1P_1$ agonist, RP-002, at 6 mg/kg by gavage to C57BL/6J mice infected with 1 × 10^5 PFU of

A/Wisconsin/WSLH/34939/09 at 1 and 25 hr postinfection. RP-002 significantly suppressed cytokine amplification (Figure 3A) as well as recruitment of myeloid cells to the infected lung, as measured by flow cytometry (Figure 3B).

Early administration of RP-002 resulted in enhanced survival time after lethal challenge with A/Wisconsin/WSLH/34939/09, with death initiating in vehicle-treated mice on day 6 postinfection compared to day 11 postinfection for RP-002-treated mice (Figure 3C, $p < 0.005$). Moreover, RP-002 treatment resulted in significant improvement in overall survival compared to vehicle-treated mice (20% mortality in RP-002 versus 80% mortality for vehicle, $p < 0.005$; Figure 3C). These findings demonstrate a significant biological phenotype resulting from early $S1P_1$ receptor agonist treatment following pathogenic influenza virus infection. By showing directly that aborting an early step in cytokine amplification significantly protects against an ordinarily lethal human pathogenic H1N1:2009 infection, the findings expand previous observations of an association between humans and animal models in which dysregulation of innate immune responses contributes to morbidity and mortality. Importantly, treatment with RP-002 ameliorated morbidity and mortality associated with human pathogenic H1N1:2009 severe influenza infection. Effects upon survival are in keeping with our previously published data (Marsolais et al., 2009) that this mechanism does not alter virus-specific neutralizing antibody response, affinity maturation, and class switching. This documents that the cytokine storm plays a direct and cardinal role in influenza-mediated lung disease. We then sought to delineate the mechanism of S1P1 suppression of cytokine amplification.

$S1P_1$ Is Expressed on Pulmonary Endothelium and Lymphocytes

To identify cells in the lung that express the $S1P_1$ receptor, we utilized an eGFP-$S1P_1$ receptor knockin mouse in which the native $S1P_1$ receptor was homologously replaced with a functional enhanced green fluorescent protein (eGFP)-tagged $S1P_1$ receptor ($S1P_1$-eGFP) (Cahalan et al., 2011). This mouse model allowed direct detection of eGFP-$S1P_1$ receptor protein expression by flow cytometry, with additional biochemical confirmation using highly specific antibodies for GFP followed by western blotting that distinguishes the fused $S1P_1$-eGFP receptor by specific molecular weight. High levels of $S1P_1$-eGFP were detected on lung lymphatic ($CD45^-$, $CD31^+$, $gp38^+$) and vascular ($CD45^-$, $CD31^+$, $gp38^-$) endothelium, whereas pulmonary epithelium ($CD45^-$, $CD31^-$, $EpCAM^+$) was negative for $S1P_1$-eGFP (Figure 4A). As reported previously, $CD4^+$ T cells ($CD4^+$, $CD3^+$), $CD8^+$ T cells ($CD4^+$, $CD3^+$), and B cells ($B220^+$, $CD19^+$) expressed $S1P_1$-eGFP (Figure 4B), whereas leukocytes, including macrophages/monocytes, a subset of alveolar macrophages ($F480^+$, $CD11c^+$, $CD11b^-$), dendritic cells ($CD11c^+$, I/A-I/E^+, $CD205^+$, $F480^-$), neutrophils, NK cells ($NK1.1^+$, $CD3^-$), and innate lymphoid cells (Lin^-Sca1^+), expressed negligible levels of $S1P_1$-eGFP (Figure 4C and data not shown). Cells from uninfected mice expressed similar levels of $S1P_1$-eGFP as infected mice, demonstrating that $S1P_1$ receptor expression is not altered following influenza virus infection (Figure 4D). To confirm the $S1P_1$-eGFP expression pattern, a western blot for $S1P_1$-eGFP was performed on cells FACS sorted from the lungs of infected

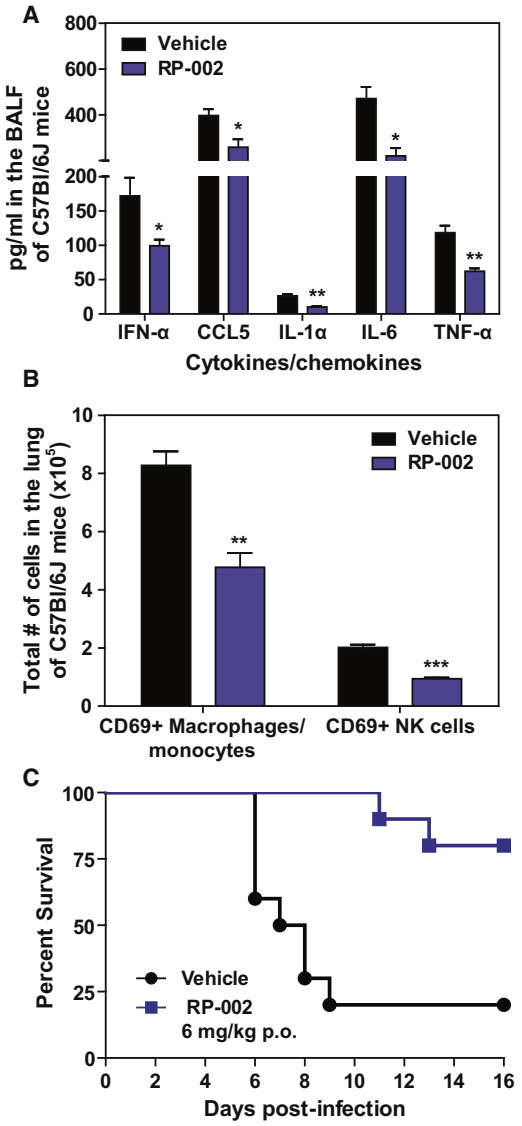


Figure 3. Oral Administration of an $S1P_1$ Agonist Suppresses Early Innate Cytokine and Chemokine Production and Significantly Improves Survival to Lethal Infection with H1N1 2009 Swine Influenza Virus

Mice were infected with 2×10^5 PFU A/Wisconsin/WSLH34939/09 and treated by gavage with either vehicle (water) or RP-002 (6 mg/kg on 1 and 25 hr postinfection for cytokine and cell recruitment assays at 48 hr, with a third dose administered at 49 hr for the 16 day survival experiment).

(A) Levels of proinflammatory cytokines and chemokines were analyzed in the BALF 48 hr postinfection by ELISA.

(B) Total numbers of activated ($CD69^+$) macrophages/monocytes and NK cells were detected in collagenase-digested lungs 48 hr postinfection by flow cytometry.

(C) Mice were monitored daily for survival for 16 days postinfection.

For (A) and (B), data represent the average ± SEM of four to five mice per group, whereas (**C**) had ten mice per group. *$p < 0.05$; **$p < 0.005$; ***$p < 0.0005$. Data are representative of three independent experiments. See also Figure S3.

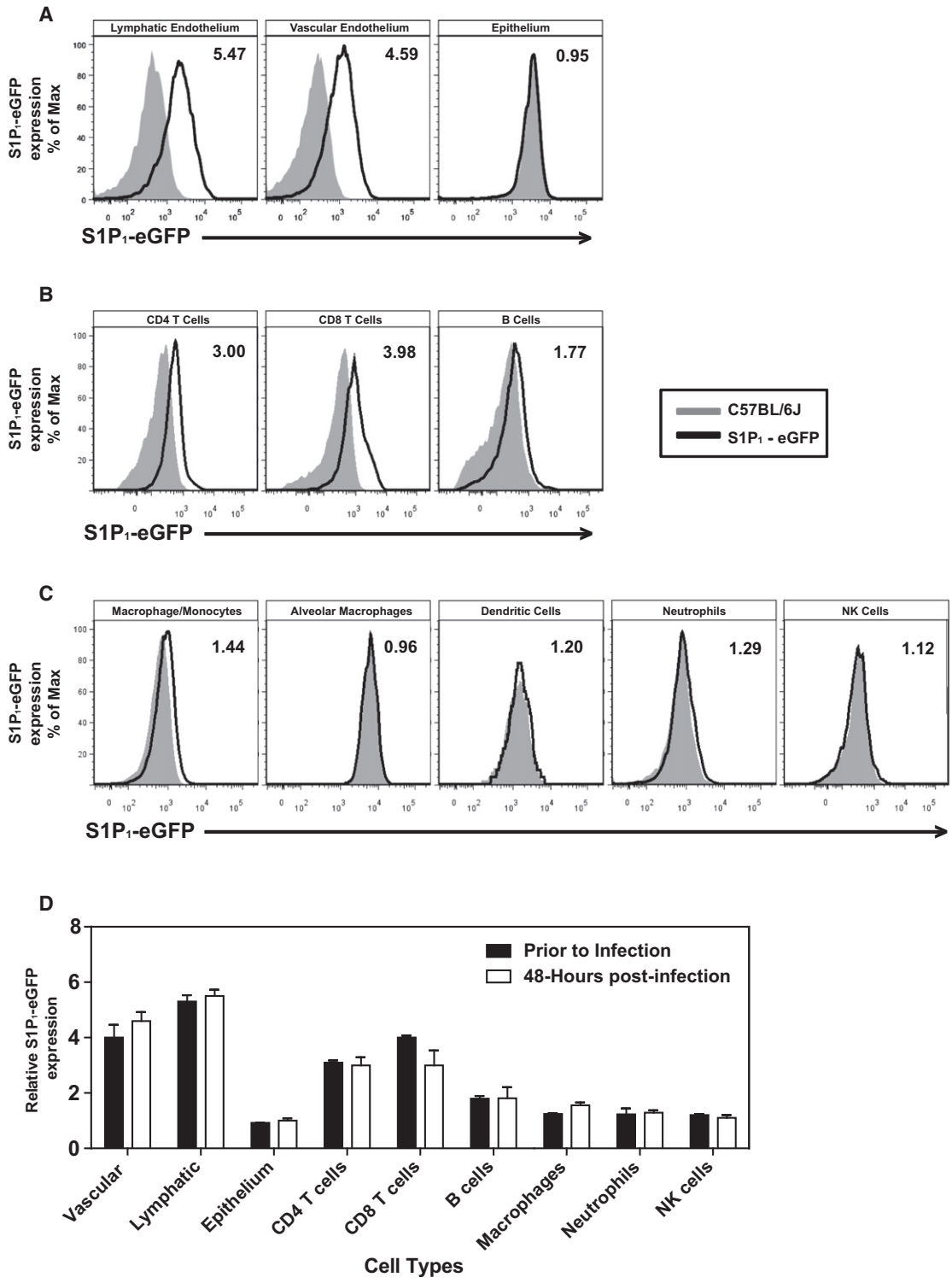


Figure 4. $S1P_1$ Receptor Is Expressed on Pulmonary Endothelium and Lymphocytes, but Not on Pulmonary Epithelial Cells

(A–C) Flow cytometry histograms showing eGFP fluorescence from heterozygous $S1P_1$-eGFP on lung endothelial and epithelial cells. (A) CD4 and CD8 T cells and B cells. (B) Macrophages, neutrophils, dendritic cells, and NK cells (C) prior to influenza virus infection.

(D) Relative expression of $S1P_1$-eGFP on lung cell populations before or 48 hr after influenza virus infection. The numbers in the upper-right corner of each plot represent the relative eGFP expression as calculated by dividing the MFI of eGFP on $S1P_1$-eGFP transgenic mice over eGFP MFI of C57BL/6J littermate controls. Data represent average ± SD from two to five mice per group. Results are representative of three independent experiments. See also Figure S4.

and uninfected mice. Endothelial cells, T cells, and B cells were positive for $S1P_1$-eGFP in uninfected mice, confirming our flow cytometry results (Figure S4A). However, though $S1P_1$-eGFP was detected by western blot in B cells from uninfected mice, we did not detect $S1P_1$-eGFP expression by western Blot in influenza virus-infected lung-resident B cells (Figure S4A). We did not detect S1P1-eGFP expression by western blot in epithelial cells, macrophages/monocytes, and alveolar macrophages either before or after influenza virus infection (Figure S4A). Moreover, CYM-5442 treatment of infected $S1P_1$-eGFP knockin mice did not diminish expression of $S1P_1$-eGFP on endothelial cells, T cells, or B cells, demonstrating that administration of this $S1P_1$-specific agonist does not induce degradation of the $S1P_1$ receptor (Figure S4B). Therefore, CYM-5442 does not exert an antagonistic effect due to receptor degradation, demonstrating functional agonism of $S1P_1$ as the mechanism by which CYM-5442 suppresses cytokine storm.

Inhibition of Cytokine Storm Is Not Due to Lymphocyte $S1P_1$ Receptor Activation

Lymphocytes and pulmonary endothelium are the only cells within the lung that express measurable amounts of $S1P_1$-eGFP as detected by flow cytometry and western blot. It is plausible that lymphocytes within the lung provide a bystander effect during the innate immune response to influenza virus infection. To rule out a role for lymphocytes during $S1P_1$ receptor-mediated inhibition of inflammation following influenza virus infection, lymphocyte-deficient *Rag2*$^{-/-}$ mice were infected and treated with vehicle or CYM-5442. Administration of CYM-5442 to influenza virus-infected *Rag2*$^{-/-}$ mice resulted in significant inhibition of expression of IFN-α, CCL2, IL-6, TNF-α, and IFN-γ (Figure 5A), in addition to CCL5, CXCL2, and CXCL10 (Figure S1B). Atypical lymphocyte populations that could, in principal, contribute to suppression of cytokine production did not express S1P1-eGFP and thus also do not contribute to this mechanism. Inflammatory cell infiltrates, including macrophages/monocytes and NK cells, were significantly reduced in CYM-5442-treated *Rag2*$^{-/-}$ mice (Figure 5B). Reduced numbers of neutrophils were also observed within the infected lung, but the difference was not statistically significant (Figure 5B). Moreover, lung-infiltrating macrophages and NK cells in *Rag2*$^{-/-}$ mice were less activated, as measured by CD69 surface expression (Figure S5). CYM-5442 treatment of influenza virus-infected lymphocyte-deficient mice inhibits the innate inflammatory response, excluding a role for lymphocytes as key regulators of influenza virus-induced cytokine storm.

$S1P_1$ Receptor Agonism Suppresses Immune Cell Recruitment through Downregulation of Chemokine Production by Lung Endothelial Cells

The strong expression of $S1P_1$-eGFP receptor on pulmonary endothelium, coupled with the inhibition of cytokine storm in RAG2$^{-/-}$ mice, suggests that $S1P_1$ receptor signaling in lung endothelial cells suppresses cytokine storm. Endothelial cells have been shown to produce various cytokines and chemokines during inflammatory processes and may be a source of cytokines and chemokines in the lung during influenza virus infection. To evaluate the effects of CYM-5442 treatment on pulmonary endothelial cell cytokine and chemokine production, we purified lung endothelial cell populations and performed mRNA and protein analyses. Endothelial cells expressed elevated levels of the chemokines CCL2, CCL5, and CXCL10 at the mRNA level (Figure 6A). Importantly, treatment with CYM-5442 significantly reduced mRNA expression of chemokines in lung endothelial cells compared to vehicle-treated mice early after influenza virus infection (Figure 6A). Examination of chemokine production in purified lung vascular and lymphatic endothelial cells revealed a significant reduction in chemokine production at the protein level after CYM-5442 treatment (Figure 6B). Analysis of pulmonary endothelial cell integrin expression did not show appreciable differences in mice infected with influenza virus treated with vehicle compared to CYM-5442 (Figure S6).

Because chemokine presentation on endothelial cells is crucial for leukocyte activation and extravasation into infected tissue, we asked whether reduction of chemokine expression following CYM-5442 treatment resulted in reduced entry of inflammatory cells into the infected lung. For these experiments,

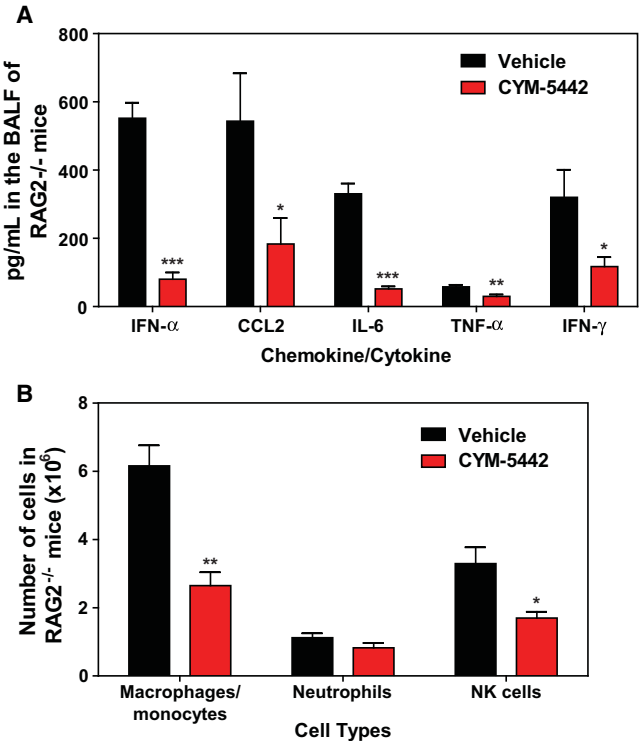


Figure 5. $S1P_1$ Receptor Agonism Suppresses Early Proinflammatory Immune Responses Independently of Lymphocytes

Rag2$^{-/-}$ mice were infected and treated with either vehicle or CYM-5442, as in Figure 1.

(A) Proinflammatory cytokines and chemokines were measured in BALF by ELISA 48 hr postinfluenza infection.

(B) Total numbers of innate immune cell recruitment were quantified from collagenase-digested lungs by flow cytometry at 48 hr postinfluenza virus infection.

Data represent average ± SEM from five mice per group. *$p < 0.05$; **$p < 0.005$; ***$p < 0.0005$. Results are representative of two independent experiments. See also Figure S1 and Figure S5.

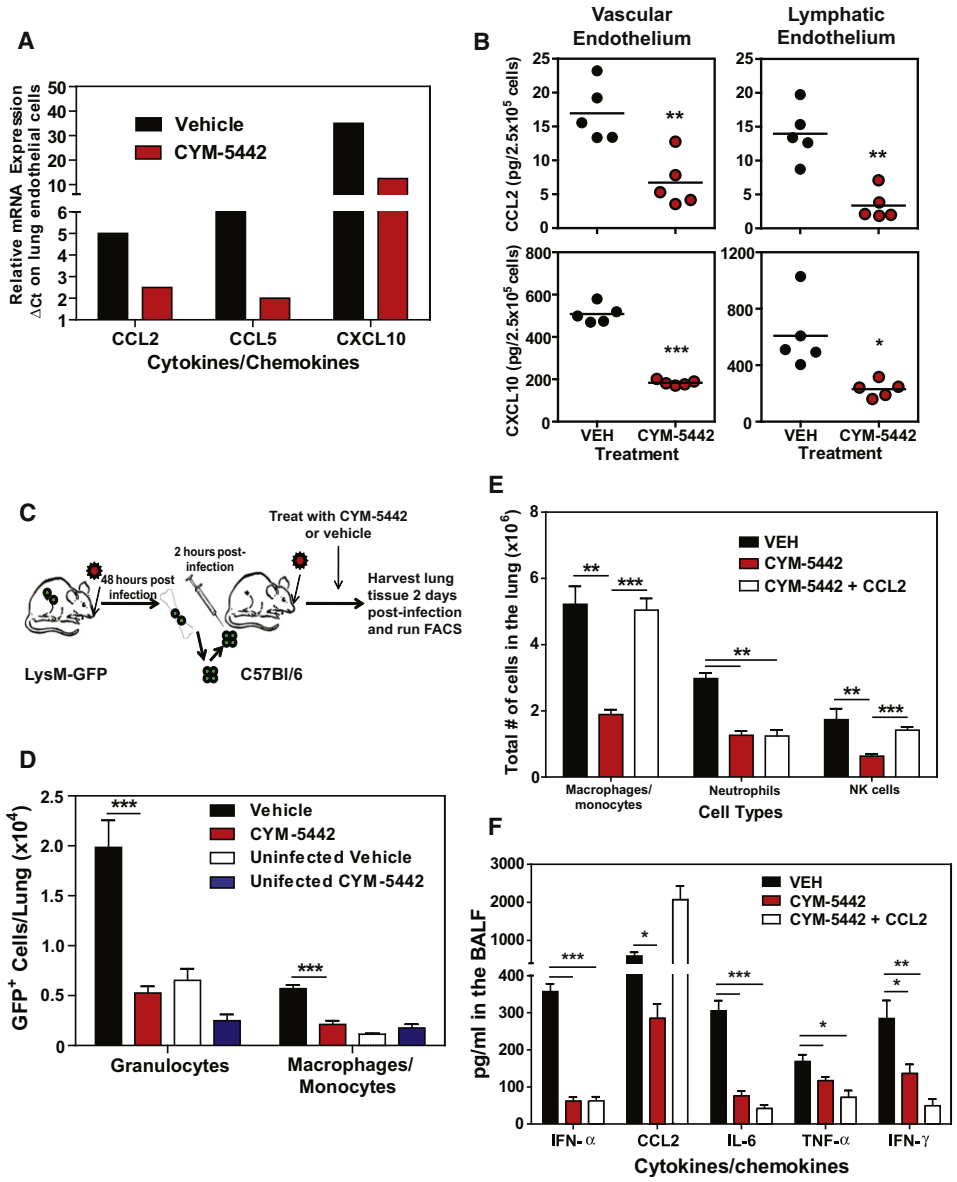


Figure 6. $S1P_1$ Receptor Agonism Actively Suppresses Recruitment of Innate Immune Cells through Downregulation of Chemokine Expression on Lung Endothelial Cells

(A) Relative mRNA expression of chemokines from purified lung endothelial cell populations at 36 hr postinfluenza virus infection.

(B) Protein expression of CCL2 and CXCL10 measured by ELISA on FACS-purified vascular and lymphatic lung endothelial cells 48 hr postinfluenza virus infection.

(C) LysM-GFP mice were infected with 1×10^4 PFU of influenza virus. At 2 days later, bone marrow cells were harvested from infected LysM-GFP mice and transferred into C57BL/6J mice that were either left uninfected or infected with 1×10^4 PFU of influenza virus 2 hr prior. Mice receiving bone marrow cells from LysM-GFP-positive mice were treated with either vehicle or CYM-5442 (2 mg/kg 1, 13, 25, and 37 hr postinfection), and at 48 hr postinfection, total lung cells were harvested from lung homogenates and the total numbers of GFP-expressing granulocytes ($CD11b^+$, $Ly6G^+$) and macrophages/monocytes ($CD11b^+F480^+$) were quantified.

(D) Total number of LysM-GFP-positive granulocytes and macrophages/monocytes per lung of uninfected or infected mice treated with either vehicle or CYM-5442.

(E) Exogenous intratracheal administration of chemokine restores recruitment of innate immune cells. However, it does not resurrect cytokine/chemokine production after CYM-5442 treatment. Mice were infected with 1×10^4 PFU WSN influenza virus, and either vehicle (water) or CYM-5442 (2 mg/kg 1, 13, 25, and 37 hr postinfection) were administered to mice i.t. Following administration of CYM-5442, recombinant mouse CCL2 was administered (50 μg) i.t. directly into the lungs. The bar graph shows the total numbers of macrophages/monocytes, neutrophils, and NK cells in the lung 48 hr postinfluenza virus infection.

(F) Proinflammatory cytokines and chemokines were measured in BALF 48 hr postinfection by ELISA in mice treated with vehicle, CYM-5442, or CYM-5442 + rmCCL2, as in (E).

Data represent average ± SEM from five mice/group. *$p < 0.05$; **$p < 0.005$; ***$p < 0.0005$. Results are representative of two independent experiments. See also Figure S1 and Figure S6.

we used LysM-GFP mice in which macrophages, monocytes, and neutrophils express green fluorescent protein, allowing us to track these cell populations after adoptive transfer into congenic C57BL/6J mice. We infected LysM-GFP mice with influenza virus and 48 hr later purified bone marrow cells for adoptive transfer. 5×10^6 LysM-GFP bone marrow cells were transferred into influenza virus-infected C57BL/6J mice treated with either vehicle or CYM-5442 (Figure 6C). The lungs of infected or uninfected C57BL/6J mice receiving LysM-GFP cells were then harvested 48 hr postinfection, and the numbers of GFP-positive cells in the lung were quantified. Treatment with CYM-5442 resulted in significantly reduced recruitment of both GFP-positive neutrophils ($CD11b^+$ $F480^-$ $Ly6G^+$) and macrophages/monocytes ($CD11b^+$ $Ly6G^-$ $F480^+$) into the lung (Figure 6D). Therefore, CYM-5442 treatment inhibits the infiltration of circulating inflammatory cells into the lung.

$S1P_1$ receptor signaling is associated with enhanced capillary integrity (Rosen et al., 2007), whereby $S1P_1$ receptor antagonists induce vascular leakage (Rosen et al., 2008; Sanna et al., 2006), and S1PR-selective agonists protect from vascular leakage induced by exogenous administration of VEGF (Sanchez et al., 2003). Though myelomonocytic recruitment is not modulated by endothelial permeability changes, we formally assessed whether suppression of inflammatory cell recruitment to the lung reflected a passive inhibition of leukocyte recruitment or whether suppression of chemokines on lung endothelial cells limits innate immune cell recruitment. Rescue experiments were performed in which mice were infected with influenza in the presence or absence of intratracheal administration of recombinant murine CCL2 (rmCCL2), with or without CYM-5442 treatment. Administration of rmCCL2 to infected mice treated with CYM-5442 restored macrophage/monocyte and NK cell numbers in the lung to levels equivalent to infected mice treated with vehicle (Figure 6E) 48 hr postinfection. These data demonstrate that CYM-5442 inhibition of cellular infiltration into the lung is not due to the enhancement of endothelial cell barrier function but due to suppression of chemokine production. Therefore, $S1P_1$ receptor agonism suppresses endothelial cell chemokine expression, resulting in diminished cell infiltration. Despite the rescue of CCL2 levels in the BALF and the restoration of macrophage/monocyte recruitment into the lung of CYM-5442-treated mice, global cytokine and chemokine production was not restored (Figure 6F and Figure S1C). Influenza virus infection of CCR2-deficient mice results in substantial reductions of infiltrating macrophages/monocytes (Dawson et al., 2000). To rule out infiltrating macrophages, we infected CCR2-deficient mice with WSN influenza virus as done above and treated with either vehicle or CYM-5442. Infection of CCR2-deficient mice resulted in significantly reduced numbers of macrophages/monocytes without altering neutrophil or NK cell numbers in the lung 48 hr postinfection (Figure S7A). Despite reduced levels of macrophages/monocytes in the lung, we still detected significant levels of IFN-α, CCL2, CCL3, CCL5, CXCL2, CXCL10, IL-1α, IL-6, and IFN-γ (Figure S7B). More importantly, treatment of CCR2-deficient mice with CYM-5442 still significantly reduced all cytokines and chemokines tested at 48 hr postinfection except for CCL5 (Figure S7B). Failure to inhibit cytokine and chemokine production in CCR2-deficient mice or to re-establish cytokine and chemokine production by the induction of macrophage/monocyte cell infiltration through rmCCL2 administration following CYM-5442 treatment suggests that monocytes and macrophages are not major sources of early cytokine and chemokine production. Furthermore, these data indicate that cell recruitment and cytokine responses may be uncoupled events.

Proinflammatory Cytokine Responses Are Independent of Innate Immune Cell Recruitment and Dependent on Type I Interferon Signaling

Infiltration of macrophages and NK cells alone does not appear to be associated with cytokine storm. Therefore, we evaluated the contribution of total inflammatory cell infiltration to cytokine production during influenza virus infection. Mice were treated with an anti-CD11b antibody (M7/80) that has been shown to inhibit the recruitment of CD11b-expressing cells into inflamed tissues (Rosen and Gordon, 1987). Though treatment with anti-CD11b significantly inhibited the recruitment of macrophages/monocytes, neutrophils, and NK cells into the lung (Figure 7A), we observed no reduction in levels of proinflammatory cytokines/chemokines, other than IFN-γ (Figure 7B), which is probably a direct result of diminished infiltration of NK cells, the primary source of IFN-γ at this time point. Moreover, treatment with CYM-5442 significantly inhibited the production of IFN-α, CCL2, IL-6, TNF-α, and IFN-γ (Figure 7B), in addition to CCL3, CCL5, CXCL2, CXCL10, and IL-1α (Figure S1D) in anti-CD11b-treated mice. These results further demonstrate that innate cell recruitment and cytokine production are independent events early after influenza virus infection, both which are inhibited by $S1P_1$ receptor agonism of pulmonary endothelial cells. Importantly, our data demonstrate that inflammatory cell infiltration into the lung is not required for cytokine and chemokine production. An additional line of evidence excluding hematopoietic cells from $S1P_1$ agonist-mediated reduction of early innate cytokine and chemokine responses following influenza virus infection was provided by irradiation resistance of S1P1 agonist suppression. C57BL/6J mice were irradiated with 500 rads and 24 hr later infected with WSN influenza virus and treated with either vehicle or CYM-5442. Irradiation reduced $CD45^+$ hematopoietic cells in lung by 90% (vehicle: 2×10^7 cell/lung versus 500 rads: 2×10^6 cells/lung), and importantly, treatment of irradiated mice with CYM-5542 significantly reduced the levels of IFN-α (irradiated vehicle: 86 pg/ml versus irradiated CYM-5442: 17.9 pg/ml, $p = 0.0006$), CCL2 (irradiated vehicle: 944 pg/ml versus irradiated CYM-5442: 452 pg/ml, $p = 0.00002$), CXCL10 (irradiated vehicle: 204 pg/ml versus irradiated CYM-5442: 83 pg/ml, $p = 0.00003$), and IL-6 (irradiated vehicle: 170 pg/ml versus irradiated CYM-5442: 70 pg/ml, $p = 0.0001$), demonstrating that a radiation-resistant cell population mediated $S1P_1$ receptor-mediated suppression of the inflammatory response. These results, coupled with the restricted $S1P_1$ receptor expression pattern, support an important role for endothelial cells in regulating pulmonary innate immune responses to influenza virus.

Type I interferons, predominantly IFN-α species, are elevated early after respiratory virus infection and are thought to be crucial for the production of proinflammatory cytokines and chemokines. Our results thus far have demonstrated that CYM-5442

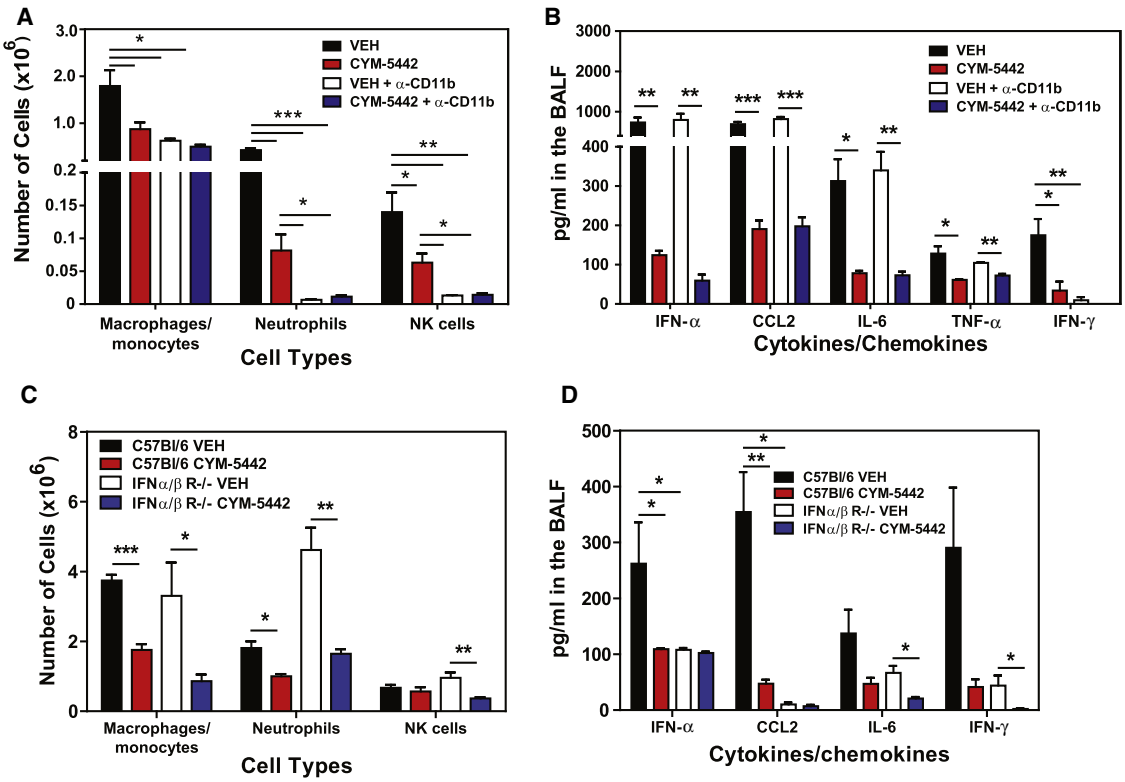


Figure 7. Proinflammatory Cytokine Responses Are Independent of Innate Immune Cell Recruitment and Dependent on Type I Interferon Signaling

(A) Total numbers of innate immune cells were quantified from lung digests by flow cytometry at 48 hr postinfluenza virus infection in mice treated 1 hr post-infection with vehicle or CYM-5442 in the presence of either anti-CD11b or isotype control antibody (M7/80) (0.5 mg/mouse) 0 and 24 hr postinfection.

(B) Proinflammatory cytokines and chemokines were measured 48 hr postinfection by ELISA in BALF in mice treated as in (A).

(C) C57BL/6J or IFN αβ receptor-deficient mice were infected with 1 × 10^4 PFU of influenza virus and treated with either vehicle or CYM-5442 (2 mg/kg 1, 13, 25, and 36 hr postinfection. Total numbers of innate immune cells were quantified from lung digests by flow cytometry at 48 hr postinfluenza virus infection, and

(D) proinflammatory cytokines and chemokines were measured 48 hr postinfection by ELISA in BALF fluid.

Data represent average ± SEM from five mice per group. *$p < 0.05$; **$p < 0.005$; ***$p < 0.0005$. Results are representative of two independent experiments. See also Figure S1 and Figure S7.

treatment consistently inhibits the production of IFN-α in the lung early after influenza virus infection (Figure 1, Figure 2, Figure 5, and Figure 6). Therefore, we postulated that blunting IFN-α production may be a mechanism by which CYM-5442 inhibits cytokine storm early after influenza virus infection. To address this, we infected IFNα/β receptor knockout mice with influenza virus, treated these mice with either vehicle or CYM-5442, and measured innate cell recruitment and cytokine/chemokine production 48 hr postinfection. Significant differences were not observed in inflammatory cell recruitment of macrophages/monocytes, neutrophils, and NK cells into the lung of IFNα/β receptor-deficient mice when compared to C57Bl/6J mice (Figure 7C). As seen with infected C57BL/6J mice, CYM-5442 treatment inhibited innate immune cell recruitment in IFNα/β receptor-deficient mice 48 hr postinfection (Figure 7C). Despite recruitment of innate immune cells into the lung of IFNα/β receptor-deficient mice 48 hr postinfection, these mice exhibited significantly reduced levels of IFN-α, CCL2, IL-6, and IFN-γ (Figure 7D) in addition to CCL5 and CXCL10 in the BALF fluid (Figure S1E). Taken together, our data demonstrate that regulation of cellular recruitment into the pulmonary tissue is mediated by endothelial cells and is independent of type I interferon signaling. Moreover, $S1P_1$ receptor agonism of endothelial cells inhibits IFN-α production and results in the dampening of global pro-inflammatory cytokine responses.

DISCUSSION

The recruitment of innate immune cells into the lungs combined with excessive proinflammatory cytokine and chemokine production are hallmarks of influenza virus infection (La Gruta et al., 2007). Contrary to current dogma, which places lung epithelium and inflammatory infiltrate at the center of influenza virus-induced cytokine storm (La Gruta et al., 2007), we demonstrate here a central role for pulmonary endothelium in regulating this process. Furthermore, we determined that both innate immune cell recruitment and early innate cytokine and chemokine production are uncoupled events, with endothelial cells at the center of both processes. The chemical and genetic approaches shown here have a potentially broad impact on

understanding disease pathogenesis. Early dysregulation of innate cellular and cytokine responses predict disease severity and death during highly pathogenic influenza virus infection (Bermejo-Martin et al., 2009; Cillóniz et al., 2009; de Jong et al., 2006; Kobasa et al., 2004, 2007). The early induction of the cytokines IFN-α, TNF-α, IL-1α, and IL-6 and the chemokines CCL2, CCL3, CXCL2 (IL-8), and CXCL10 are associated with symptom formation in humans (Hayden et al., 1998; Kaiser et al., 2001). TNF-α, IL-1, and IL-6 possess multifunctional activities and are associated with morbidity during influenza virus infection. Chemokines such as CCL2, CCL3, CXCL2, and CXCL10 induce the recruitment of innate immune cells into the lung, which can release more cytokines exacerbating cytokine storm and further damage the lung. We demonstrate that suppression of early innate immune responses through $S1P_1$ signaling on endothelial cells results in significantly reduced mortality during infection of mice with a human pathogenic strain of influenza virus (Figure 3). Importantly, our results identify a pulmonary cell type, endothelial cells, as potential targets for suppressing excessive innate inflammatory responses.

The ability of CYM-5442 to diminish the production of IFN-α early following influenza virus infection is broadly relevant. The requirement of type I interferon signaling for the early production of multiple cytokines and chemokines after influenza virus infection is striking. Moreover, production of proinflammatory cytokines requires type I interferon signaling, but the type I interferon pathway is dispensable for cell recruitment. Though type I interferon signaling is well known to inhibit viral replication (García-Sastre and Biron, 2006), evidence also points to pathogenic roles for IFN-α during viral infection. Several proinflammatory cytokines and chemokines are downstream of type I interferon receptor signaling. Moreover, disease onset correlates directly with local respiratory production of IFN-α in humans (Hayden et al., 1998). Thus, type I interferon signaling may play dual roles in viral pathogenesis and viral clearance. $S1P_1$ receptor signaling on pulmonary endothelial cells in vivo blunts but does not abolish IFN-α production and may explain why pathology is lessened without compromising the host's ability to clear the virus.

The ability of $S1P_1$ receptor activation to suppress cytokine and chemokine production in the lung in irradiated mice, coupled to the pattern of $S1P_1$ receptor expression in the lung, strongly indicates a central endothelial component to early innate immune responses. An interesting question is whether lung lymphatic, vascular, or both endothelial cell populations are directly responsible for the observed immune modulation. Moreover, of the large receptor reserve in blood, endothelium ensures that $S1P_1$ receptor expression is maintained on both basolateral and luminal surfaces of the targeted endothelial cell populations (Cahalan et al., 2011) and retains a responsiveness to agonist. Additional approaches to separate blood and lymphatic contributions are still needed. Though we know that endothelial cells produce chemokines, endothelial-mediated regulation of cytokine production in the lung still needs to be determined. It is possible that endothelial cells may regulate cytokine production in the lung through a complex crosstalk mechanism with lung epithelium or resident hematopoietic cells. Nevertheless, the identification of endothelial cells as central orchestrators of early innate immune-mediated inflammation is of fundamental interest and has broad implications for treating multiple diseases. The contribution of aberrant proinflammatory cytokine and chemokine production to pathogenesis has been reported for viral and bacterial diseases, including HIV (Stacey et al., 2009), Hantavirus (Borges et al., 2006), severe acute respiratory syndrome (SARS) (Thiel and Weber, 2008), and pneumococcal bacterial pneumonia (Bergeron et al., 1998). Further, the etiology of several autoimmune conditions has been directly associated with excessive innate immune responses (Kawane et al., 2010; Link, 1998). Thus, understanding the cellular pathways that regulate cytokine storm and developing appropriate chemical signaling tools to identify pathophysiological points of control not only provide insight into microbial-host interactions, but also may ultimately reveal additional approaches to achieve effective immunotherapy in multiple diseases. In addition, our results present a nonlymphopenic mechanism by which sphingosine analog treatment suppresses pathogenic immune responses. Lastly, these data suggest that endogenous S1P acting on endothelial $S1P_1$ receptor could be a negative regulator of cytokine amplification and raise the possibility that heterogeneities in S1P metabolism between individuals could contribute to the advantages or disadvantages conferred by genetic individuality to host survival in a number of diseases.

EXPERIMENTAL PROCEDURES

Mice, Virus, Compounds, and Reagents

Six- to eight-week-old C57Bl/6 male mice and $S1P_1$-eGFP mice described elsewhere (Cahalan et al., 2011) were bred and maintained in a closed breeding facility at The Scripps Research Institute. Influenza A/WSN/33 (WSN; H1N1) and the human H1N1 2009 isolate A/Wisconsin/WSLH34939/09 (a kind gift from Yoshihiro Kawaoka, University of Madison, WI) were amplified and plaqued on Madin-Darby Canine Kidney (MDCK) cells. Mice were infected intratracheally (i.t.) with 1×10^4 PFU of influenza A/WSN/33 virus or intranasally with 1×10^5 PFU of A/Wisconsin/WSLH34939/09 under isoflurane anesthesia. At 1 hr postinfection, mice were anesthetized by isoflurane inhalation for i.t. delivery of vehicle (100 ul of water), AAL-R (0.2 mg/kg dissolved in water), CYM-5442 (2 mg/kg dissolved in water), or RP-002 (3 mg/kg i.t. or 6 mg/kg orally dissolved in water). Multiple doses of compound were administered at the specific times listed in the figurte legends. (AAL-R and CYM-5442 were synthesized according to published methods (Jo et al., 2005). RP-002, (R)-2-(4-(5-(3-cyano-4-isopropoxyphenyl)-l,2,4-oxadiazol-3-yl)-2,3-dihydro-lH-inden-l-ylamino)-N,N-dimethylacetamide hydrochloride, was synthesized according to the published method (Martinborough et al., 2011). The compound had an EC_{50} for S1P1 of 0.13 nM, was > 100-fold selective versus S1P5, and was 10,000-fold selective versus S1P2, 3, and 4 respectively, when assayed as described (Cahalan et al., 2011). Recombinant murine CCL2 (rmCCL2) was purchased from Shenandoah Biotechnology Inc. (Warwick, PA).

Cytokine and Chemokine Analysis

The trachea of euthanized mice was exposed, transected, and incubated with a blunt 18 gauge needle. One milliliter of phosphate-buffered saline supplemented with Complete Mini, EDTA-free Protease Inhibitor Cocktail (Roche) was infused and recovered four times. The recovered bronchoalveolar lavage fluid was spun at 3000 × *g* for 3 min at 4°C and stored at −80°C until use. Multiplex ELISA was performed on supernatant by Quansys Biosciences (Logan, UT) to detect IL-1α, IL-1β, IL-2, IL-3, IL-4, IL-5, IL-6, IL-9, IL-10, IL-12p70, TNF-α, MIP-1α, MCP-1, GM-CSF, IFN-γ, and RANTES. ELISAs were also performed using CCL2 (MCP-1), CCL5 (RANTES), CXCL10 (IP-10), IL-1α, IL-6, TNF-α, and IFN-γ Duoset kits (R&D systems) as well as the *VeriKine* Mouse Interferon-Alpha and Interferon-Beta ELISA Kits (Pestka Biomedical Laboratories, Inc). For the quantification of mRNA chemokine expression, lung

endothelial cells were FACS purified to > 90%–95% purity, and mRNA was purified using the RNeasy mini kit from QIAGEN. Prior to real-time PCR, genomic DNA was digested and cDNA was made using the RT^2 First Strand Kit (C-03) according to manufacturer's instructions (SA Biosciences, Frederick, MD). cDNA (200 ng) was added to individual wells and quantified using the RT^2 profiler PCR Array according to the manufacturer's instructions (SA Biosciences, Frederick, MD).

Cellular Analysis by Flow Cytometry

Lungs were harvested from PBS-perfused mice and mechanically diced into small tissue pieces using surgical scissors. Diced lungs were suspended in 4 ml of CDTI buffer (0.5 mg/ml collagenase from Clostridium histolyticum type IV [Sigma], 0.1 mg/ml Dnase I from bovine pancreas grade II [Roche], 1 mg/ml trypsin inhibitor type Ii-s [Sigma] in DMEM) for 1 hr at 37°C. Lung was then disrupted mechanically through a 100 μm filter, and red blood cells were lysed using red blood cell lysis buffer (0.02 Tris-HCL and 0.14 NH_4Cl). Inflammatory cells were purified by centrifugation in 35% PBS-buffered Percoll (GE Healthcare Life Sciences) at 1,500 rpm for 15 min. Cell pellets were resuspended in staining buffer, and Fc receptors were blocked using 25 μg/ml anti-mouse CD16/32 (BD Biosciences). Cells were stained with the following anti-mouse antibodies: AlexaFluor 488-conjugated gp38 (eBioscience; clone eBio8.1.1), PE-conjugated (BioLegend, Inc.; clone ME13.3) and APC-conjugated (eBioscience; clone 390) CD31, PE-Cy7-conjugated EpCAM (BioLegend, Inc.; clone 68.8), Pacific blue-conjugated CD45.2 (BioLegend, Inc.; clone 104), PerCP-Cy5.5-conjugated NK1.1 (BD Biosciences; clone PK136), PE-Cy7-conjugated CD3e (eBioscience; clone 145-2C11), e450-conjugated CD4 (eBioscience; clone L3T4), PE-conjugated CD8α (BD Biosciences; clone 53-6.1), Pacific blue-conjugated B220 (BD Biosciences; clone RA3-6B2), PE-conjugated CD19 (BD Biosciences, clone 1D3), PE-Cy7-conjugated CD11b (eBiosciences; clone M1/70), PerCP-Cy5.5-conjugated CD11c (eBiosciences; clone N418), APC-conjugated Gr-1 (BD Biosciences; clone RB6-8C5), Pacific blue and PE-conjugated Ly6G (BD Biosciences; clone IA8), APC-conjugated F480 (eBioscience; clone BM8), PE-conjugated 7/4 (AbD Serotec; clone 7/4), PE-conjugated I/A-I/E (BD Biosciences; clone M5/114.15.2), PE-Cy7 conjugated CD205 (eBiosciences; clone 205yekta), Fitc-conjugated CD69 (BD Biosciences; clone H1.2F3), and APC-conjugated CD25 (eBiosciences; clone PC61.5). Flow cytometry acquisition was performed with BD FACSDiva-driven BD LSR II flow cytometer (Becton, Dickinson and Company). Data were then analyzed with FlowJo software (Treestar Inc.).

Western Blot

FACS-purified lung cell populations (1×10^5 cells) using the antibodies described above were homogenized in RIPA buffer supplemented with protease inhibitors (Pierce). Lysates were centrifuged at 50,000 × g for 30 min, and the protein concentration in the supernatant was determined by BCA assay (Pierce). Equal amounts of protein from cell lysates were loaded in nondenaturing conditions and separated by SDS-PAGE in 4%–12% NuPAGE (Novex) Bis-Tris gels. Gels were transferred to PVDF membranes followed by probing for GFP using an anti-GFP antibody (Abcam) and an anti-rabbit Ig light-chain HRP secondary (ELC Biosciences) using chemiluminescence autoradiography.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and can be found with this article online at doi:10.1016/j.cell.2011.08.015.

ACKNOWLEDGMENTS

This is Publication Number 21112 from the Department of Immunology and Microbial Science and the Department of Chemical Physiology and The Scripps Research Institute Molecular Screening Center, The Scripps Research Institute (TSRI). This work was supported, in part, by USPHS grants AI074564 (M.B.A.O., H.R., K.B.W., and J.R.T.), AI009484 (M.B.A.O.), AI05509 (H.R.), MH084512 (H.R.), and NIH training grants NS041219 (K.B.W.), AI007244 (K.B.W.), and AI007364 (J.R.T). We thank Marcus Boehm, Li-ming Huang, and Bryan Clemons (Receptos, Inc.) for helping provide RP-002 as a chemical tool. Hugh Rosen is a founder of Receptos. Edward Roberts is a consultant to Receptos. Fiona Scott, Esther Martinborough, and Robert Peach are employees of Receptos.

Received: February 23, 2011
Revised: March 27, 2011
Accepted: August 13, 2011
Published: September 15, 2011

REFERENCES

Bergeron, Y., Ouellet, N., Deslauriers, A.M., Simard, M., Olivier, M., and Bergeron, M.G. (1998). Cytokine kinetics and other host factors in response to pneumococcal pulmonary infection in mice. Infect. Immun. *66*, 912–922.

Bermejo-Martin, J.F., Ortiz de Lejarazu, R., Pumarola, T., Rello, J., Almansa, R., Ramírez, P., Martin-Loeches, I., Varillas, D., Gallegos, M.C., Serón, C., et al. (2009). Th1 and Th17 hypercytokinemia as early host response signature in severe pandemic influenza. Crit. Care *13*, R201.

Borges, A.A., Campos, G.M., Moreli, M.L., Souza, R.L., Aquino, V.H., Saggioro, F.P., and Figueiredo, L.T. (2006). Hantavirus cardiopulmonary syndrome: immune response and pathogenesis. Microbes Infect. *8*, 2324–2330.

Brinkmann, V., Billich, A., Baumruker, T., Heining, P., Schmouder, R., Francis, G., Aradhye, S., and Burtin, P. (2010). Fingolimod (FTY720): discovery and development of an oral drug to treat multiple sclerosis. Nat. Rev. Drug Discov. *9*, 883–897.

Cahalan, S.M., Gonzalez-Cabrera, P.J., Sarkisyan, G., Nguyen, N., Schaeffer, M.T., Huang, L., Yeager, A., Clemons, B., Scott, F., and Rosen, H. (2011). Actions of a picomolar short-acting $S1P_1$ agonist in $S1P_1$-eGFP knock-in mice. Nat. Chem. Biol. *7*, 254–256.

Cillóniz, C., Shinya, K., Peng, X., Korth, M.J., Proll, S.C., Aicher, L.D., Carter, V.S., Chang, J.H., Kobasa, D., Feldmann, F., et al. (2009). Lethal influenza virus infection in macaques is associated with early dysregulation of inflammatory related genes. PLoS Pathog. *5*, e1000604.

Dawson, T.C., Beck, M.A., Kuziel, W.A., Henderson, F., and Maeda, N. (2000). Contrasting effects of CCR5 and CCR2 deficiency in the pulmonary inflammatory response to influenza A virus. Am. J. Pathol. *156*, 1951–1959.

de Jong, M.D., Simmons, C.P., Thanh, T.T., Hien, V.M., Smith, G.J., Chau, T.N., Hoang, D.M., Chau, N.V., Khanh, T.H., Dong, V.C., et al. (2006). Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. Nat. Med. *12*, 1203–1207.

García-Sastre, A. (2010). Influenza virus receptor specificity: disease and transmission. Am. J. Pathol. *176*, 1584–1585.

García-Sastre, A., and Biron, C.A. (2006). Type 1 interferons and the virus-host relationship: a lesson in détente. Science *312*, 879–882.

Gonzalez-Cabrera, P.J., Jo, E., Sanna, M.G., Brown, S., Leaf, N., Marsolais, D., Schaeffer, M.T., Chapman, J., Cameron, M., Guerrero, M., et al. (2008). Full pharmacological efficacy of a novel S1P1 agonist that does not require S1P-like headgroup interactions. Mol. Pharmacol. *74*, 1308–1318.

Hayden, F.G., Fritz, R., Lobo, M.C., Alvord, W., Strober, W., and Straus, S.E. (1998). Local and systemic cytokine responses during experimental human influenza A virus infection. Relation to symptom formation and host defense. J. Clin. Invest. *101*, 643–649.

Itoh, Y., Shinya, K., Kiso, M., Watanabe, T., Sakoda, Y., Hatta, M., Muramoto, Y., Tamura, D., Sakai-Tagawa, Y., Noda, T., et al. (2009). In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. Nature *460*, 1021–1025.

Jo, E., Sanna, M.G., Gonzalez-Cabrera, P.J., Thangada, S., Tigyi, G., Osborne, D.A., Hla, T., Parrill, A.L., and Rosen, H. (2005). S1P1-selective in vivo-active agonists from high-throughput screening: off-the-shelf chemical probes of receptor interactions, signaling, and fate. Chem. Biol. *12*, 703–715.

Kaiser, L., Fritz, R.S., Straus, S.E., Gubareva, L., and Hayden, F.G. (2001). Symptom pathogenesis during acute influenza: interleukin-6 and other cytokine responses. J. Med. Virol. *64*, 262–268.

Kawane, K., Tanaka, H., Kitahara, Y., Shimaoka, S., and Nagata, S. (2010). Cytokine-dependent but acquired immunity-independent arthritis caused by DNA escaped from degradation. Proc. Natl. Acad. Sci. USA *107*, 19432–19437.

Kobasa, D., Jones, S.M., Shinya, K., Kash, J.C., Copps, J., Ebihara, H., Hatta, Y., Kim, J.H., Halfmann, P., Hatta, M., et al. (2007). Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. Nature *445*, 319–323.

Kobasa, D., Takada, A., Shinya, K., Hatta, M., Halfmann, P., Theriault, S., Suzuki, H., Nishimura, H., Mitamura, K., Sugaya, N., et al. (2004). Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. Nature *431*, 703–707.

La Gruta, N.L., Kedzierska, K., Stambas, J., and Doherty, P.C. (2007). A question of self-preservation: immunopathology in influenza virus infection. Immunol. Cell Biol. *85*, 85–92.

Link, H. (1998). The cytokine storm in multiple sclerosis. Mult. Scler. *4*, 12–15.

Marsolais, D., Hahm, B., Edelmann, K.H., Walsh, K.B., Guerrero, M., Hatta, Y., Kawaoka, Y., Roberts, E., Oldstone, M.B., and Rosen, H. (2008). Local not systemic modulation of dendritic cell S1P receptors in lung blunts virus-specific immune responses to influenza. Mol. Pharmacol. *74*, 896–903.

Marsolais, D., Hahm, B., Walsh, K.B., Edelmann, K.H., McGavern, D., Hatta, Y., Kawaoka, Y., Rosen, H., and Oldstone, M.B. (2009). A critical role for the sphingosine analog AAL-R in dampening the cytokine response during influenza virus infection. Proc. Natl. Acad. Sci. USA *106*, 1560–1565.

Martinborough, E., Boehm, M., Yeager, A., Yamiyo, J., Huang, L., Brahmachary, E., Moorjani, M., Timony, G., Brooks, J., Peach, R., et al. (2011). Selective sphingosine 1-phosphate receptor modulators and methods of chiral synthesis, World International Patent Organization: WO/2011/060392 p. 92.

Openshaw, P.J., and Dunning, J. (2010). Influenza vaccination: lessons learned from the pandemic (H1N1) 2009 influenza outbreak. Mucosal Immunol. *3*, 422–424.

Rivera, J., Proia, R.L., and Olivera, A. (2008). The alliance of sphingosine-1-phosphate and its receptors in immunity. Nat. Rev. Immunol. *8*, 753–763.

Rosen, H., and Gordon, S. (1987). Monoclonal antibody to the murine type 3 complement receptor inhibits adhesion of myelomonocytic cells in vitro and inflammatory cell recruitment in vivo. J. Exp. Med. *166*, 1685–1701.

Rosen, H., and Liao, J. (2003). Sphingosine 1-phosphate pathway therapeutics: a lipid ligand-receptor paradigm. Curr. Opin. Chem. Biol. *7*, 461–468.

Rosen, H., Sanna, M.G., Cahalan, S.M., and Gonzalez-Cabrera, P.J. (2007). Tipping the gatekeeper: S1P regulation of endothelial barrier function. Trends Immunol. *28*, 102–107.

Rosen, H., Gonzalez-Cabrera, P., Marsolais, D., Cahalan, S., Don, A.S., and Sanna, M.G. (2008). Modulating tone: the overture of S1P receptor immunotherapeutics. Immunol. Rev. *223*, 221–235.

Rosen, H., Gonzalez-Cabrera, P.J., Sanna, M.G., and Brown, S. (2009). Sphingosine 1-phosphate receptor signaling. Annu. Rev. Biochem. *78*, 743–768.

Sanchez, T., Estrada-Hernandez, T., Paik, J.H., Wu, M.T., Venkataraman, K., Brinkmann, V., Claffey, K., and Hla, T. (2003). Phosphorylation and action of the immunomodulator FTY720 inhibits vascular endothelial cell growth factor-induced vascular permeability. J. Biol. Chem. *278*, 47281–47290.

Sanna, M.G., Wang, S.K., Gonzalez-Cabrera, P.J., Don, A., Marsolais, D., Matheu, M.P., Wei, S.H., Parker, I., Jo, E., Cheng, W.C., et al. (2006). Enhancement of capillary leakage and restoration of lymphocyte egress by a chiral S1P1 antagonist in vivo. Nat. Chem. Biol. *2*, 434–441.

Stacey, A.R., Norris, P.J., Qin, L., Haygreen, E.A., Taylor, E., Heitman, J., Lebedeva, M., DeCamp, A., Li, D., Grove, D., et al. (2009). Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. J. Virol. *83*, 3719–3733.

Thiel, V., and Weber, F. (2008). Interferon and cytokine responses to SARS-coronavirus infection. Cytokine Growth Factor Rev. *19*, 121–132.

Tscherne, D.M., and García-Sastre, A. (2011). Virulence determinants of pandemic influenza viruses. J. Clin. Invest. *121*, 6–13.

LETTER

doi:10.1038/nature10002

Determinants of nucleosome organization in primary human cells

Anton Valouev[1], Steven M. Johnson[2], Scott D. Boyd[1], Cheryl L. Smith[1], Andrew Z. Fire[1,3] & Arend Sidow[1,3]

Nucleosomes are the basic packaging units of chromatin, modulating accessibility of regulatory proteins to DNA and thus influencing eukaryotic gene regulation. Elaborate chromatin remodelling mechanisms have evolved that govern nucleosome organization at promoters, regulatory elements, and other functional regions in the genome[1]. Analyses of chromatin landscape have uncovered a variety of mechanisms, including DNA sequence preferences, that can influence nucleosome positions[2–4]. To identify major determinants of nucleosome organization in the human genome, we used deep sequencing to map nucleosome positions in three primary human cell types and *in vitro*. A majority of the genome showed substantial flexibility of nucleosome positions, whereas a small fraction showed reproducibly positioned nucleosomes. Certain sites that position *in vitro* can anchor the formation of nucleosomal arrays that have cell type-specific spacing *in vivo*. Our results unveil an interplay of sequence-based nucleosome preferences and non-nucleosomal factors in determining nucleosome organization within mammalian cells.

Previous studies in model organisms[3–7] as well as initial analyses in human cells[8] have identified fundamental aspects of nucleosome organization. Here we focus on the dynamic relationships between sequence-based nucleosome preferences and chromatin regulatory function in primary human cells. We mapped tissue-specific and DNA-encoded nucleosome organization across granulocytes and two types of T cells ($CD4^+$ and $CD8^+$) isolated from the blood of a single human donor, by isolating cellular chromatin and treating it with micrococcal nuclease (MNase) followed by deep sequencing of the resulting nucleosome-protected fragments (Methods, Supplementary Fig. 1). To provide sufficient depth for both local and global analyses, we used high-throughput SOLiD technology, generating 584, 342 and 343 million mapped reads for granulocytes, $CD4^+$ and $CD8^+$ T cells, respectively. These are equivalent to 16–28× genome coverage by 147 bp nucleosome footprints (cores; see Methods). The depth of sequence was critical for our subsequent analysis: although shallower coverage can illuminate features of nucleosome positions through statistical analysis (for example, refs 6, 8), any definitive map and thus comparison of static and dynamic positioning requires high sequence coverage throughout the genome.

To provide complementary data on purely sequence-driven nucleosome positioning in the absence of cellular influences, we reconstituted genomic DNA *in vitro* with recombinantly derived histone octamers to produce *in vitro* nucleosomes (Methods, Supplementary Fig. 2), and generated over 669 million mapped reads, representing 32× core coverage of the genome. To identify primary nucleosome positioning sites in DNA, the reconstitution was performed under conditions of DNA excess (see methods). We also generated a control data set of 321 million mapped reads from MNase-digested naked DNA. In the population of granulocytes (our deepest *in vivo* data set), over 99.5% of the mappable genome is engaged by nucleosomes (Methods), and 50 percent of nucleosome-depleted bases occur in regions shorter than 160 bp.

We first focused on global patterns of nucleosome positioning and spacing by calculating fragment distograms and phasograms[6,7,9]. Distograms (histograms of distances between mapped reads' start positions aligning in opposing orientation, Supplementary Fig. 3a) reveal the average core fragment size as a peak if there are many sites in the genome that contain consistently positioned nucleosomes. A positioning signal that is strongly amplified by conditioning the analysis on sites with three or more read starts (reflecting a positioning preference; 3-pile subset), is present not only *in vivo* (Fig. 1a), but also *in vitro* (Fig. 1b), demonstrating that many genomic sites bear intrinsic, sequence-driven, positioning signals. Phasograms (histograms of distances between mapped reads' start positions aligning in the same orientation, Supplementary Fig. 3b) reveal consistent spacing of positioned nucleosomes by exhibiting a wave-like pattern with a period that represents genome-average internucleosome spacing. In granulocytes, the wave peaks are 193 bp apart (Fig. 1c, adjusted $R^2 = 1$, P-value $< 10^{-15}$), which, given a core fragment length of 147 bp, indicates an internucleosome linker length of 46 bp. By contrast, the phasograms of both types of T cells have spacing that is wider by 10 bp (Fig. 1d), equivalent to a 56 bp average linker length. These results are consistent with classical observations of varying nucleosome phases in different cell types[10,11]. Linker length differences have been tied to differences in linker histone gene expression[12,13], which we found to be 2.4 times higher in T cells compared to granulocytes (84 reads per kilobase of mature transcript per million mapped reads (RPKM)[14] vs 35 RPKM). The *in vitro* phasogram (Fig. 1e) reveals no detectable stereotypic spacing of positioned nucleosomes, demonstrating a lack of intrinsic phasing among DNA-encoded nucleosome positioning sites.

Using a positioning stringency metric (Methods; Supplementary Fig. 4) that quantifies the fraction of defined nucleosome positions within a given segment, we calculated the fraction of the genome that is occupied by preferentially positioned nucleosomes at different stringency thresholds. The maximum number of sites at which some positioning preference can be detected statistically is 120 million, covering just over 20% of the genome (Supplementary Fig. 5) at the low stringency of 23%. Thus, the majority of nucleosome positioning preferences is weak, and nucleosomes across the majority of the human genome are not preferentially positioned, either by sequence or by cellular function.

Next we focused on how transcription and chromatin functions affect nucleosome organization regionally. For each cell type, we generated deep RNA-seq data and binned genes into groups according to their expression levels. The average spacing of nucleosomes was greatest within silent genes ($CD4^+$ T cells, 206 bp, Fig. 2a) and decreased by as much as 11 bp as the expression levels went up (t-statistic P-value $= 6.5 \times 10^{-34}$). This suggests that transcription-induced cycles of nucleosome eviction and reoccupation cause denser packing of nucleosomes and slight reduction in nucleosome occupancy (Supplementary Fig. 6). On the basis of this result, we hypothesized that higher-order chromatin organization as implied by specific

[1]Department of Pathology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305, USA. [2]Department of Microbiology and Molecular Biology, Brigham Young University, 757 WIDB, Provo, Utah 84602-5253, USA. [3]Department of Genetics, Stanford University School of Medicine, Pasteur Drive, Stanford, California 94305, USA.

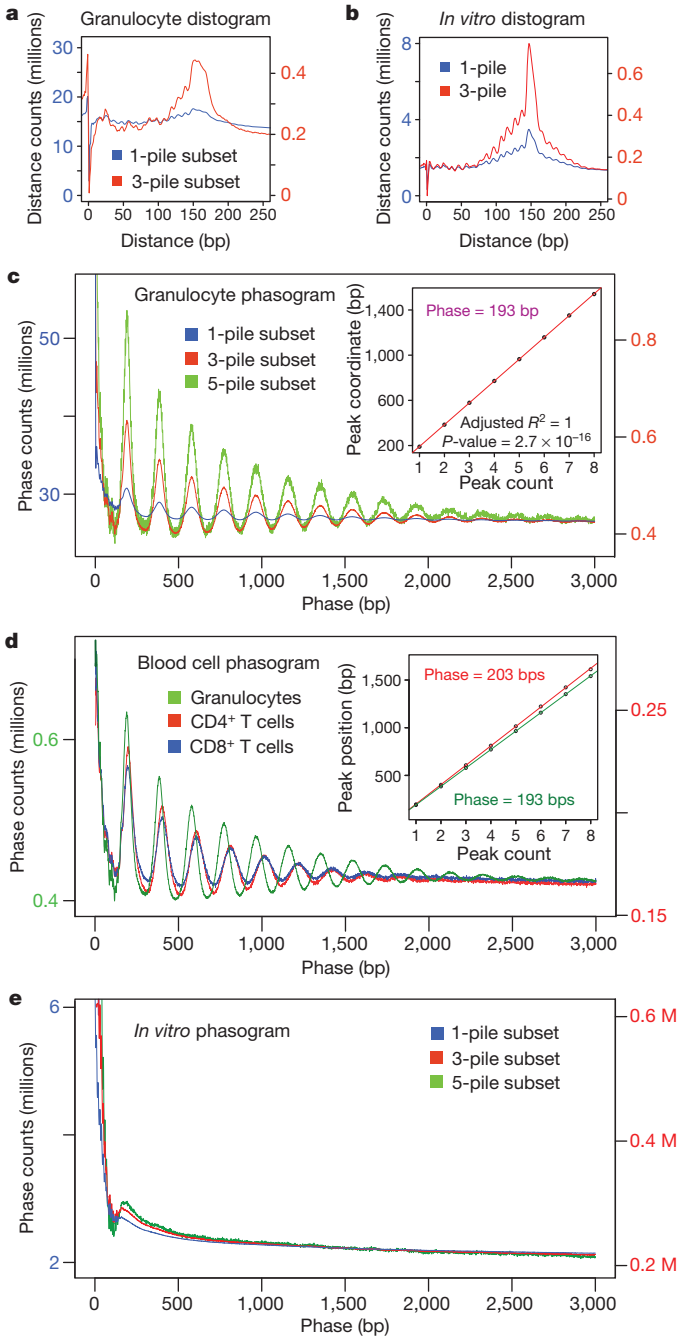


Figure 1 | Global parameters of cell-specific nucleosome phasing and positioning in human. **a**, *In vivo* granulocyte distogram (calculation explained in Supplementary Fig. 3a). *x*-axis represents the range of recorded distances. *y*-axis represents frequencies of observed distances within 1-pile (blue) and 3-pile (red) subsets. 1-pile subset represents the entire data set, 3-pile subset represents a subset of sites containing three or more coincident read starts. **b**, Distogram of the *in vitro* reconstituted nucleosomes showing 1-pile and 3-pile subsets as in (**a**). **c**, *In vivo* granulocyte phasogram (calculation explained in Supplementary Fig. 3b). *x*-axis shows the range of recorded phases. *y*-axis shows frequencies of corresponding phases. Phasograms of 1-pile, 3-pile and 5-pile subsets are plotted. Inset, linear fit to the positions of the phase peaks within 3-pile subsets (slope = 193 bp). **d**, Phasograms of blood cell types. Inset, linear fits in CD4$^+$ T cells (203 bp) and granulocytes (193 bp). **e**, Phasograms of 1-pile, 3-pile and 5-pile subsets in the *in vitro* data.

chromatin modifications might be associated with specific spacing patterns. Using previously published ChIP-seq data, we identified regions of enrichment[15] for histone modifications that are found within heterochromatin (H3K27me3, H3K9me3)[16], gene-body euchromatin

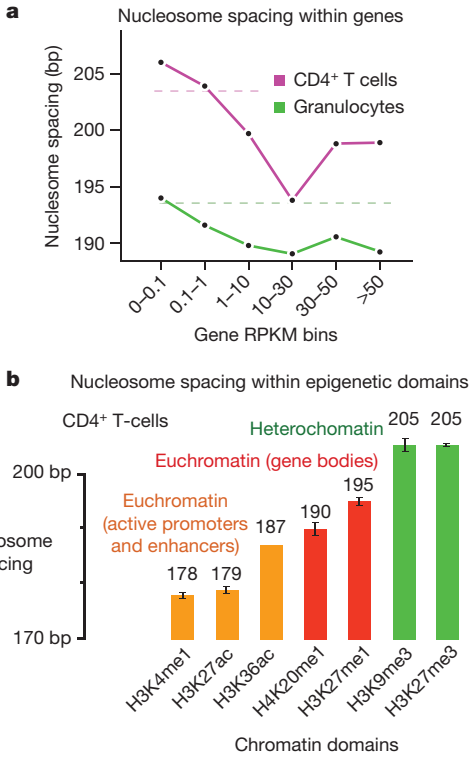


Figure 2 | Transcription and chromatin modification-dependent nucleosome spacing. **a**, Nucleosome spacing as a function of transcriptional activity. *x*- axis represents gene expression values binned according to RPKM values. Internucleosome spacing is plotted along the *y*-axis. Dashed lines represent genome-wide average spacing for each cell type. **b**, Nucleosome spacing within genomic regions marked by specific histone marks in CD4$^+$ T cells. Bar height plots estimated nucleosome spacing for each histone modification. Bar colours differentiate chromatin types (euchromatin vs heterochromatin).

(H4K20me1, H3K27me1)[16], or euchromatin associated with promoters and enhancers (H3K4me1, H3K27ac, H3K36ac)[17], and estimated spacing of nucleosomes for each of these epigenetic domains. We found that active promoter-associated domains contained the shortest spacing of 178–187 bp, followed by a larger spacing of 190–195 bp within the body of active genes, whereas heterochromatin spacing was largest at 205 bp (Fig. 2b). These results reveal striking heterogeneity in nucleosome organization across the genome that depends on global cellular identity, metabolic state, regional regulatory state, and local gene activity.

To characterize DNA signals responsible for consistent positioning of nucleosomes, we identified 0.3 million sites occupied *in vitro* by nucleosomes at high stringency (>0.5; Methods). The region occupied by the centre of the nucleosome (dyad) exhibits a significant increase in G/C usage (Poisson *P*-value $< 10^{-100}$; Fig. 3a). Flanking regions increase in A/T usage as the positioning strength increases (Fig. 3b). A subset of *in vitro* positioned nucleosomes (stringency > 0.5) which are also strongly positioned *in vivo* (stringency > 0.4) revealed increased A/T usage within the flanks (Fig. 3c) compared to *in vitro*-only positioning sites (Fig. 3a), which underscores the importance of flanking repelling elements for positioning *in vivo*. We term such elements with strong G/C cores and A/T flanks 'container sites' to emphasize the proposed positioning mechanism (Fig. 3d). This positioning signal is different from a 10-bp dinucleotide periodicity observed in populations of nucleosome core segments isolated from a variety of species[18,19] and proposed to contribute to precise positioning and/or rotational setting of DNA on nucleosomes[19] on a fine scale (Supplementary Fig. 7). G/C-rich signals are known to promote nucleosome occupancy[20,21], whereas AA-rich sequences repel nucleosomes[4], and our data demonstrate that precise arrangement of a core-length attractive segment flanked by repelling sequences can produce a strongly positioned nucleosome (Fig. 3d).

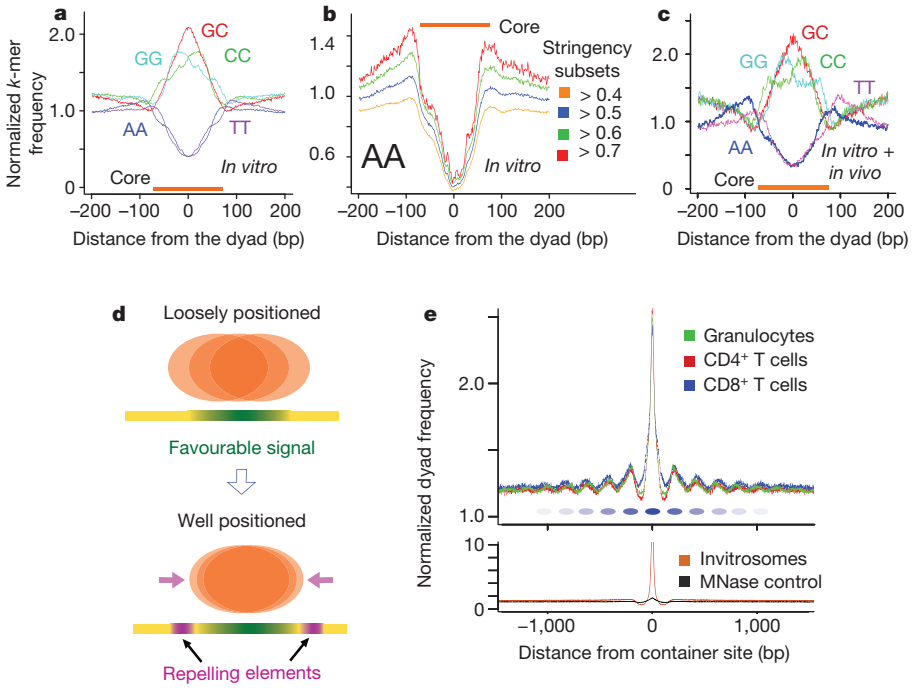


Figure 3 | Sequence signals that drive nucleosome positioning. **a**, Sequence signals within sites containing moderately positioned *in vitro* nucleosomes (stringency > 0.5). Distance from the positioned dyad to a given dinucleotide is plotted along the *x*-axis; *y*-axis represents frequency of a given *k*-mer divided by its genome-wide expectation. The 147-bp footprint of a nucleosome is indicated by an orange band. **b**, Changes in AA dinucleotide usage with increasing positioning stringency. *x* and *y* axes same as in (**a**). Curves of AA usage within the sites of increasingly positioned dyads are shown (stringency cutoffs of 0.4, 0.5, 0.6, 0.7). **c**, Sequence signals within sites containing *in-vitro*-positioned nucleosomes (stringency > 0.5) that also have high *in vivo* stringency (stringency > 0.4). *x* and *y* axes same as in (**a**). **d**, Schematic depiction of the container site positioning mechanism. The C/G-rich core area (green) favours occupancy, but does not precisely position the nucleosome (top). Adding flanking A/T-rich repelling elements (purple, bottom) restricts the position of the nucleosome. **e**, Nucleosome organization around container sites *in vivo* and *in vitro*. *x*-axis represents distances from the dyads to container sites (based on 300,000 container sites). Frequencies of nucleosome dyads around those sites are plotted along the *y*-axis. The upper plot shows distribution of *in vivo* dyads across $CD4^+$ cells, $CD8^+$ cells and granulocytes. The ovals depict hypothetical nucleosome positions across the site with colour intensities reflecting their positioning strength. The lower plot shows distribution of dyads *in vitro* and in MNase control.

Dyad frequencies around container sites (Fig. 3e) show a strong peak of enrichment *in vivo*, confirming that DNA positions nucleosomes *in vivo* over these sites. Additionally, wave-like patterns emanate from these sites *in vivo* (but not *in vitro*), reflecting the nucleation of phased arrays by positioned cellular nucleosomes. Viewing these results in light of the nucleosome barrier model[22], which proposes that nucleosomes are packed into positioned and phased arrays against a chromatin barrier, we conclude that sequence-positioned nucleosome can initiate propagation of adjacent stereotypically positioned nucleosomes. Importantly, wave periods around container sites are shorter in granulocytes than in T cells, allowing tissue-specific variation in linker length (Fig. 1d) to alter placement of nucleosomes over distances of as much as 1 kilobase from an initial container site. Functional consequences of such rearrangements might include global shifts in regulatory properties that could contribute to distinct transcription factor accessibility profiles in different cell types.

The cellular environment can drive nucleosomes to sequences not intrinsically favourable to being occupied, as is evident in a genome-wide comparison of observed nucleosome coverage of all possible tetranucleotides between the granulocyte and the *in vitro* data (Fig. 4a). *In vitro*, nucleosome occupancy is strongly associated with AT/GC content, but this preference is abolished *in vivo*; the exception are C/G rich tetramers that contain CpG dinucleotides, which show a 30% reduction in apparent nucleosome occupancy despite having high core coverage *in vitro*. Consistent with this, CpG islands are fivefold depleted for observed nucleosome coverage *in vivo* (Fig. 4b). No such decrease is observed in the *in vitro* data set.

The decreased nucleosome occupancy of promoters could be due to promoter-related functions of mammalian CpG islands, similar to promoter-associated nucleosome-free regions observed in flies[23] and yeast[5], which do not have CpG islands. We therefore analysed transcription-dependent nucleosome packaging around promoters. As in other organisms[23–27], promoters of active genes have a nucleosome-free region (NFR) of about 150 bp overlapping the transcriptional start site and arrays of well-positioned and phased nucleosomes that radiate from the NFR (Fig. 4c). A notable reduction in apparent nucleosome occupancy extends up to 1 kb into the gene body. We also observed consistent nucleosome coordinates in an independent data set of H3K4me3-bearing nucleosomes[16] (Fig. 4d). Comparison of the nucleosome data (Fig. 4d) with binding patterns of RNA polymerase II[16] (Fig. 4d) around active promoters indicates that phasing of positioned nucleosomes can be explained by packing of nucleosomes against Pol II stalled at the promoter, with Pol II potentially acting as the 'barrier'. The set of inactive promoters, by contrast, exhibits neither a pronounced depletion of nucleosomes, nor a positioning and phasing signal (Fig. 4c). The transition of an inactive promoter to an active one is therefore likely to involve eviction of nucleosomes, coupled with positioning and phasing of nucleosomes neighbouring RNA Pol II (Fig. 4e). These results indicate that CpG-rich segments in mammalian promoters override intrinsic signals of high nucleosome affinity (Supplementary Fig. 8) to become active; this would be in contrast to fly and yeast, where AT-rich promoters may comprise intrinsic sequence signals that are particularly prone to nucleosome eviction[28].

To explore how regulatory factors interact with sequence signals to influence nucleosome organization outside of promoters, we focused on binding sites of the NRSF/REST repressor protein[15] and the insulator protein CTCF. NRSF and CTCF sites are flanked by arrays of positioned nucleosomes (Fig. 4f and Supplementary Fig. 9), consistent with barrier-driven packing previously reported for CTCF[29,30]. Both proteins occupy additional linker space, with NRSF taking up an extra

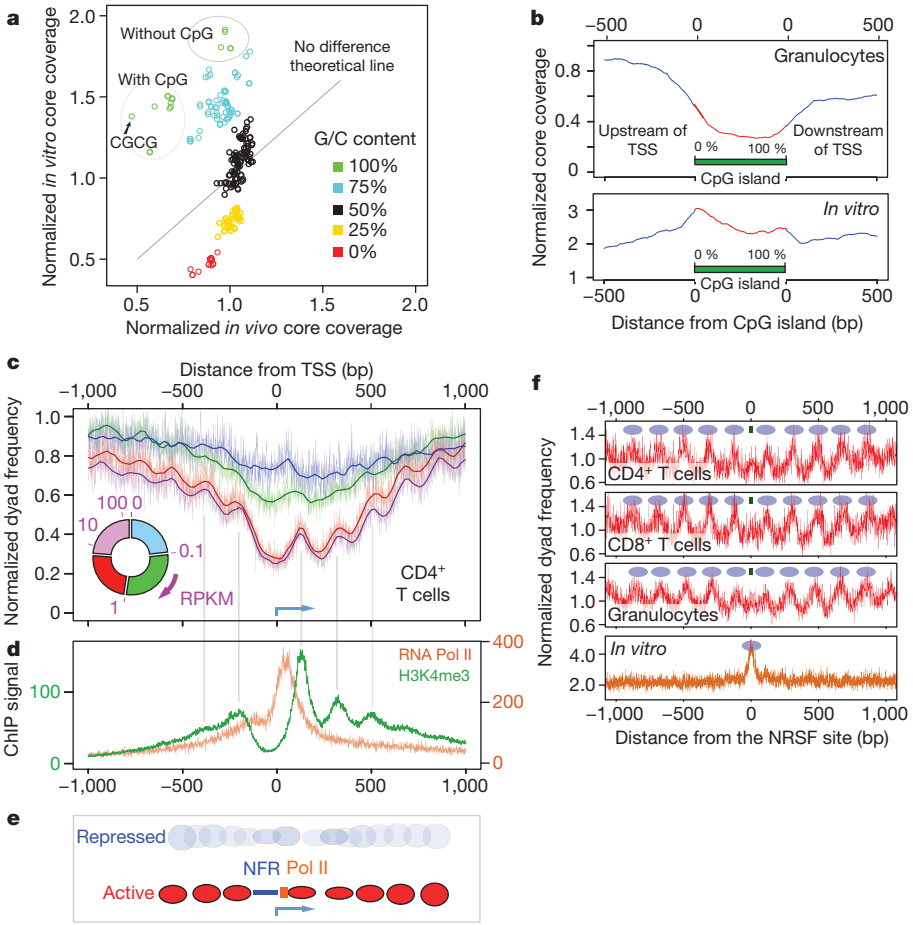


Figure 4 | Influence of gene regulatory function on nucleosome positioning. **a**, Comparison of sequence preferences of nucleosomes *in vivo* and *in vitro*. Normalized nucleosome core coverage *in vivo* (granulocytes) for a given sequence 4-mer is plotted along the *x*-axis. *In vitro* core coverage is plotted along the *y*-axis. Each data point on the plot represents one of the 256 possible 4-mers (coloured according to their G/C content). The diagonal line depicts the positions in the plot for which sequence-based preferences of nucleosomes would be the same *in vivo* and *in vitro*. **b**, Nucleosome core coverage over CpG islands *in vivo* and *in vitro*. *x*-axis represents coordinates within CpG islands (0–100%) and flanking upstream of the transcriptional start sites (TSS) (left) and downstream of the TSS (right). Normalized frequencies of nucleosome cores *in vivo* (upper plot) and *in vitro* (lower plot) are plotted along the *y*-axis. **c**, *In vivo* $CD4^+$ T-cell nucleosome organization around promoters. *x*-axis represents distance from the TSS (blue arrow). Normalized frequencies of nucleosome dyads are plotted along the *y*-axis. Nucleosome arrangements within four gene groups are shown (not expressed 0–0.1 RPKM, low expressed 0.1–1 RPKM, moderately expressed 1–8 RPKM, highly expressed > 8 RPKM). Pie chart depicts distribution of RPKM values across gene groups. **d**, RNA Pol II binding signal within highly expressed genes (orange curve) and H3K4me3-marked nucleosome dyad frequency (green curve) within highly expressed genes (>8 RPKM). Nucleosomes show consistent positions, indicated by grey lines pointing to nucleosome centres. **e**, Schematic depiction of nucleosome organization around promoters of repressed and active genes. Promoters of repressed genes do not have a well-defined nucleosome organization, whereas promoters of active genes have a nucleosome-free region (NFR, blue), RNA Pol II (orange) localized at the NFR boundary, and positioned nucleosomes (red) radiating from the NFR. Height of the ovals represents nucleosome frequency (inferred from **c**). **f**, Nucleosome distribution around the top 1,000 NRSF sites *in vivo* and *in vitro*. Distances from the NRSF binding sites are plotted along the *x*-axis. *y*-axis represents the normalized frequency of nucleosome dyads. Blue ovals depict hypothetical nucleosome positions. NRSF binding site is shown by the green rectangle.

37 bp and CTCF 74 bp. In agreement with sequence-based predictions[21], both CTCF and NRSF sites intrinsically encode high nucleosome occupancy as can be seen from the *in vitro* data (Fig. 4f and Supplementary Fig. 9), but this signal is overridden *in vivo* by occlusion of these sites from associating with nucleosomes. Additionally, phasing of nucleosomes around these regulatory sites is more compact in granulocytes compared to T cells (Supplementary Fig. 9), again exemplifying the importance of cellular parameters for placement of nucleosomes.

Our genome-wide, deep sequence data of nucleosome positions facilitated an initial characterization of the determinants of nucleosome organization in primary human cells. Spacing of nucleosomes differs between cell types and between distinct epigenetic domains in the same cell type, and is influenced by transcriptional activity. We confirm positioning preferences in regulatory elements such as promoters and chromatin regulator binding sites, but find that the majority of the human genome exhibits little if any detectable positioning. The influence of sequence on positioning of nucleosomes *in vivo* is modest but detectable. Despite DNA sequence being a potent driver of nucleosome organization at certain sites, the cellular environment often overrides sequence signals and can drive nucleosomes to occupy intrinsically unfavourable DNA elements or evict nucleosomes from intrinsically favourable sites. We find evidence for the barrier model for nucleosome organization, and that barriers can be nucleosomes (positioned by container sites), RNA polymerase II (stalled at the promoter), or sequence-specific regulatory factors. Our nucleosome maps should be useful for investigating how nucleosome organization affects gene regulation and vice versa, as well as for pinpointing the mechanisms driving regional heterogeneity of nucleosome spacing.

METHODS SUMMARY

Neutrophil granulocytes, $CD4^+$ and $CD8^+$ T cells were isolated from donor blood using Histopaque density gradients and Ig-coupled beads against blood cell surface makers (pan T and $CD4^+$ microbeads, Miltenyi Biotec). Nucleosome cores

were prepared as described previously[7]; cells were snap-frozen and crushed to release chromatin, followed by micrococcal nuclease treatment. *In vitro* nucleosomes were prepared by combining human genomic DNA with recombinantly-derived histone octamers at an average ratio of 1 octamer per 850 bp. Unbound DNA was then digested using micrococcal nuclease. After digestion, reactions were stopped with EDTA, samples were treated with proteinase K, and nucleosome-bound DNA was extracted with phenol-chloroform and precipitated with ethanol (Supplementary methods). Purified DNA was size-selected (120–180 bp) on agarose to obtain mononucleosome cores, followed by sequencing library construction. RNA was isolated by homogenizing purified cells in TRIzol, poly-A RNA was purified using a Qiagen Oligotex kit and RNA-seq libraries were constructed using a SOLiD Whole Transcriptome Analysis kit. All sequence data was obtained using the SOLiD 35 bp protocol and aligned using the SOLiD pipeline against the human hg18 reference genome. Downstream analyses were all conducted using custom scripts (Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 August 2010; accepted 18 March 2011.
Published online 22 May; corrected 23 June 2011 (see full-text HTML version for details).

1. Mellor, J. The dynamics of chromatin remodeling at promoters. *Mol. Cell* **19,** 147–157 (2005).
2. Radman-Livaja, M. & Rando, O. J. Nucleosome positioning: how is it established, and why does it matter? *Dev. Biol.* **339,** 258–266 (2010).
3. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458,** 362–366 (2009).
4. Berstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O. & Schreiber, S. L. Global nucleosome occupancy in yeast. *Genome Biol.* **5,** R62 (2004).
5. Yuan, G.-C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309,** 626–630 (2005).
6. Johnson, S. M., Tan, F. J., McCullough, H. L., Riordan, D. P. & Fire, A. Z. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin. *Genome Res.* **16,** 1505–1516 (2006).
7. Valouev, A. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18,** 1051–1063 (2008).
8. Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132,** 887–898 (2008).
9. Trifonov, E. N. & Sussman, J. L. The pitch of chromatin DNA is reflected in it its nucleotide sequence. *Proc. Natl Acad. Sci. USA* **77,** 3816–3820 (1980).
10. Kornberg, R. D. Structure of chromatin. *Ann. Rev. Biochem.* **46,** 931–954 (1977).
11. Widom, J. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc. Natl Acad. Sci. USA* **89,** 1095–1099 (1992).
12. Schlegel, R. A., Haye, K. R., Litwack, A. H. & Phelps, B. M. Nucleosome repeat lengths in the definitive erythroid series of the adult chicken. *Biochim. Biophys. Acta* **606,** 316–330 (1980).
13. Fan, Y. *et al.* Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **29,** 1199–1212 (2005).
14. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5,** 621–628 (2008).
15. Valouev, A. *et al.* Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nature Methods* **5,** 829–834 (2008).
16. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
17. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genet.* **40,** 897–903 (2008).
18. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191,** 659–675 (1986).
19. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442,** 772–778 (2006).
20. Hughes, A. & Rando, O. J. Chromatin 'programming' by sequence - is there more to the nucleosome code than %GC? *J. Biol.* **8,** 96 (2009).
21. Tillo, D. *et al.* High nucleosome occupancy is encoded at human regulatory sequences. *PLoS ONE* **5,** e9129 (2010).
22. Mavrich, T. N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18,** 1073–1083 (2008).
23. Mavrich, T. N. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature* **453,** 358–362 (2008).
24. Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genet.* **39,** 1235–1244 (2007).
25. Gu, S. G. & Fire, A. Partitioning the *C. elegans* genome by nucleosome modification, occupancy, and positioning. *Chromosoma* **119,** 73–87 (2010).
26. Sasaki, S. *et al.* Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. *Science* **323,** 401–404 (2009).
27. Zhang, Y. *et al.* Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. *Nature Struct. Mol. Biol.* **16,** 847–852 (2009).
28. Field, Y. *et al.* Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nature Genet.* **41,** 438–445 (2009).
29. Chuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19,** 24–32 (2009).
30. Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4,** e1000138 (2008).
31. Albert, I. *et al.* Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446,** 572–576 (2007).
32. Wellinger, R. E. & Thoma, F. Nucleosome structure and positioning modulate nucleotide excision repair in the non-transcribed strand of an active gene. *EMBO J.* **16,** 5046–5056 (1997).
33. Sha, K. *et al.* Distributed probing of chromatin structure *in vivo* reveals pervasive chromatin accessibility for expressed and non-expressed genes during tissue differentiation in *C. elegans*. *BMC Genomics* **11,** 465 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Stanford Genetics/Pathology Sequencing Initiative. We thank G. Narlikar for help with *in vitro* experiments, Life Technologies, especially J. Briggs, for help with generating sequencing data, P. Lacroute for help with sequence alignment, S. Galli for valuable discussions, L. Gracey for critical reading of the manuscript, and members of the Sidow and Fire labs for valuable feedback and discussions. Work in the Fire lab was partially supported by NIGMS (R01GM37706). A.V. was partially supported by an ENCODE subcontract to A.S. (NHGRI U01HG004695). S.M.J. was partially supported by the Stanford Genome Training program (NHGRI T32HG00044).

Author Contributions A.V., S.M.J., A.S. and A.Z.F. designed the experiments. S.M.J., A.V., C.L.S. and S.D.B. performed the experiments. A.V. designed and carried out analyses with input from A.S., A.Z.F. and S.M.J.; A.V., A.S. and A.Z.F. wrote the manuscript.

Author Information All sequence data were submitted to Sequence Read Archive (accession number GSE25133). Sites containing strongly positioned *in vitro* nucleosomes are available as a supplementary data file. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.S. (arend@stanford.edu) or A.Z.F. (afire@stanford.edu).

METHODS

Cell purification. Blood samples were obtained from the Stanford Blood Center. Samples were screened for any medical history of malignancy or signs of infectious disease, and tested for serologic evidence of viral infections to ensure that samples came from healthy donors. The Stanford Blood Center procedures used for the cells in this study are the same as those used for transfusion of patients and are routinely inspected by the FDA, the American Association of Blood Banks, and the College of American Pathologists. The blood for the experiments was processed immediately upon donation to avoid any change in quality as a result of sample storage.

Buffy coat (36 ml) from a blood donor was diluted in PBS to a total volume of 200 ml. The cells were layered on a Histopaque gradient with densities 1.119 and 1.077 g ml^{-1} according to manufacturer's instructions (Sigma HISTOPAQUE-1119 and 1077) and separated by centrifugation to yield granulocytes and mononuclear fractions. T cells were isolated from mononuclear cells using a Pan T isolation Kit (Miltenyi Biotec), followed by separation into CD4$^+$ and CD8$^+$ fractions using CD4$^+$ microbeads (Miltenyi Biotec).

Isolation of mononucleosome core DNA fragments from human cells. To isolate mononucleosome core DNA from human cells, neutrophil granulocytes, CD4$^+$ lymphocytes and CD8$^+$ lymphocytes were flash-frozen in liquid nitrogen in 0.34 M sucrose Buffer A and ground, digested on different days, and isolated as described in ref 7. By carrying out an MNase digestion in a short time frame (12 min at 16 °C) following grinding of the samples, we minimize the potential for nucleosome mobility. To maximize uniformity of representation, we use an extraction protocol after MNase digestion that does not rely on solubility of the individual core particles; this resulted in recovery of the bulk of input DNA as a mono-nucleosome band (Supplementary Fig. 10), limiting the degree to which the protocol might select for specific (for example, accessible) chromosomal regions. The mean nucleosome core length obtained for analysis (153 nucleotides) indicates an average overhang of 3 nucleotides on each side of individual cores (147 bp + 2 × 3 bp = 153 bp). Subsequent analyses assign nucleosome positions accounting for this mean overhang and making use of the ability to define location based on interpolation between values calculated from plus-oriented and minus-oriented reads (see below).

Preparation of *in vitro* nucleosomes. Naked genomic DNA isolated from neutrophil granulocytes from our *in vivo* studies was sheared by sonication using a Covaris sonicator and separated on a 1% UltraPure Agarose (Invitrogen) gel run at 100 V for 1 h. A smear of fragments with lengths from 850–2,000 bp (the bulk of the sheared DNA) was isolated and extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen). DNA fragment lengths several-fold larger than nucleosome cores were chosen for this analysis to minimize any end-effects that could have contributed an end-based signal at shorter fragment sizes. Lack of end-preference in the reconstitutions was then confirmed under the conditions of these assays using a series of defined restriction fragments as templates for assembly (S.M.J. and A.F., results not shown).

The ends of the sheared DNA fragments were repaired as described below and then were assembled with recombinant *Xenopus* histones into nucleosomes as described previously[34] at a 1.1:1 molar ratio of DNA to histone octamer such that on average one nucleosome would occupy 850 bp of DNA. Specifically, 4.9 μg of DNA and 0.80 μg of octamer were reconstituted in a total volume of 200 μl.

The ref. 34 conditions (in which DNA was not limiting) were used for our analysis in order to focus specifically on primary sequence effects on nucleosome position. We note that two recent studies in yeast use somewhat different conditions, with a higher ratio of nucleosomes to DNA[3,27]. Assays at high nucleosome:DNA ratio provide a composite readout reflecting both (1) primary preferences of nucleosomes (caused by sequence signals within the nucleosome-bound DNA) and (2) secondary effects due to steric hindrance as a result of dense packing of nucleosomes. Although such data are certainly valuable in modelling chromosome dynamics, the goals of our study (definition of individual sequence elements that can initiate positioning) were best served with the lower nucleosome:DNA assay conditions[34].

Isolation of *in vitro* nucleosome core DNA fragments. *In vitro* nucleosome core DNAs were isolated by diluting 70 μl of the reconstituted *in vitro* nucleosome into a total volume of 200 μl containing 5 mM $MgCl_2$, 5 mM $CaCl_2$, 70 mM KCl and 10 mM Hepes at pH 7.9 (final concentrations) and digesting with 20 units of micrococcal nuclease (Roche) resuspended at 1 U μl^{-1} for 15 min at room temperature. The digestion was stopped by adding an equal volume of 3% SDS, 100 mM EDTA and 50 mM Tris. Octamer proteins were removed by treating with one-tenth volume proteinase K (20 mg ml^{-1} in TE at pH 7.4) for 30 min at 50 °C followed by phenol/chloroform and chloroform extractions and ethanol precipitation. This procedure was repeated twice to process the entire *in vitro* sample, and then *in vitro* DNA cores were isolated on a 2% UltraPure Agarose (Invitrogen) gel run at 100 V for 1 h followed by DNA extraction from the gel using a QIAquick Gel Extraction Kit (Qiagen) following the standard protocol with the exception of allowing the isolated gel sample to incubate in Buffer QG at room temperature until dissolved.

Genomic MNase digest control library preparation. For control libraries, genomic DNA (20 μg) from human neutrophil granulocytes in 0.34 M sucrose Buffer A with 1× BSA (New England Biolabs) and 1 mM $CaCl_2$ was digested with 200 units of micrococcal nuclease (Roche) (0.4 U μl^{-1} final concentration) in a total volume of 500 μl for 10 min at 23 °C. The digestion was stopped by addition of 10 μl 0.5 M EDTA, followed by ethanol precipitation. The digested DNA was run on a 2.5% agarose gel and the smear of DNA fragments from 135–225 bp was excised from the gel and purified using a QIAquick Gel Extraction Kit (Qiagen) as noted above.

End repair, linker ligation and library amplification. The ends of isolated mononucleosome core DNAs (granulocytes, CD4$^+$ lymphocytes and CD8$^+$ lymphocytes), *in vitro* core DNAs and genomic control DNAs were processed by treating 0.3–0.5 μg of the DNA samples with T4 polynucleotide kinase (New England Biolabs) at 37 °C for 2.5 h followed by ethanol precipitation and subsequent treatment with T4 DNA polymerase (New England Biolabs) in the presence of dNTPs for 15 min at 12 °C. After purification using either a QIAquick Gel Extraction Kit as described above or a QIAquick PCR Purification Kit (Qiagen), linkering of previously annealed duplexes AF-SJ-47 (5′-OH-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3′)/AF-SJ-48 (5′-P-ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT-3′) and AF-SJ-49 (5′-OH-CTGCCCCGGGTTCCTCATTCTCT-3′)/AF-SJ-50 (5′-P-AGAGAATGAGGAACCCGGGGCAGTT-3′) to the samples was accomplished with T4 DNA ligase during a 6.5-h room-temperature incubation. The ligation reactions were separated on a 2% agarose gel, and the relevant band isolated as described above. Amplification of the linkered libraries was accomplished with 8 (granulocyte mononucleosome library), 10 (CD4$^+$ lymphocytes, CD8$^+$ lymphocytes and genomic control libraries) or 12 (*in vitro* library) cycles of polymerase chair reaction (PCR) using primers AF-SJ-47 (SOLiD P1 primer) and AF-SJ-49 (SOLiD P2 primer) with subsequent separation and purification using a 2% agarose gel and the QIAquick Gel Extraction Kit as described above. The number of cycles used in the PCR amplification were monitored and selected as described in ref. 25.

RNA-seq library preparation. Cells were homogenized in TRIzol using an 18G needle, followed by total RNA extraction using phenol-chloroform-isoamyl alcohol. Poly-A RNA was isolated from total RNA using a Qiagen Oligotex kit according to the manufacturer's instructions. The RNA-seq SOLiD sequencing library was built from 100 ng of poly-A RNA according to the manufacturer's instructions (SOLiD whole transcriptome analysis kit).

DNA sequencing and mapping. Both nucleosome fragment and RNA-seq libraries were sequenced using the SOLiD DNA sequencing platform to produce 35 bp reads. All sequence data was mapped using SOLiD software pipeline against the human hg18 assembly using the first 25 bp from each read. This was done to maximize the number of the reference-mapped reads, as the higher error rate in read positions 26–35 of that version of the SOLiD chemistry prevented a substantial fraction of reads from mapping to the genome. For the genome-wide analysis we retained only unambiguously mapped reads.

Genome coverage by nucleosome cores was calculated as: core coverage = (number of mapped reads) × (147)/(genome size)

mRNA sequencing and data analysis. RNA-seq libraries were sequenced on the SOLiD platform to produce 35 bp reads and then the first 25 bp of each read were mapped to hg18 using the SOLiD mapping pipeline which resulted between 77 and 99 million mapped reads for each cell type. RPKM values were calculated as in ref. 14, with a modification that adjusted for transcript length, which was calculated according to the formula $L' = L - 50 \times (E - 1)$, where L is the actual transcript length, and E is the number of exons in the gene. This modification is needed because of the lack of mappings across splice junctions.

Mathematical notations. Start counts: $S_{+/-}(j)$ represent counts of 5′ coordinates of reads that map in + or – orientation at the j-th position of the reference strands. For example, if read maps to the interval $[x,y)$ on the + strand, then its 5′ coordinate is x, if it maps to – strand, then it's $y - 1$.

Indicator functions: I(condition) = 1 if condition is satisfied, 0 otherwise.

Nucleosome positioning stringency metric: nucleosome positioning stringency metric quantifies the fraction of nucleosomes covering a given position that are 'well positioned'. The stringency at position i of the genome is calculated according to the formula:

$$S(i, w=30) = \frac{D(i, w=30)}{\sum_{i-150}^{i+150} \frac{1.09}{w} D(j, w=30)},$$

where $D(i,w)$ is a kernel-smoothed dyad count calculated according to the formula:

$$D(i,w) = \sum_{j=0}^{L} K(i-j,w) \cdot d(j),$$

where L is the size of a given chromosome, and $K(u,w)$ is a smoothing kernel function of the form:

$$K(u,w) = (1-(u/w)^2)^3 I\{|u| < w\},$$

and

$$\int_{-1}^{1} (1-u^2)^3 du = 1/1.09,$$

and $d(j)$ represents the number of dyads that occurs at the position j:

$$d(j) = s_+(j-l/2) + s_-(j+l/2).$$

Here l is the average library size ($l = 153$ for *in vivo* data sets, 147 for *in vitro* data set). The core size is inferred from the 3-pile distogram peak position in the range of 100–200 bp.

The numerator of the stringency formula represents a kernel-smoothed count of nucleosome centres (dyads) at position i in the genome, whereas the denominator represents the count of nucleosome centres that infringe on the nucleosome centred at that position, which is inferred by integration of the dyad density estimate over an area of nucleosome infringement. The stringency is constructed in such a way that it would achieve a maximum of 1 if all nucleosomes were perfectly centred at that position (Supplementary Fig. 4). If two alternative, mutually exclusive, equally frequent nucleosome positions are observed in the data, then the stringency would be 0.5 or 50% for each alternative site (illustrated in Supplementary Fig. 4).

Application of the Kernel Density Estimation allowed obtaining smooth estimates of the stringency, which was useful for detection of nucleosome centres and robustly estimating the degree of positioning. We experimented with other smooth kernels and obtained highly consistent results. In principle, the kernel choice should not affect the results substantially as long as there is sufficient nucleosome core coverage (which follows from the convergence property of Kernel Density Estimation).

The kernel bandwidth w is an important parameter of the stringency formula and provides a means to control the smoothness of the stringency profile. Larger values of w provide higher smoothing but result in less accurate estimates of positioning centres, which is acceptable in cases of low core coverage. On the other hand, lower values of w result in less smoothing but more accurate estimation of the positioning centres, which is desirable in cases when nucleosome core coverage is high. We decided to use $w = 30$ in our calculation as it provided a sufficient amount smoothing across all of our data sets without sacrificing the sharpness of the positioning estimate.

Nucleosome positioning stringency was used for calculation of the fraction of the genome containing preferentially positioned nucleosomes (Supplementary Fig. 5). Positioned nucleosomes used in the container site analysis (Fig. 3a–c) were identified with the positioning stringency metric (as shown) and additional filters on nucleosome occupancy (*in vitro* occupancy > 30) to improve the statistical confidence of the positioning estimates.

Nucleosome dyad coordinates. Nucleosome dyads were inferred from 5′ coordinates of reads by shifting them by half the average nucleosome core size towards the 3′ end. The average nucleosome cores size was estimated by a maximum value of the 3-pile distogram in a size range of 100–200 bp.

Rotational positioning analysis. We examined oligonucleotide preferences of rotational positioning of nucleosomes, which is associated with 10-bp patterning of short k-mers within nucleosome cores[18,31]. Plotting the frequencies of dyads around specific oligomers within the genome showed that the strongest patterning was exhibited by C-polymers (CC,CCC) with an exact helical period of 10.15 bp (Supplementary Fig. 7a, P-value $< 2 \times 10^{-16}$), indicating that they are important for rotational positioning. *In vivo*, such rotational preferences are much less pronounced (Supplementary Fig. 7b), indicating that cellular factors or conditions often override the sequence-encoded rotational settings.

Characterization MNase cleavage patterns. MNase is known to have sequence preferences that can affect both individual and bulk analyses of chromatin structure. Previous studies comparing MNase with alternative probes in model systems, both at specific loci (for example, ref. 32) and genome wide (for example, ref. 33), support the correspondence between the patterns of nucleosomes inferred from MNase digestion of chromatin and the *in vivo* chromatin landscapes. Nonetheless, it remained important to characterize the patterns of MNase activity in our data.

We investigated the extent of cleavage bias by MNase by examining sequence preferences within the cleavage sites, which correspond to 5′ end read positions in our data (Supplementary Figure 7a–e). Consistent with previous observations, MNase exhibits a pronounced but imperfect tendency to cleave at A or T nucleotides in naked DNA (Supplementary Fig. 11a). This same bias is detectable but, importantly, weaker when nucleosomes occupy the DNA, both *in vivo* and *in vitro* (1-pile subsets, top row b–e). Sites of more frequent cleavage (3-pile and 5-pile subsets, middle and bottom rows) revealed preferences that were virtually indistinguishable from the single-site preference.

The fact that the cleavage bias does not extend beyond 1–2 base pairs suggests that our analyses of nucleosome positioning preferences, which have substantially less than single-base resolution, should be robust to biases introduced by the MNase digestion. A case in point is the above-discussed rotational positioning analysis, whose resolution is on the order of 10 bp and which involves oligonucleotides that do not resemble the MNase cleavage site (Supplementary Fig. 7a).

To investigate whether the sequence-driven nucleosome positioning element identified by the *in vitro* reconstitution experiment (Fig. 3) was a result of particularly pronounced MNase digestion bias within specific sites, we examined nucleotide preferences of nucleosome fragments overlapping sites of medium (>0.5) and high (>0.7) positioning stringency (Supplementary Fig. 11f, g). Preferences within these sites are identical to genome-wide preferences, ruling out the possibility that their positioning is an artefact of MNase digestion. In addition, we observe wave-like patterns *in vivo* around these sites (Fig. 3e) consistent with existence of a chromatin barrier in the form of a well-positioned nucleosome.

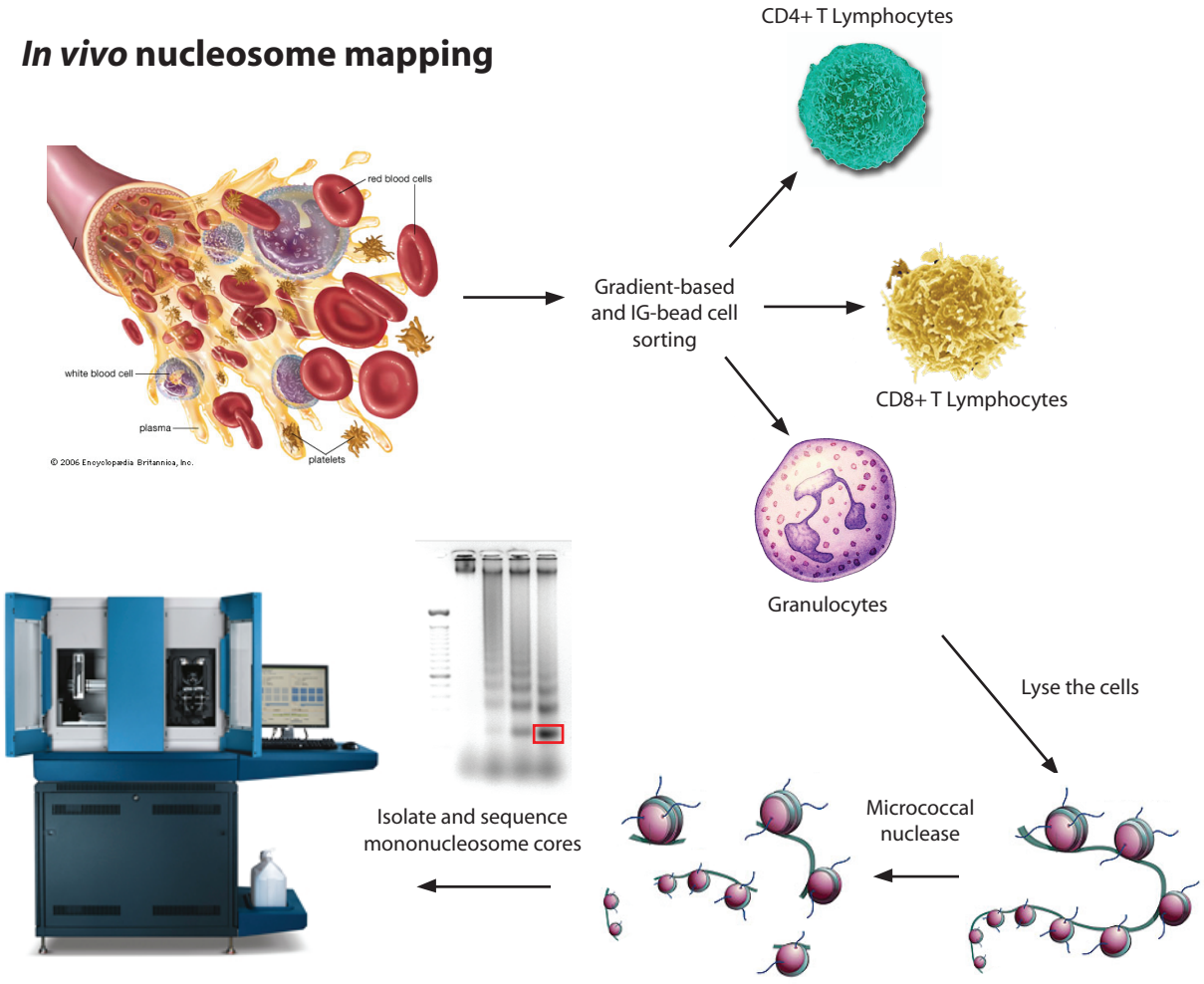
The lack of systematic differences in cleavage bias in our experimental data sets, in conjunction with the fact that naked DNA is affected most by the cleavage bias, suggests that our conclusions are robust to the use of MNase.

Analyses of independent data sets. We conducted additional analyses on independent data not generated by us to address any lingering concerns about biases or reproducibility. First, we sought to confirm independently that MNase cuts the linker DNA separating nucleosomes. In our data, CTCF sites (Supplementary Fig. 9) are surrounded by arrays of highly positioned and phased nucleosomes extending at least 1 kb in each direction. We investigated the frequency of cleavage by DNase I, a nuclease with preferences different from those of MNase, around CTCF sites within lymphoblastoid cell lines, using publicly available data from the ENCODE project. In agreement with our MNase results, we observed strongly phased peaks in the DNase I ENCODE data that align with linker DNA sites in our nucleosome data (Supplementary Fig. 12).

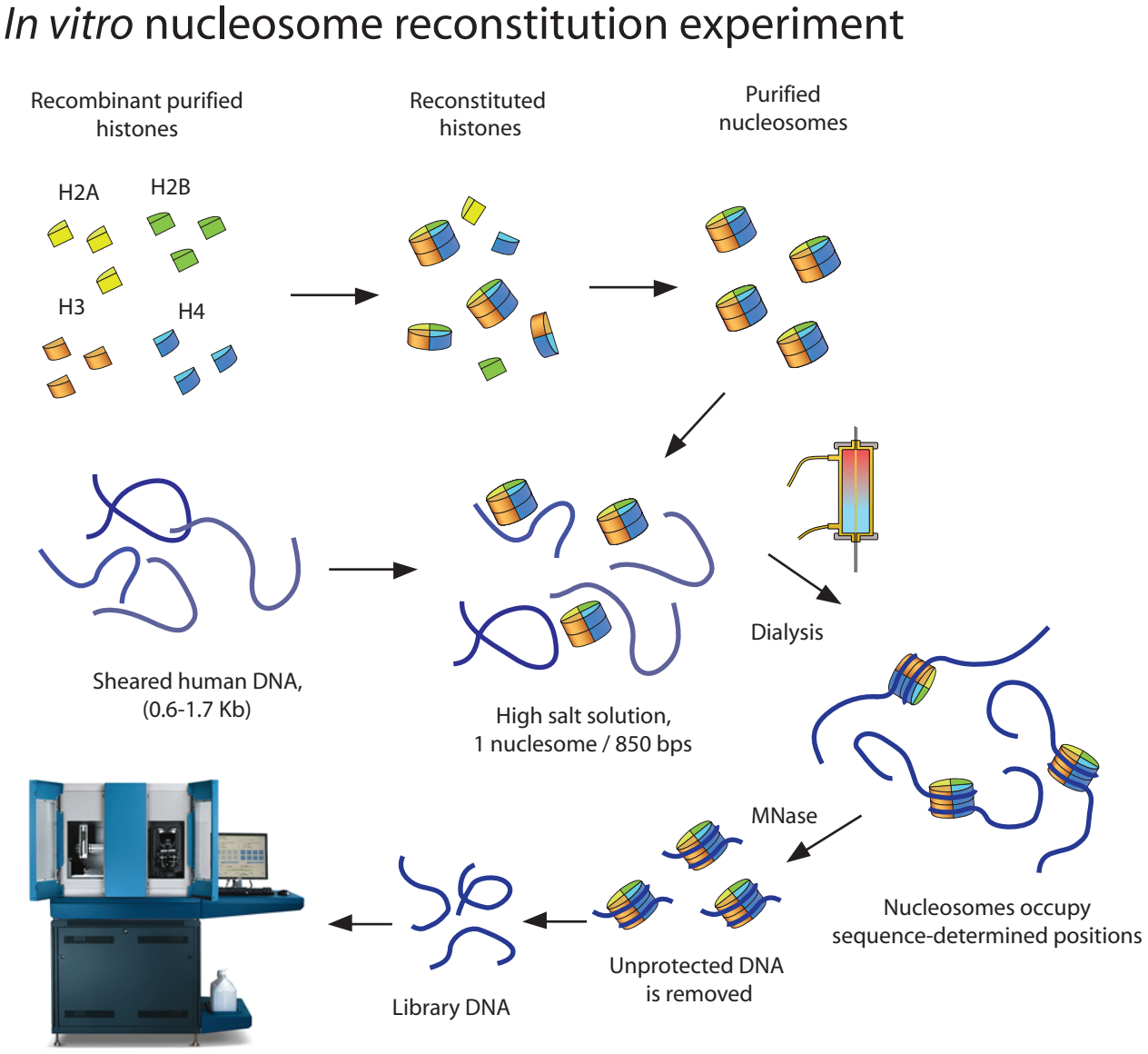
The estimates of spacing between nucleosomes as depicted in Fig. 1d are consistent between the two types of T cells we analysed. To ask whether these estimates were also reproducible by a different approach, we turned to a published data set that was generated for a different purpose, and by different means. Ref. 8 compared nucleosome distribution between resting and activated $CD4^+$ T cells using MNase treatment of the cellular chromatin. We analysed spacing of nucleosomes in their data and obtained a highly concordant estimate of 202 and 203 bp (Supplementary Fig. 13) which is in agreement with the 203 bp spacing we see in our data (Fig. 1d).

34. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304,** 3–19 (1999).

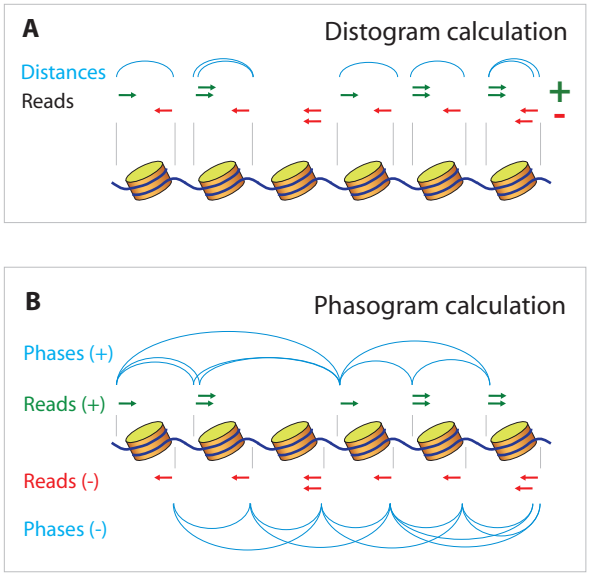
doi:10.1038/nature10002



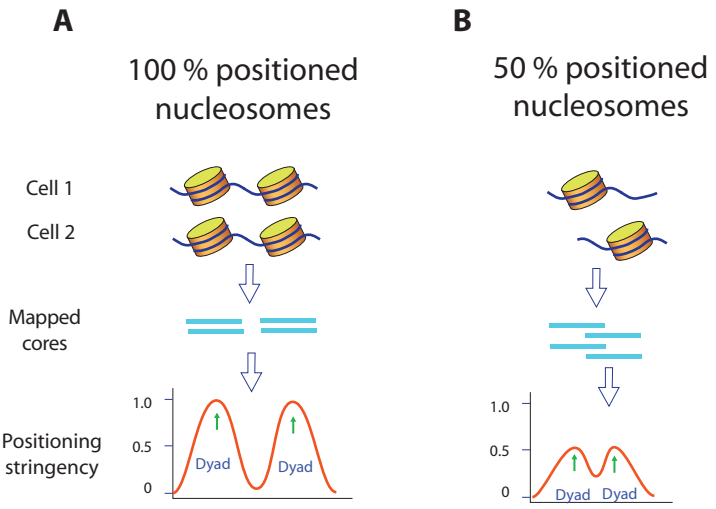
Supplementary Figure 1. Schematic depiction of *in vivo* nucleosome mapping experiment. Blood cells were isolated from a human donor blood and sorted into populations representing CD4+ T-cells, CD8+ T-cells and granulocytes. Nuclear chromatin was released by crushing the cells, followed by Micrococcal nuclease treatment. Mononucleosome fraction was isolated by gel electrophoresis and sequenced to high depth using SOLiD platform.



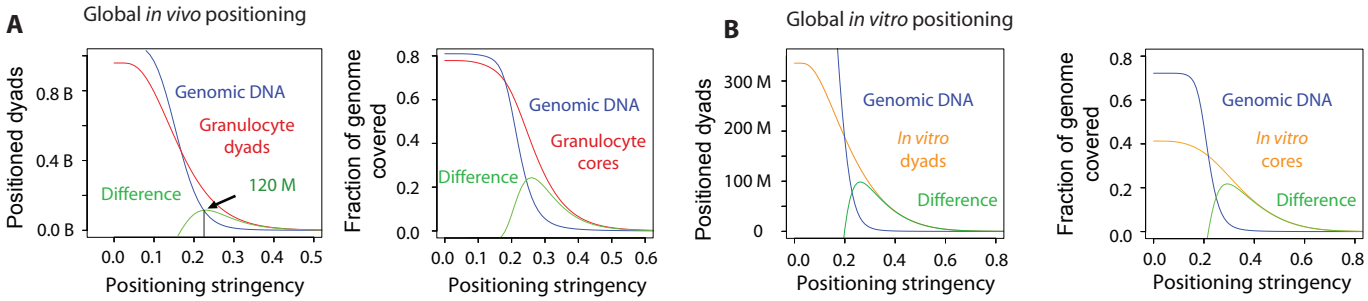
Supplementary Figure 2. Schematic representation of *in vitro* reconstitution experiment. Recombinant histones were assembled to produce the histone octamer particles. Human genomic DNA was sheared to a range of 0.6-1.7 Kb and combined with octamers at a ratio of one octamer per 850 bps of DNA. The salt was gradually dialyzed away and unbound DNA was removed by Micrococcal nuclease treatment. Nucleosome-bound DNA was purified and sequenced on the SOLiD platform.



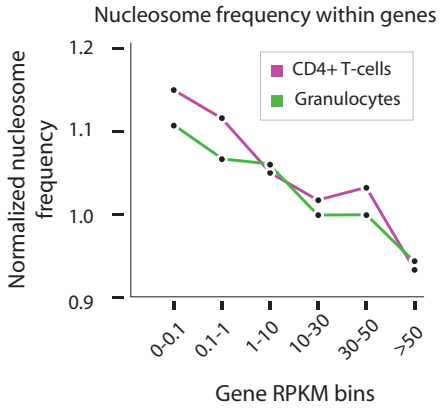
Supplementary Figure 3. Distograms and phasograms. (A) Schematic depiction of the distogram calculation. Blue arcs represent recorded distances between nucleosome reads that map on opposite strands. Distance frequencies are represented as a histogram (distogram, see Fig. 1A-B of the main text). Distograms are used to reveal the existence of consistently positioned nucleosomes in the main data. **(B)** Schematic depiction of the phasogram calculation. Blue arcs represent recorded phases between the nucleosome reads mapping on the same strand of the reference genome. Phase frequencies are represented as a histogram (phasogram, see Fig 1C-D). Phasograms are used to reveal the existance of consistently spaced nucleosomes forming regular arrays.



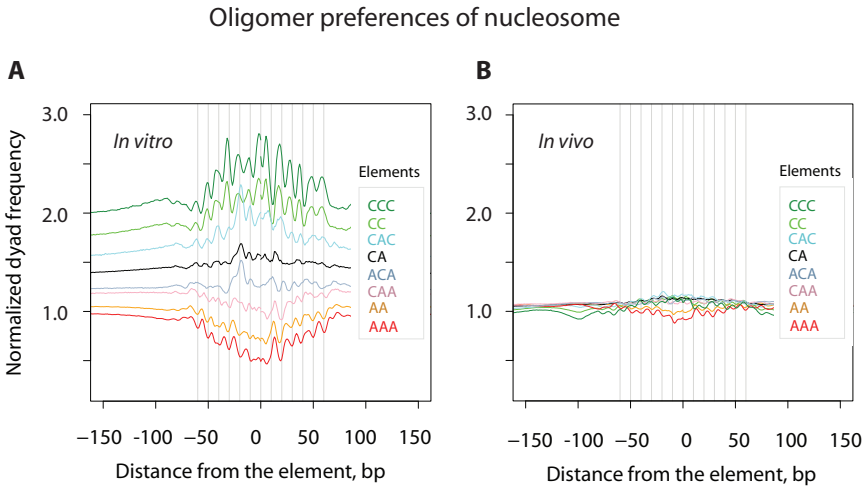
Supplementary Figure 4. Schematic depiction of the nucleosome positioning stringency metric. At the sites containing perfrectly positioning nucleosomes (panel **A**) the stringency values are 1.0 (100% positioning), and at the sites containing two mutually exclusive nucleosome positions which are utilized with 50% frequency across cells (panel **B**), the stringency values are 0.5 (50% positioning frequency at each of the two sites). Nucleosome dyad positions are identified as the local maxima of the stringency profile (green arrows).



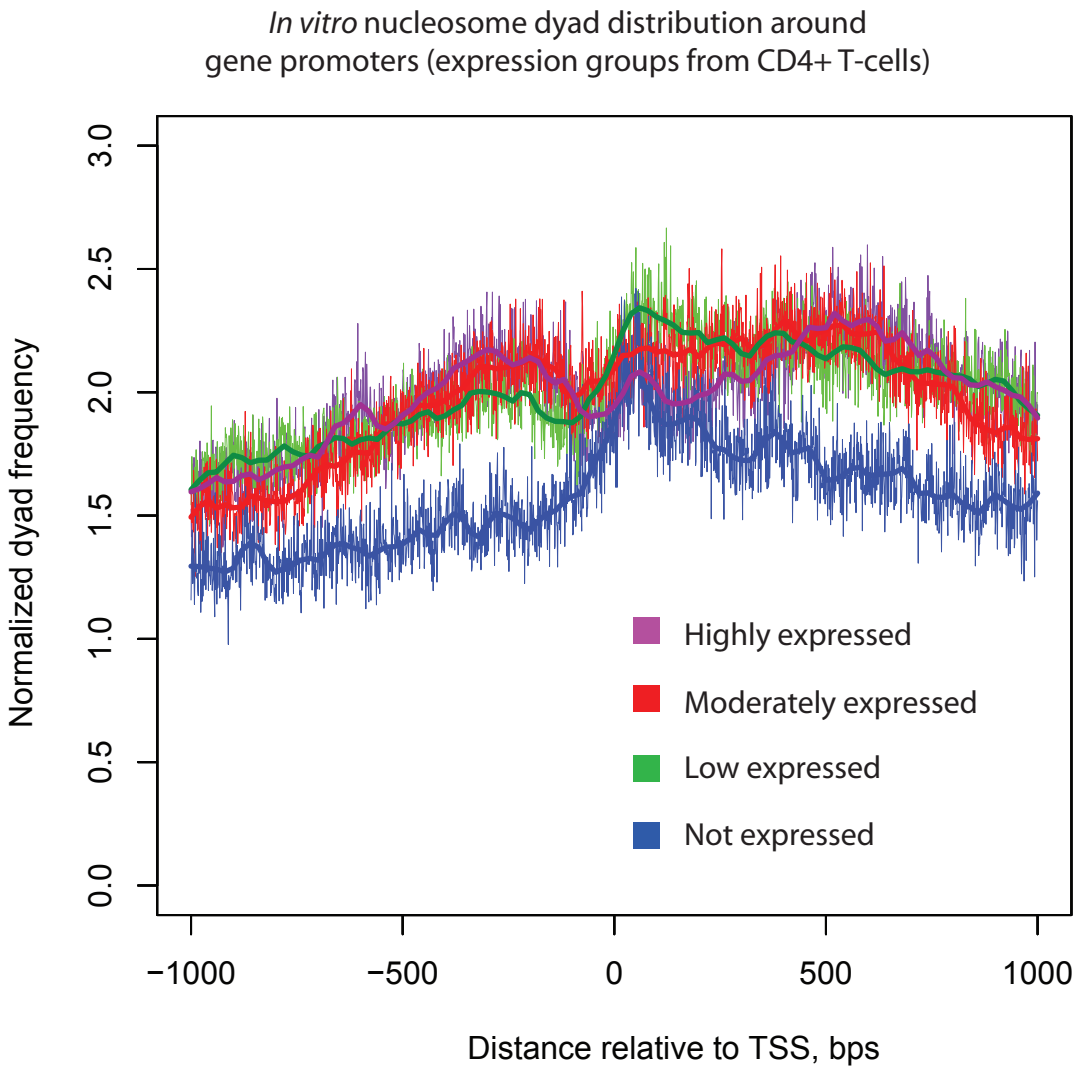
Supplementary Figure 5. Genome-wide positioning of nucleosomes. (A) Global *in vivo* nucleosome positioning of granulocytes. In both panels, X axis represents a range of positioning stringency cutoffs. In the left panel, Y axis represents the number of positioned dyads at a given positioning stringency cutoff. The red curve represents granulocyte data, the blue curve represents genomic DNA control matched to the number of granulocyte reads, the green curve represents the difference curve that provides the number of statistically positioned dyads at a given stringency cutoff. In the right panel, Y axis represents the fraction of the genome covered by 147 bp nucleosome cores centered at the dyad positions exceeding a given stringency. The red curve represents granulocyte nucleosome data, the blue curve represents genomic control matching the granulocyte data read number, and the green curve represents the difference between granulocytes and control curves and gives the fraction of the genome covered by statistically positioned nucleosomes. **(B)** Global *in vitro* nucleosome positioning. The data are plotted as in **(A)** using *in vitro* data and control matching the read number of the *in vitro* data set.



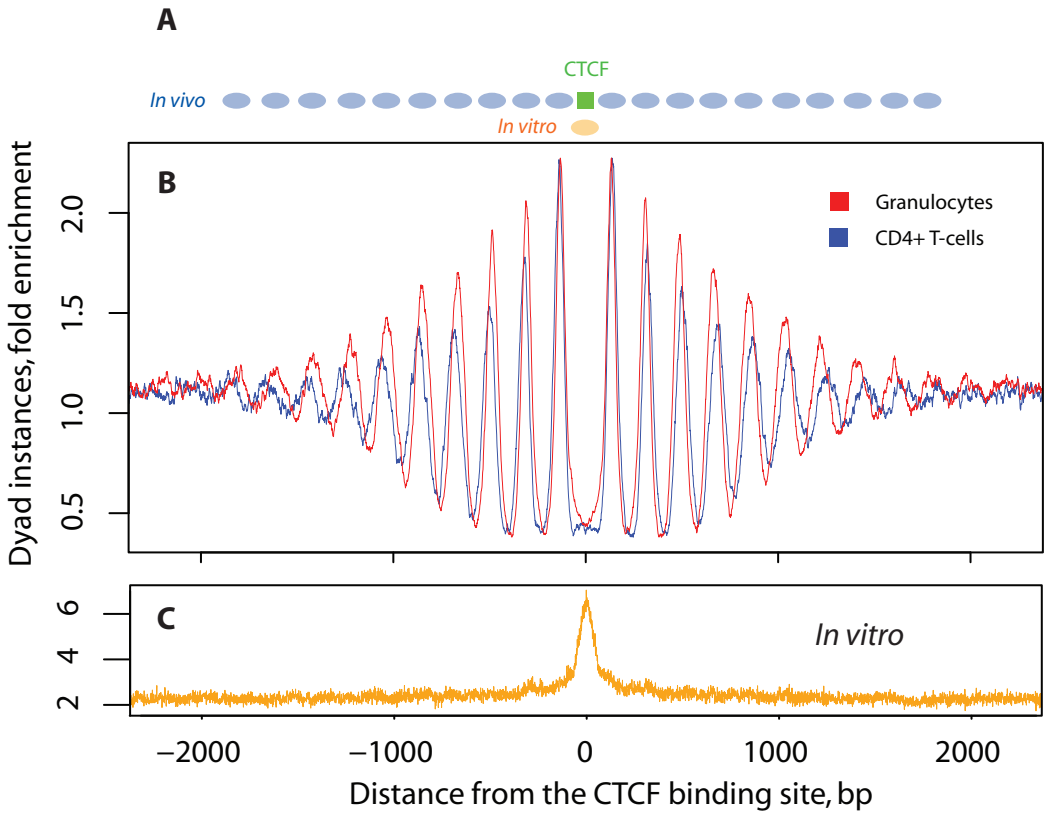
Supplementary Figure 6. Association between transcriptional levels and measured nucleosome occupancy. X axis represents gene expression values binned according to their RPKM values. Y axis represents normalized frequencies of observed nucleosome coverage within the regions occupied by genes in each bin.



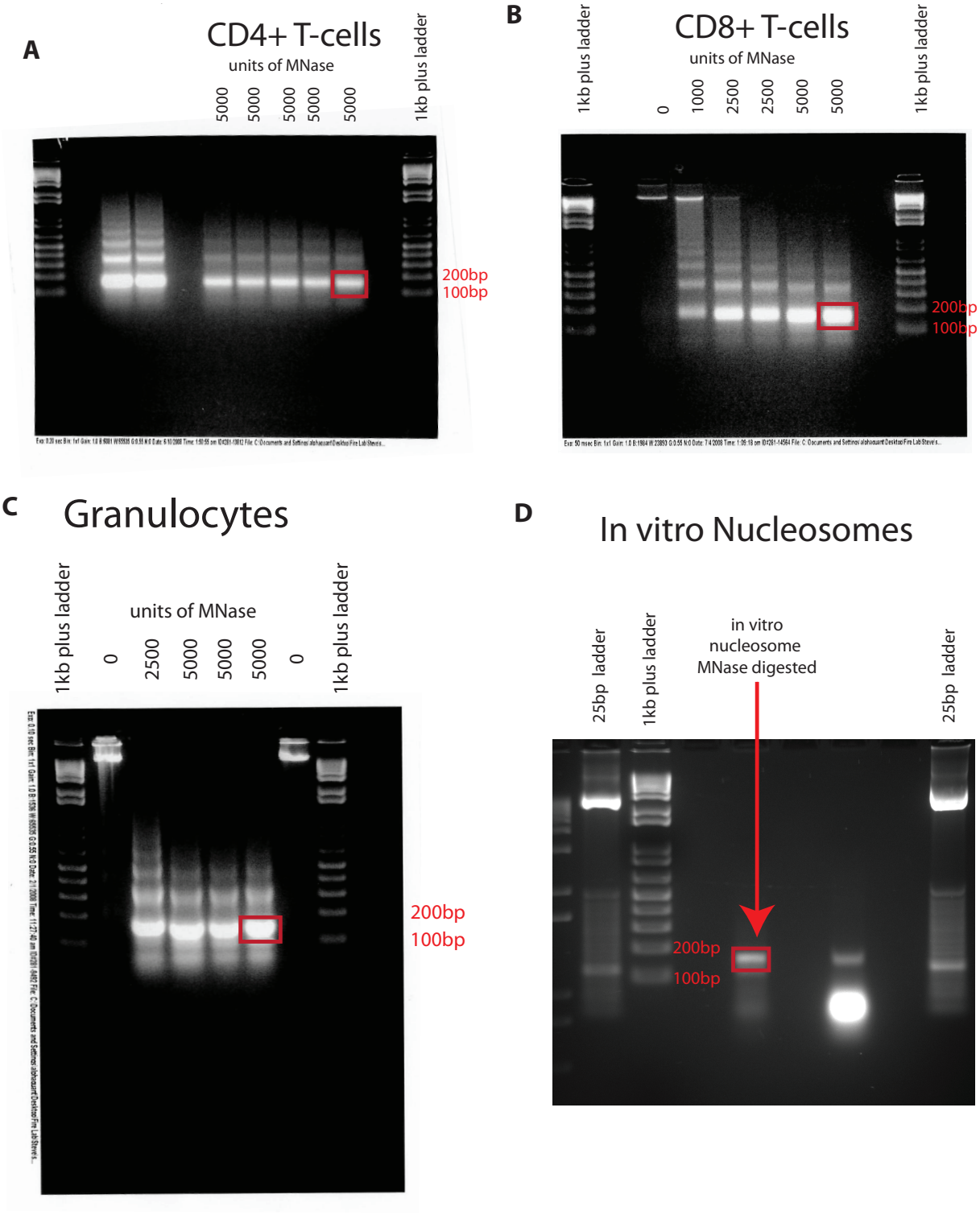
Supplementary Figure 7. **(A)** Signatures of rotational positioning of *in vitro* nucleosomes. Shown are preferences relative to most dimers and trimers composed of Cs and As. X axis represents a distance from a given oligomer to a dyad inferred from mapped sequence reads. Y axis represents the frequency of dyads at a given distance normalized to the expected frequency. 10bp-spaced peaks represent helical rotational preferences of oligomers relative to nucleosome surface. **(B)** Signatures of rotational positioning of *in vivo* granulocyte nucleosomes against the same panel of oligomers.



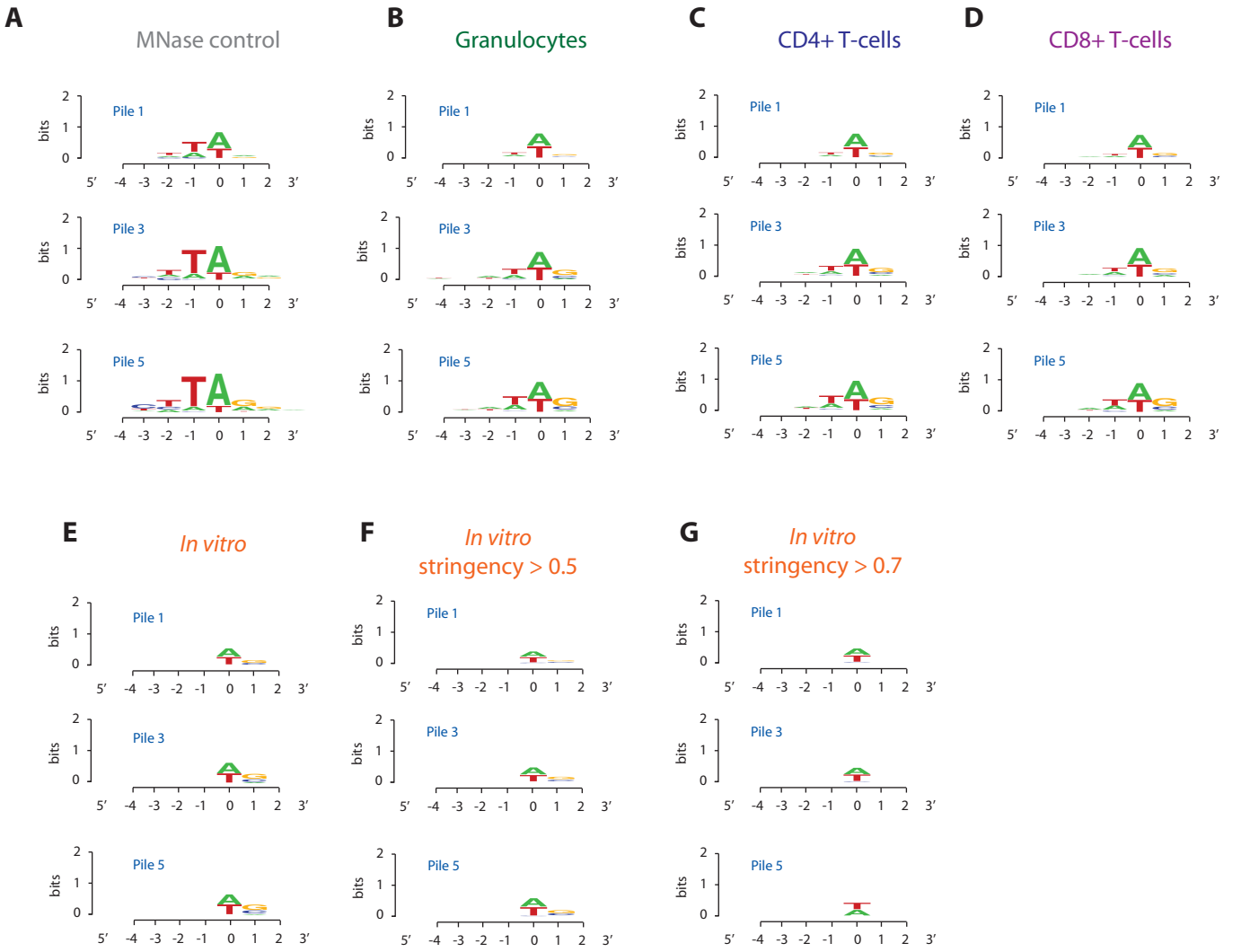
Supplementary Figure 8. Sequence-encoded nucleosome organization around TSS. Plotted are frequencies of *in vitro* nucleosome dyads around promoters of genes binned according to their expression levels in CD4+ T-cells. X axis represents the distances relative to the TSS (left of zero is away from the gene). Y-axis represents frequencies of nucleosome dyads normalized to the genome-wide average. Each of the 4 gene bins is represented by a line of a corresponding color displayed in the legend.



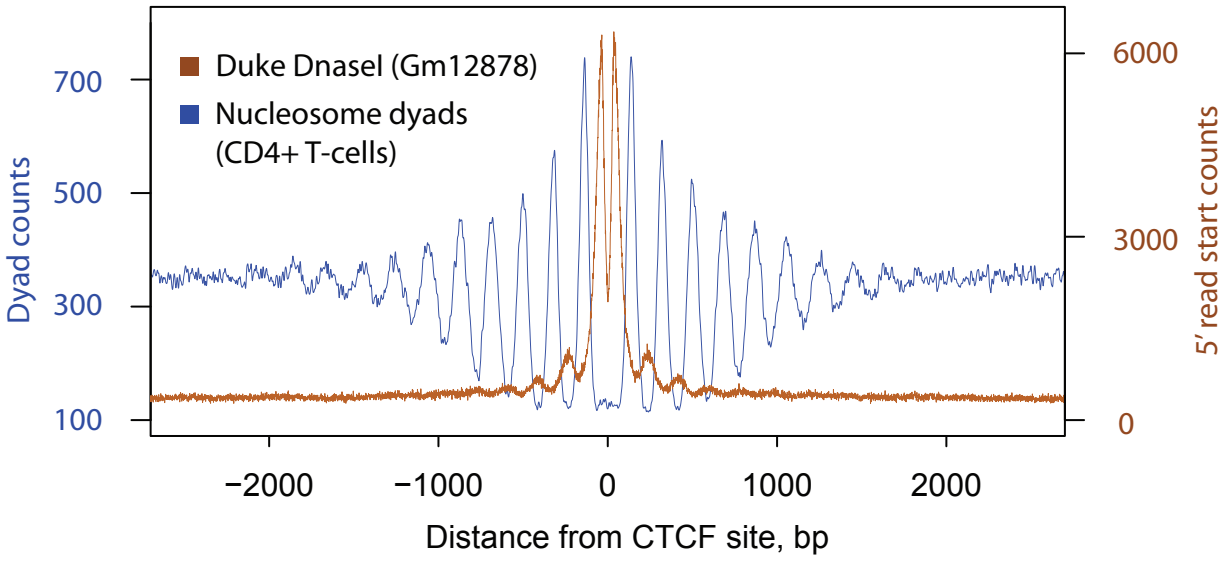
Supplementary Figure 9. Nucleosome organization around CTCF binding sites. (**A**) Schematic depiction of nucleosome organization inferred from the data. The blue ovals represent *in vivo* nucleosome positions, the green square represents binding of CTCF protein which is flanked by two well-positioned nucleosoomes. The orange oval represents preferred position of nucleosomes *in vitro*. (**B**) Dyad frequencies around CTCF binding site. Binidng sites were aligned so that position 0 represents coordinate of CTCF binding inferred from CTCF data in CD4+ T-cells. X-axis represents 4 Kbp window around CTCF binding site, Y-axis represents normlized frequencies of dyads across the regions. The red curve represents smoothed frequency of nucleosome dyads from granulocytes, the blue curve represents smoothed nucleosome dyad frequency in CD4+ T-cells. (**C**) Dyad frequencies in the *in vitro* reconstitution data around CTCF binding sites.



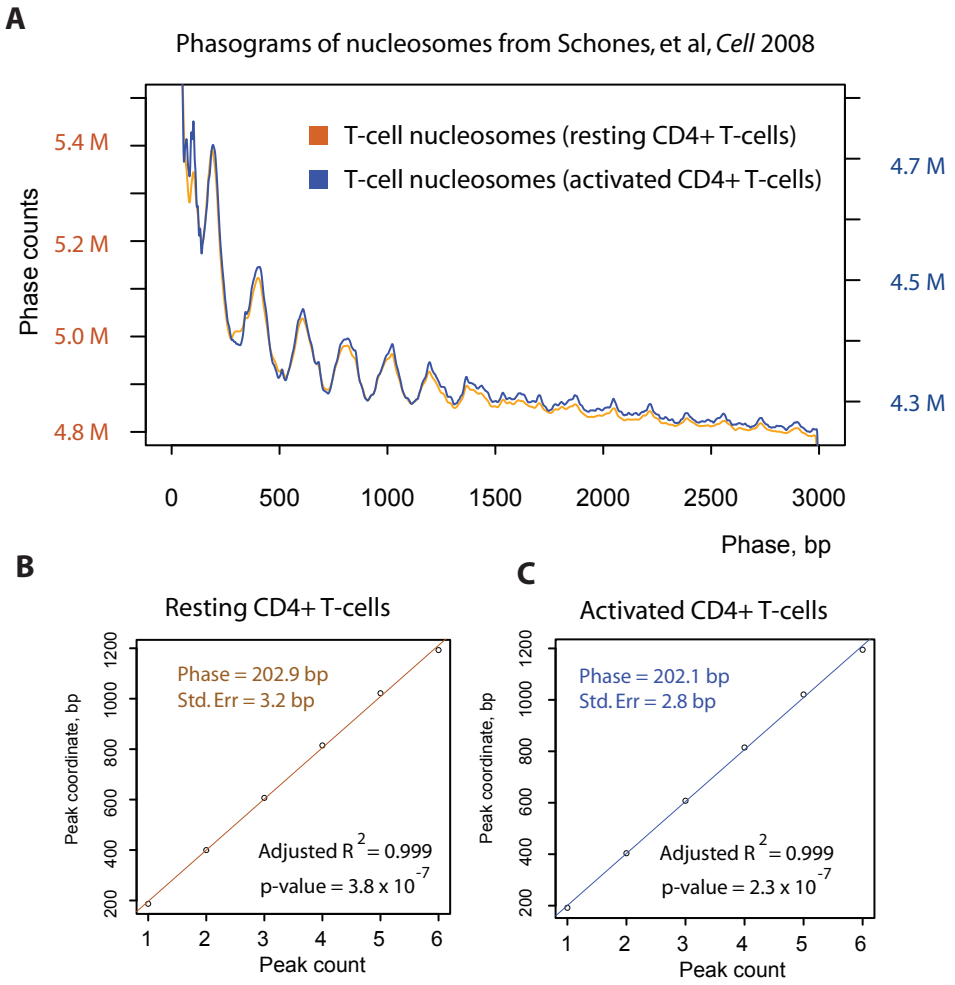
Supplementary Figures 10. Isolation of nucleosome-bound DNA. Agarose gels of nucleosome-bound DNA after micrococcal-treatment in CD4+ T-cells (**A**), CD8+ T-cells (**B**), Granulocytes (**C**), and in vitro reconstituted nucleosomes (**D**). Bands isolated for sequencing are marked by red rectangles.



Supplementary Figure 11. Micrococcal nuclease sequence bias anaysis. Shown are Weblogos (Crooks et al 2004) across sites cleaved by micrococcal nuclease in the control data (A), *in vivo* nucleosome data (B-D), and *in vitro* nucleosome data (E-G). We examined sites containing nucleosomes of increasing positioning strength (Pile1, sites with 1 or more read starts on the same strand; Pile3, sites with 3 or more read starts; Pile5, sites with 5 or more reads starts). For each subset, we aligned start positions and plotted nucleotide frequency at corresponding sites, with 0 representing the first sequenced base of the fragments. For the sites containing positioned *in vitro* nucleosomes (stringency > 0.5 and > 0.7), we plotted nucleotide frequencies from overlapping nucleosome fragments.



Supplementary Figure 12. Chromatin structure around CTCF sites. We plotted Dnase I cutting frequency (brown) and dyad frequencies (blue) around CTCF binding sites. Dnase I cleavage frequency is represented by plotting frequency of 5′ ends from Dnase I sequence reads using Duke Dnase-seq protocol (Song and Crawford, 2010) in the lymphoblastoid cell line. Peaks of Dnase I are in strong counter-phase with dyads, representing cleavage sites localizing within the nucleosome linker DNA. In addition, a strong peak of Dnase I can be seen between the CTCF binding site and the first wel-positioned nuclesome.



Supplementary Figure 13. Nucleosome spacing in resting and activated T-cells. (**A**) Phasograms of nucleosomes in resting and activated T-cells (Schones et al, 2008). Nucleosome spacing was estimated using a linear fit to peak positions in the corresponding phasograms. (**B**) Spacing was estimated to be 202.9 bps in resting T-cells, and (**C**) 202.1 bps in activated T-cells. These results provide independent replication of phasing estimates in CD4+ and CD8+ T cells (Fig. 1D).