Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position

Jason D Buenrostro[1–3], Paul G Giresi[2,3], Lisa C Zaba[2,3], Howard Y Chang[2,3] & William J Greenleaf[1]

We describe an assay for transposase-accessible chromatin using sequencing (ATAC-seq), based on direct *in vitro* transposition of sequencing adaptors into native chromatin, as a rapid and sensitive method for integrative epigenomic analysis. ATAC-seq captures open chromatin sites using a simple two-step protocol with 500–50,000 cells and reveals the interplay between genomic locations of open chromatin, DNA-binding proteins, individual nucleosomes and chromatin compaction at nucleotide resolution. We discovered classes of DNA-binding factors that strictly avoided, could tolerate or tended to overlap with nucleosomes. Using ATAC-seq maps of human CD4$^+$ T cells from a proband obtained on consecutive days, we demonstrated the feasibility of analyzing an individual's epigenome on a timescale compatible with clinical decision-making.

Eukaryotic genomes are hierarchically packaged into chromatin[1], and the nature of this packaging plays a central role in gene regulation[2,3]. Major insights into the epigenetic information encoded within the nucleoprotein structure of chromatin have come from high-throughput, genome-wide methods for separately assaying chromatin accessibility ('open chromatin')[4,5], nucleosome positioning[6–8] and transcription factor (TF) occupancy[9]. Though they are powerful, published protocols for existing methods require millions of cells as starting material, involve complex and time-consuming sample preparations, and cannot probe the interplay of nucleosome positioning, chromatin accessibility and TF binding simultaneously. These limitations give rise to three shortcomings. First, current methods can average over and 'drown out' heterogeneity in cellular populations. Second, cells must often be grown *ex vivo* to obtain sufficient material, and such conditions perturb the *in vivo* context and modulate the epigenetic state in unknown ways. Third, input requirements often prevent application of these assays to well-defined clinical samples, thereby precluding generation of 'personal epigenomes' on diagnostic timescales.

Here we report a robust and sensitive method for epigenomic profiling that can provide a multidimensional portrait of gene regulation. We used ATAC-seq to identify regions of open chromatin, identify nucleosome-bound and nucleosome-free positions in regulatory regions, and infer the positions of DNA-binding proteins using 'footprinting' in a B-cell line. We demonstrated that this method is compatible with clinical timescales and standard blood draws by observing the open-chromatin landscape of a healthy volunteer.

RESULTS

ATAC-seq probes chromatin accessibility with transposomes

Hyperactive Tn5 transposase[10,11], loaded *in vitro* with adaptors for high-throughput DNA sequencing, can simultaneously fragment and tag a genome with sequencing adaptors (previously described as "tagmentation"[11]). Because transposons have been shown to integrate into active regulatory elements *in vivo*[12], we hypothesized that transposition by purified Tn5, a prokaryotic transposase, on small numbers of unfixed eukaryotic nuclei would interrogate regions of accessible chromatin. In ATAC-seq, Tn5 transposase integrates its adaptor payload into regions of accessible chromatin, whereas steric hindrance in less accessible chromatin makes such transposition less probable. Therefore, amplifiable DNA fragments suitable for high-throughput sequencing are preferentially generated at locations of open chromatin (**Fig. 1a**). The entire assay and library construction can be carried out in a simple protocol involving Tn5 insertion followed by PCR. In contrast, published DNase-seq and formaldehyde-assisted isolation of regulatory elements with sequencing (FAIRE-seq) methods for assaying chromatin accessibility involve multistep protocols[13,14] and many potentially loss-prone steps, such as adaptor ligation, gel purification and cross-link reversal.

We carried out ATAC-seq on sets of 50,000 and 500 unfixed nuclei isolated from the GM12878 human lymphoblastoid cell line (Encyclopedia of DNA Elements (ENCODE) Tier 1; ref. 15) for comparison and validation with chromatin-accessibility data sets, including DNase-seq[13] and FAIRE-seq[16], which were generated from 1–50 million cells (**Fig. 1b**). At a locus previously highlighted by others[5] (**Fig. 1c**), ATAC-seq had a signal-to-noise

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA. [2]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California, USA. [3]Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California, USA. Correspondence should be addressed to W.J.G. (wjg@stanford.edu) or H.Y.C. (howchang@stanford.edu).

RECEIVED 20 JUNE; ACCEPTED 29 AUGUST; PUBLISHED ONLINE 6 OCTOBER 2013; DOI:10.1038/NMETH.2688

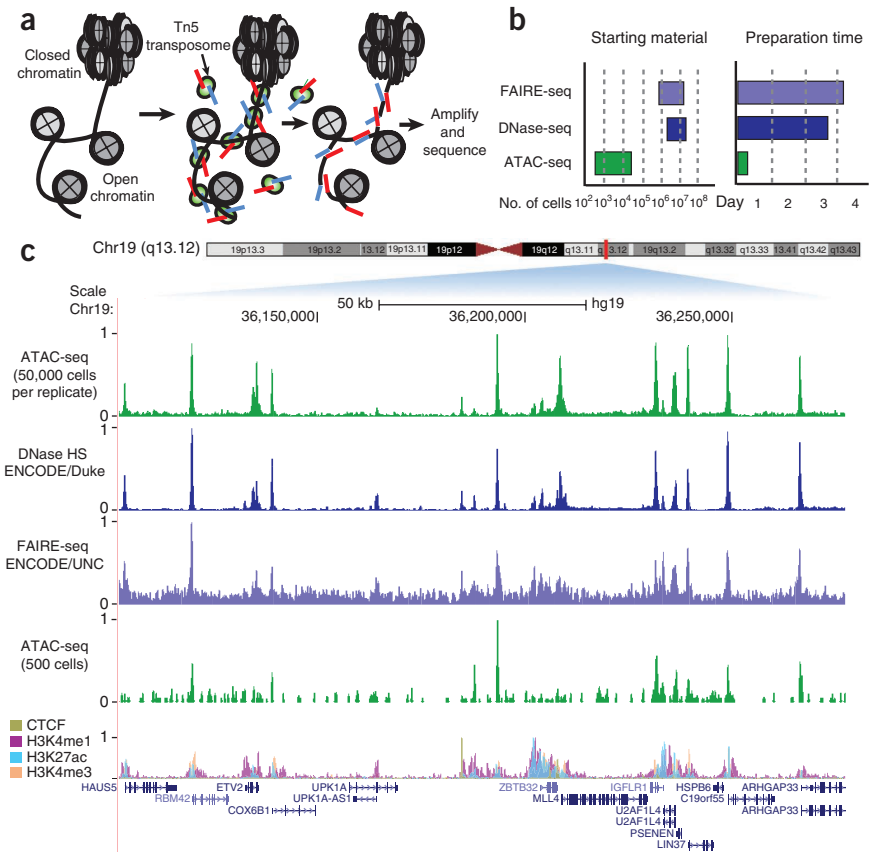

Figure 1 | ATAC-seq probes open-chromatin state. (**a**) ATAC-seq reaction schematic. Transposase (green), loaded with sequencing adaptors (red and blue), inserts only in regions of open chromatin (between nucleosomes in gray) and generates sequencing-library fragments that can be PCR-amplified. (**b**) Approximate reported input material and sample preparation time requirements for genome-wide methods of open-chromatin analysis. (**c**) Comparison of ATAC-seq to other open-chromatin assays at a locus in GM12878 lymphoblastoid cells. The lower ATAC-seq track was generated from 500 FACS-sorted cells. Bottom, composite locations of CTCF and histone modifications associated with active enhancers and promoters. DNase-seq (Duke) and FAIRE-seq (University of North Carolina, UNC) data were from the indicated ENCODE production centers.

ratio similar to that of DNase-seq, whose data were generated from ~3–5 orders of magnitude more cells[13]. Peak intensities were highly reproducible between technical replicates (R = 0.98) and highly correlated between ATAC-seq and two different sources of DNase-seq data (R = 0.79 and R = 0.83; **Supplementary Fig. 1**), and the majority of reads within peaks came from intersections of DNase and ATAC-seq peaks (**Supplementary Fig. 2**). We compared our data to DNase-hypersensitive sites identified in ENCODE DNase-seq data: receiver operating characteristic curves demonstrated a similar sensitivity and specificity for ATAC-seq and DNase-seq (**Supplementary Fig. 3**). ATAC-seq peak intensities also correlated well with markers of active chromatin and not with transposase sequence preference (**Supplementary Figs. 4** and **5**). Highly sensitive open-chromatin detection was maintained even when using 5,000 or 500 human nuclei as starting material (**Supplementary Figs. 3** and **6**), although sensitivity was diminished for smaller numbers of input material.

ATAC-seq insert sizes disclose nucleosome positions

We found that ATAC-seq paired-end reads produced detailed information about nucleosome packing and positioning. The insert size distribution of sequenced fragments from human chromatin had clear periodicity of approximately 200 bp, suggesting many fragments are protected by integer multiples of nucleosomes (**Fig. 2a**). This fragment size distribution also showed clear periodicity equal to the helical pitch of DNA[11]. By partitioning the insert size distribution according to functional classes of chromatin as defined by previous models[17] and normalizing to the global insert distribution (Online Methods), we observed clear class-specific enrichments across this insert size distribution (**Fig. 2b**). This result demonstrated that these functional states of chromatin have an accessibility fingerprint that can be read out with ATAC-seq. The differential fragmentation patterns were consistent with the putative functional state of these classes: CCCTC-binding factor (CTCF)-bound regions were enriched for short fragments of DNA, whereas transcription start sites (TSSs) were differentially depleted for mono-, di- and trinucleosome–associated fragments. Transcribed and promoter-flanking regions were enriched for longer multinucleosomal fragments, a result suggesting that such regions represent a form of chromatin that is less accessible than that of TSSs. Finally, prior studies have shown that certain DNA sequences are refractory to

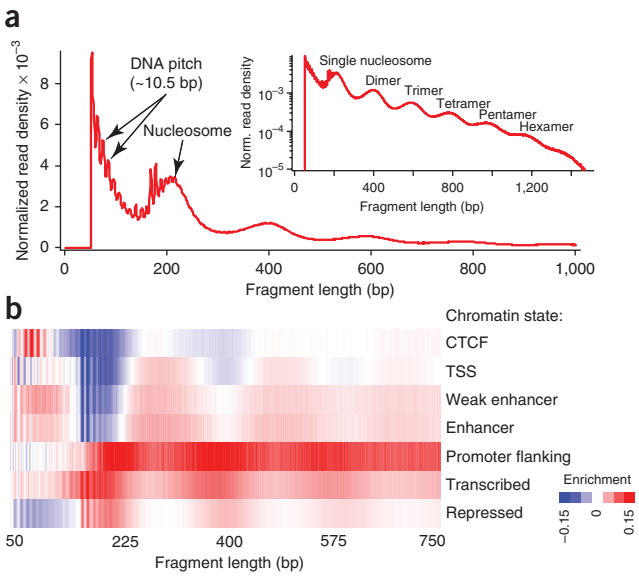

Figure 2 | ATAC-seq provides genome-wide information on chromatin compaction. (**a**) ATAC-seq fragment sizes generated from GM12878 nuclei. Inset, log-transformed histogram shows clear periodicity persists to six nucleosomes. (**b**) Normalized read enrichments for seven classes of chromatin state previously defined[17].

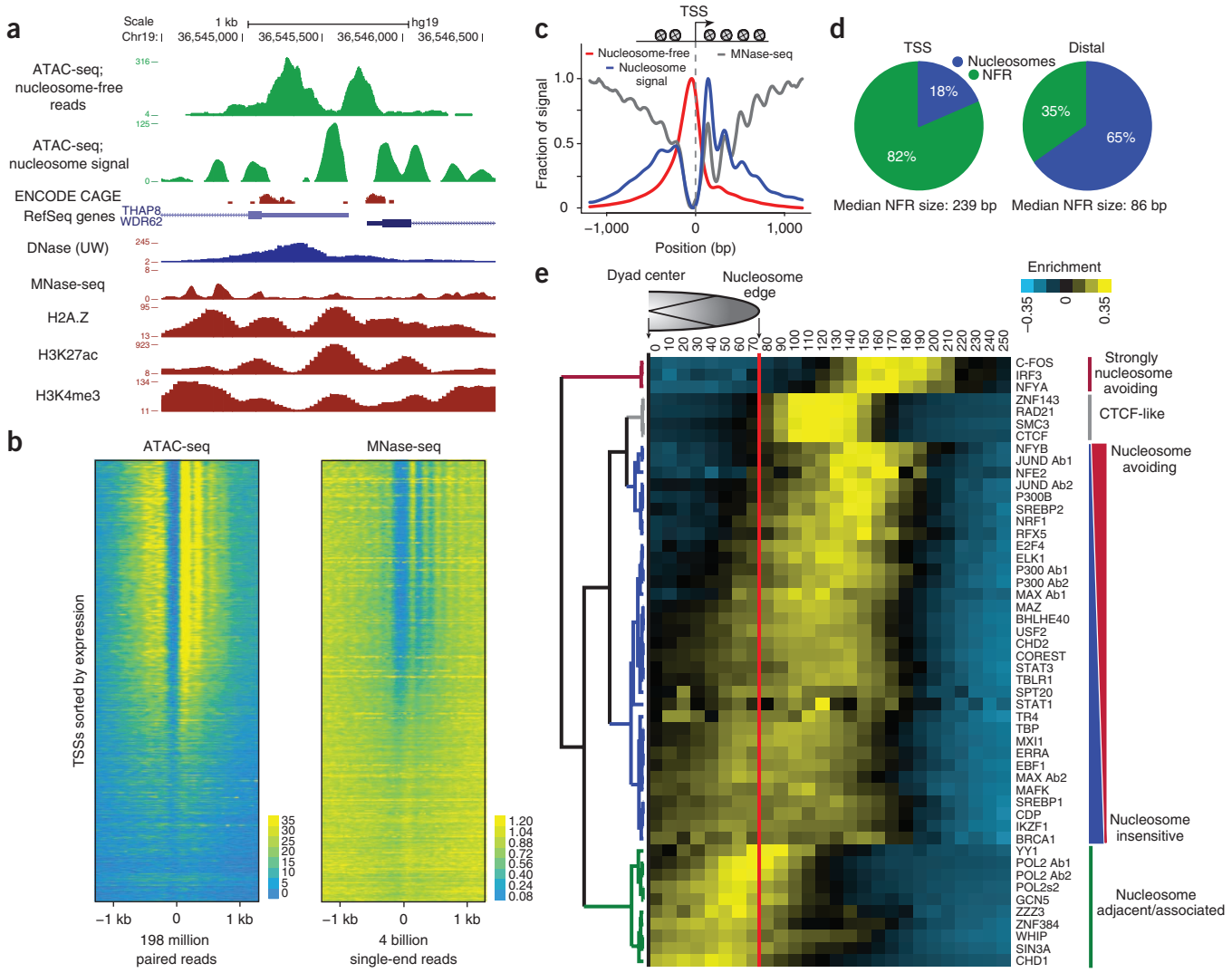

Figure 3 | ATAC-seq provides genome-wide information on nucleosome positioning in regulatory regions. (**a**) Locus containing two transcription start sites (TSSs) showing a nucleosome-free read track and calculated nucleosome track (Online Methods) as well as DNase, MNase, H3K27ac, H3K4me3 and H2A.Z tracks for comparison. DNase-seq (University of Washington, UW) data were from the indicated ENCODE production center. (**b**) ATAC-seq (198 million paired reads) and MNase-seq (4 billion single-end reads from ref. 23) nucleosome signal for all active TSSs (n = 64,836). TSSs are sorted by cap analysis of gene expression (CAGE) values. (**c**) TSSs are enriched for nucleosome-free fragments and show phased nucleosomes similar to those seen by MNase-seq at the –2, –1, +1, +2, +3 and +4 positions. (**d**) Relative fraction of nucleosome associated versus nucleosome-free (NFR) bases in TSS and distal sites (Online Methods). (**e**) Hierarchical clustering of DNA-binding factor position with respect to the nearest nucleosome dyad within accessible chromatin.

nuclease digestion and are released as large, multinucleosome-sized fragments[18]; subsequent studies showed that such fragments are condensed heterochromatin[19]. Indeed we found that repressed, untranscribed regions (as defined by the chromatin state model) are depleted for short fragments and enriched for phased multinucleosomal inserts, which is consistent with their expected inaccessible state. These data suggest that ATAC-seq reveals differentially accessible forms of chromatin, which have been long hypothesized to exist *in vivo*[2,20,21].

To explore nucleosome positioning within accessible chromatin in the GM12878 cell line, we partitioned our data into reads generated from putative nucleosome-free regions of DNA and reads likely derived from nucleosome-associated DNA (Online Methods and **Supplementary Fig. 7**). Using a simple heuristic that positively weights nucleosome-associated fragments and negatively weights nucleosome-free fragments (Online Methods), we calculated a data track used to call nucleosome positions within regions of accessible chromatin[22]. An example locus (**Fig. 3a**) contains a putative bidirectional promoter with cap analysis of gene expression (CAGE) data showing two TSSs separated by ~700 bp. ATAC-seq revealed in fact two distinct nucleosome-free regions separated by a single well-positioned mononucleosome (**Fig. 3a**). Compared to micrococcal nuclease digestion followed by sequencing (MNase-seq)[23] data, ATAC-seq data were more amenable to detecting nucleosomes within putative regulatory regions, as the majority of reads were concentrated within accessible regions of chromatin (**Fig. 3b**). By averaging signal across all active TSSs, we found that nucleosome-free fragments were enriched at a canonical nucleosome-free promoter region overlapping the TSS, whereas our nucleosome signal was

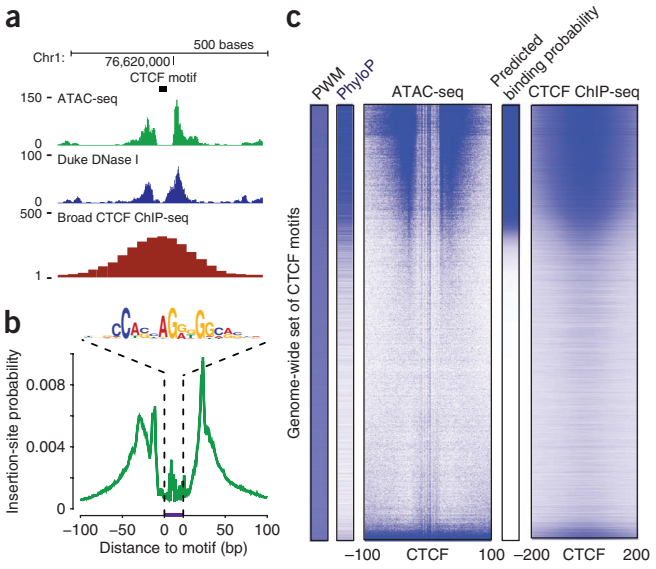

Figure 4 | ATAC-seq assays genome-wide factor occupancy. (**a**) CTCF footprints observed in ATAC-seq and DNase-seq data, at a specific locus on chromosome 1. DNase-seq (Duke) and ChIP-seq (Broad Institute) data were from the indicated ENCODE production centers. (**b**) Aggregate ATAC-seq footprint for CTCF (motif shown) generated over binding sites within the genome. (**c**) CTCF predicted binding probability inferred from ATAC-seq data, position weight matrix (PWM) scores for the CTCF motif, and evolutionary conservation (PhyloP). Shown in the rightmost column are the CTCF ChIP-seq data (ENCODE) for this GM12878 cell line.

enriched both upstream and downstream of the active TSS and displayed characteristic phasing of upstream and downstream nucleosomes[6,7,23] (**Fig. 3c**). Because ATAC-seq reads are concentrated at regions of open chromatin, we saw a strong nucleosome signal at the +1 nucleosome that decreased at the +2, +3 and +4 nucleosomes. In contrast, MNase-seq nucleosome signal increased at larger distances from the TSS, likely owing to overdigestion of more accessible nucleosomes. Additionally, MNase-seq (4 billion reads) assays all nucleosomes, whereas reads generated from ATAC-seq (198 million paired reads) are concentrated at regulatory nucleosomes (**Fig. 3b**,**c**). Using our nucleosome calls, we further partitioned putative distal regulatory regions and TSSs into regions that were nucleosome free and regions that were predicted to be nucleosome bound. We note that TSSs were enriched for nucleosome-free regions when compared to distal elements, which tend to remain nucleosome rich (**Fig. 3d**). These data suggest ATAC-seq can provide high-resolution readouts of nucleosome-associated and nucleosome-free regions in regulatory elements genome wide.

ATAC-seq reveals patterns of nucleosome-TF spacing

Using chromatin immunoprecipitation followed by sequencing (ChIP-seq) data, we plotted the position of a variety of DNA-binding factors with respect to the dyad of the nearest nucleosome derived from ATAC-seq data. Unsupervised hierarchical clustering (**Fig. 3e**) revealed major classes of binding with respect to the proximal nucleosome, including (i) a strongly nucleosome-avoiding group of factors with binding events stereotyped at ~180 bases from the nearest nucleosome dyad (comprising C-FOS, NFYA and IRF3), (ii) a class of factors that 'nestle up' next to the expected end of nucleosome DNA contacts, which notably includes chromatin looping factor CTCF and cohesin-complex subunits RAD21 and SMC3, (iii) a large class primarily of TFs that have gradations of nucleosome-avoiding or nucleosome-overlapping binding behavior and (iv) a class whose binding sites tend to overlap nucleosome-associated DNA. Interestingly, this final class includes chromatin-remodeling factors such as CHD1 and SIN3A as well as RNA polymerase II, which appears to be enriched at the nucleosome boundary[8]. The interplay between precise nucleosome positioning and locations of DNA-binding factor immediately suggests specific hypotheses for mechanistic studies.

ATAC-seq footprints infer factor occupancy genome wide

We reasoned that DNA sequences directly occupied by DNA-binding proteins are protected from transposition; the resulting sequence 'footprint' reveals the presence of a DNA-binding protein at each site, analogous to DNase digestion footprints[24]. At a specific CTCF-binding site on chromosome 1, we observed a clear footprint (a deep notch of ATAC-seq signal) similar to footprints seen with DNase-seq[25,26], at the precise location of the CTCF motif that coincides with the summit of the CTCF ChIP-seq signal in GM12878 cells (**Fig. 4a**). We averaged ATAC-seq signal over all expected locations of CTCF within the genome and observed a well-stereotyped footprint (**Fig. 4b**). Similar results were obtained for a variety of common TFs (**Supplementary Fig. 8**). We inferred the CTCF binding probability from motif consensus score, evolutionary conservation and ATAC-seq insertion data to generate a posterior probability of CTCF binding at all loci[27] (**Fig. 4c**). Results using ATAC-seq closely recapitulated ChIP-seq binding data in this cell line and compared favorably to DNase-based factor occupancy inference (**Supplementary Fig. 9**), suggesting that factor occupancy data can be extracted from these ATAC-seq data, which would allow reconstruction of regulatory networks.

ATAC-seq enables epigenomic analysis on clinical timescales

Because ATAC-seq is rapid, information rich and compatible with small numbers of cells, we reasoned it may serve as a powerful tool to look at aspects of an individual's epigenome on a timescale that would make its application to the clinic feasible. We applied ATAC-seq to assay the epigenome of a healthy volunteer's T cells obtained via standard serial blood draws. With rapid T-cell enrichment and sample handling protocols (Online Methods), the total required time from blood draw to sequencing was approximately 275 min (**Fig. 5a**). We performed this procedure on three consecutive days (**Fig. 5b**) and investigated the ATAC-seq profile at the *IL2* locus to show how regulatory information on an individual epigenome can be obtained. IL-2 is a key cytokine that drives T-cell growth and functions in inflammatory and autoimmune diseases[28]. Furthermore, distinct drugs[29–31] inhibit the activities of different TFs that bind putative *IL2* enhancers in a context-dependent manner. One might wish to identify the causal TF pathway in order to rationally target inhibition without exposing the patient to drugs unlikely to serve the therapeutic goal of IL-2 blockade. ATAC-seq showed that in the proband's T cells, only nuclear factor of activated T cells (NFAT), but not two other drug targets, engaged *IL2* (**Fig. 5c**). Thus, ATAC-seq was able to provide clinically relevant information on the regulatory state of this individual.

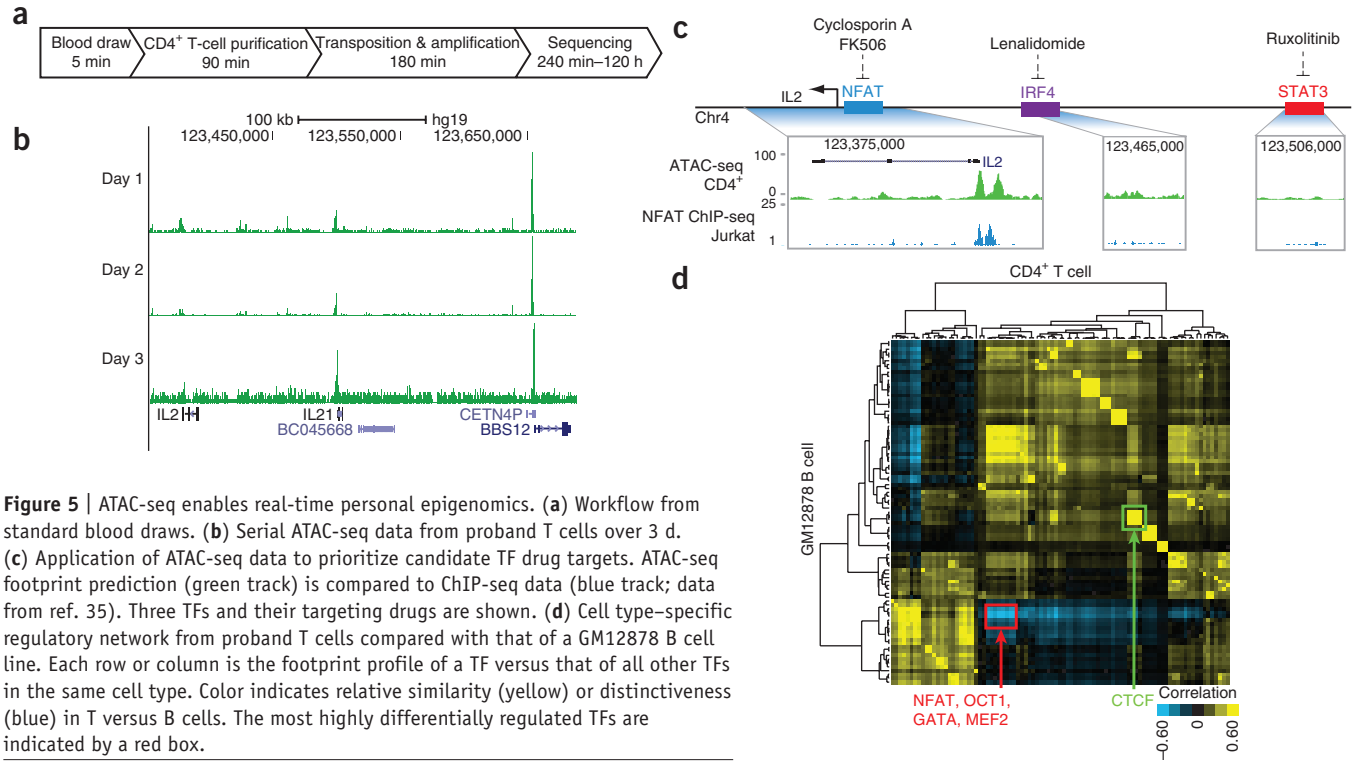

Figure 5 | ATAC-seq enables real-time personal epigenomics. (**a**) Workflow from standard blood draws. (**b**) Serial ATAC-seq data from proband T cells over 3 d. (**c**) Application of ATAC-seq data to prioritize candidate TF drug targets. ATAC-seq footprint prediction (green track) is compared to ChIP-seq data (blue track; data from ref. 35). Three TFs and their targeting drugs are shown. (**d**) Cell type–specific regulatory network from proband T cells compared with that of a GM12878 B cell line. Each row or column is the footprint profile of a TF versus that of all other TFs in the same cell type. Color indicates relative similarity (yellow) or distinctiveness (blue) in T versus B cells. The most highly differentially regulated TFs are indicated by a red box.

Using ATAC-seq footprints, we generated the occupancy profiles of 89 TFs in proband T cells, enabling systematic reconstruction of regulatory networks. With this personalized regulatory map, we compared the genomic distribution of the same 89 TFs between GM12878 and proband CD4$^+$ T cells. TFs that exhibited large variation in distribution between T cells and B cells were enriched for T cell–specific factors (**Fig. 5d**). This analysis showed that NFAT was differentially regulating, whereas canonical CTCF occupancy was highly correlated within these two cell types (**Fig. 5d**). Supporting this interpretation, we noted specific loci where NFAT was localized to known T cell–specific genes such as *CD28* and the uncharacterized large intergenic noncoding RNA RP11-229C3.2 (**Supplementary Fig. 10**). Additionally, ATAC-seq of CD4$^+$ and CD8$^+$ T cells and monocytes isolated by FACS from a single blood draw created a map of accessible chromatin and demonstrated that ATAC-seq is compatible with cellular enrichment using surface markers (**Supplementary Fig. 11**). Separately, allele-specific chromatin accessibility has been shown to be particularly relevant to our understanding of human disease[32]. As a proof of principle we also used ATAC-seq to identify candidate allele-specific open-chromatin regions within the GM12878 cell line (**Supplementary Fig. 12**). These results demonstrate the feasibility of generating detailed personalized gene regulatory networks from clinical samples, opening the door for future diagnostic applications.

DISCUSSION

ATAC-seq allows simultaneous interrogation of factor occupancy, nucleosome positions in regulatory sites, and chromatin accessibility genome wide by considering the position of insertion and the distribution of insert lengths captured during the transposition reaction. Although extant methods such as DNase-seq and MNase-seq can also provide this information, they each require separate assays with large cell numbers. ATAC-seq also provides insert-size fingerprints of biologically relevant genomic regions, which suggests that it captures information on chromatin compaction. We expect ATAC-seq to have broad applicability, to considerably add to the genomics toolkit and to improve our understanding of gene regulation, particularly when integrated with other powerful rare-cell techniques such as FACS or laser-capture microdissection and recent advancements in RNA-seq[33,34].

One potentially exciting application of ATAC-seq is to generate personal epigenomic profiles on a timescale compatible with clinical decision-making. We envision that when ATAC-seq is coupled with ongoing improvements to sequencing and analysis turnaround times, this technique will offer the possibility of a day-long turnaround time for a personal epigenomic map. Deeper analyses to generate global regulatory networks and other inferences will take additional time, and we anticipate that bioinformatic analyses—not the molecular biology or sequencing—will become the bottleneck of epigenomic studies in the future. As ATAC-seq is compatible with FACS, it may enable studies on carefully sorted and rare subpopulations from primary tissues. We expect that ATAC-seq can also be applied to study select cellular subpopulations at different points during development and aging and during the progression of human diseases, including cancer, autoimmunity and neuropsychiatric disorders.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. NCBI Gene Expression Omnibus: raw data are available at accession number GSE47753.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of Greenleaf and Chang labs for discussion, A. Burnet and S. Kim lab and the Stanford flow-cytometry core facility for assistance with FACS sorting, A. Schep for modeling Tn5 insertion preference, and V. Risca for graphics. This work was supported by the US National Institutes of Health (H.Y.C., W.J.G. and J.D.B.), including RC4NS073015, U01DK089532 and U19AI057229; Scleroderma Research Foundation (H.Y.C.); and California Institute for Regenerative Medicine (H.Y.C.). H.Y.C. acknowledges support as an Early Career Scientist of the Howard Hughes Medical Institute. GM12878 cells were a gift from the Snyder laboratory (Stanford University).

AUTHOR CONTRIBUTIONS

J.D.B., P.G.G. and L.C.Z. performed the research. All authors designed experiments and interpreted the data. H.Y.C. and W.J.G. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Kornberg, R.D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868–871 (1974).
2. Kornberg, R.D. & Lorch, Y. Chromatin structure and transcription. *Annu. Rev. Cell Biol.* **8**, 563–587 (1992).
3. Mellor, J. The dynamics of chromatin remodeling at promoters. *Mol. Cell* **19**, 147–157 (2005).
4. Boyle, A.P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
5. Thurman, R.E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
6. Schones, D.E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
7. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011).
8. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
9. Gerstein, M.B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
10. Goryshin, I.Y. & Reznikoff, W.S. Tn5 *in vitro* transposition. *J. Biol. Chem.* **273**, 7367–7374 (1998).
11. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol.* **11**, R119 (2010).
12. Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S.J. & Craig, N.L. DNA transposon *Hermes* inserts into DNA in nucleosome-free regions *in vivo*. *Proc. Natl. Acad. Sci. USA* **107**, 21966–21972 (2010).
13. Song, L. & Crawford, G.E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* **2010** pdb.prot5384 (2010).
14. Simon, J.M., Giresi, P.G., Davis, I.J. & Lieb, J.D. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nat. Protoc.* **7**, 256–267 (2012).
15. The ENCODE Project Consortium. A user's guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
16. Giresi, P.G. & Lieb, J.D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods* **48**, 233–239 (2009).
17. Hoffman, M.M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* **41**, 827–841 (2013).
18. Prioleau, M.-N., Nony, P., Simpson, M. & Felsenfeld, G. An insulator element and condensed chromatin region separate the chicken β-globin locus from an independently regulated erythroid-specific folate receptor gene. *EMBO J.* **18**, 4035–4048 (1999).
19. Ghirlando, R., Litt, M.D., Prioleau, M.-N., Recillas-Targa, F. & Felsenfeld, G. Physical properties of a genomic condensed chromatin fragment. *J. Mol. Biol.* **336**, 597–605 (2004).
20. Kornberg, R.D. & Lorch, Y. Chromatin and transcription: where do we go from here. *Curr. Opin. Genet. Dev.* **12**, 249–251 (2002).
21. Zhou, J., Fan, J.Y., Rangasamy, D. & Tremethick, D.J. The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat. Struct. Mol. Biol.* **14**, 1070–1076 (2007).
22. Chen, K. *et al.* DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351 (2013).
23. Kundaje, A. *et al.* Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res.* **22**, 1735–1747 (2012).
24. Hesselberth, J.R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
25. Boyle, A.P. *et al.* High-resolution genome-wide *in vivo* footprinting of diverse transcription factors in human cells. *Genome Res.* **21**, 456–464 (2011).
26. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
27. Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21**, 447–455 (2011).
28. Fraser, J.D., Irving, B.A., Crabtree, G.R. & Weiss, A. Regulation of interleukin-2 gene enhancer activity by the T cell accessory molecule CD28. *Science* **251**, 313–316 (1991).
29. Flanagan, W.M., Corthésy, B., Bram, R.J. & Crabtree, G.R. Nuclear association of a T-cell transcription factor blocked by FK-506 and cyclosporin A. *Nature* **352**, 803–807 (1991).
30. Lopez-Girona, A. *et al.* Lenalidomide downregulates the cell survival factor, interferon regulatory factor-4, providing a potential mechanistic link for predicting response. *Br. J. Haematol.* **154**, 325–336 (2011).
31. Verstovsek, S. *et al.* Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. *N. Engl. J. Med.* **363**, 1117–1127 (2010).
32. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory. *Science* **337**, 1190–1195 (2012).
33. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
34. Shalek, A.K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
35. Jolma, A. *et al.* Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.* **20**, 861–873 (2010).

ONLINE METHODS

ATAC-seq protocol. The ATAC-seq protocol has three major steps.

1. Prepare nuclei. To prepare nuclei, we spun 50,000 cells at 500*g* for 5 min, which was followed by a wash using 50 µL of cold 1× PBS and centrifugation at 500*g* for 5 min. Cells were lysed using cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% IGEPAL CA-630). Immediately after lysis, nuclei were spun at 500*g* for 10 min using a refrigerated centrifuge. To avoid losing cells during the nuclei prep, we used a fixed-angle centrifuge and carefully pipetted away from the pellet after centrifugations.

2. Transpose and purify. Immediately following the nuclei prep, the pellet was resuspended in the transposase reaction mix (25 µL 2× TD buffer, 2.5 µL transposase (Illumina) and 22.5 µL nuclease-free water). The transposition reaction was carried out for 30 min at 37 °C. Directly following transposition the sample was purified using a Qiagen MinElute kit.

3. PCR. Following purification, we amplified library fragments using 1× NEBnext PCR master mix and 1.25 µM of custom Nextera PCR primers 1 and 2 (**Supplementary Table 1**), using the following PCR conditions: 72 °C for 5 min; 98 °C for 30 s; and thermocycling at 98 °C for 10 s, 63 °C for 30 s and 72 °C for 1 min. To reduce GC and size bias in our PCR, we monitored the PCR reaction using qPCR in order to stop amplification before saturation. To do this, we amplified the full libraries for five cycles, after which we took an aliquot of the PCR reaction and added 10 µl of the PCR cocktail with Sybr Green at a final concentration of 0.6×. We ran this reaction for 20 cycles to determine the additional number of cycles needed for the remaining 45-µL reaction. The libraries were purified using a Qiagen PCR cleanup kit yielding a final library concentration of ~30 nM in 20 µL. Libraries were amplified for a total of 10–12 cycles.

Low–cell number protocol. To prepare the 500- and 5,000-cell reactions, we used the same protocol with some notable exceptions: the transposition reaction was done in a 5-µL instead of 50-µL reaction. Also, we eliminated the Qiagen MinElute purification before PCR and instead took the 5-µL reaction immediately after transposition directly into the 50-µL PCR.

Library QC and quantitation. During the ATAC-seq protocol, we chose to avoid a size-selection step to maximize the library complexity. The sequenced insert size is a distribution between 40 bp and 1 kb with a mean of ~120 bp. From the bioanalyzer and gels we observed fragments >2 kb, which would make Qubit and other mass-based quantitation methods hard to interpret. For this reason we quantified our libraries using qPCR-based methods.

$CD4^+$ enrichment from peripheral blood. One green-top tube of whole blood was obtained from one normal volunteer three times over a 72-h period, under a Stanford University IRB-approved protocol. Informed consent was obtained. 5 mL of blood at each time point was negatively selected for $CD4^+$ cells using RosetteSep Human $CD4^+$ T Cell Enrichment Cocktail (StemCell Technologies). RosetteSep cocktail was incubated with the blood at 50 µL/mL for 20 min, diluted in an equal volume of PBS with 2% FBS and underlaid with 15 mL Ficol-Paque Plus (GE). Blood was centrifuged for 20 min at 1,200*g* without break, negatively selected cells were removed from the density medium–plasma interface and cells were washed twice in PBS with 2% FBS.

FACS of peripheral blood leukocytes and GM cells. GM 12878 cells were stained with DAPI NucBlue Fixed Cell Stain (Molecular Probes), and live cells were sorted using a FACSAria (BD Biosciences) with a 100-µm nozzle. One peripheral blood sample (buffy coat) was stained with BD Bioscience antibodies CD14-A-488 (M5E2, diluted 1:20), CD3-PE-Cy7 (SK7, 1:20), CD4-APC-Cy7 (RPA-T4, 1:20) and CD8 (RPA-T8, 1:20) for 20 min in the dark at RT. Cells were lysed using BDpharmLyse at 1:10 dilution in diH_20 (BD) for 15 min, centrifuged for 5 min, washed with PBS 2% FBS twice and resuspended in PBS with 2% FBS. 50,000 $CD3^+CD8^+$, $CD3^+CD4^+$ and $CD14^+$ cell populations were sorted into PBS with 10% FBS.

Primary data processing. Data were collected using either 34 × 8 × 34 reads from a MiSeq or 50 × 8 × 50 reads on a HiSeq. Reads were aligned to hg19 using Bowtie[36] with the parameters -X2000 and -m1. These parameters ensured that fragments up to 2 kb were allowed to align (-X2000) and that only unique aligning reads were collected (-m1). For all data files, duplicates were removed using Picard.

For peak-calling and footprinting, we adjusted the read start sites to represent the center of the transposon binding event. Previous descriptions of the Tn5 transposase show that the transposon binds as a dimer and inserts two adaptors separated by 9 bp (ref. 11). Therefore, all reads aligning to the + strand were offset by +4 bp, and all reads aligning to the – strand were offset –5 bp.

ATAC-seq peak-calling. We used ZINBA to call all reported ATAC-seq peaks in this manuscript. ZINBA was run using a window size of 300 bp and an offset of 75 bp. Alignability was used to model the zero-inflated component, and the ATAC-seq read count for the background and enriched components. Enriched regions were identified as those with a posterior probability of >0.8.

ATAC-seq insertion size enrichment analysis within chromatin annotations. First, the distribution of paired-end sequencing fragment sizes overlapping each chromatin state (http://www.ensembl.org/info/docs/funcgen/regulatory_segmentation.html) were computed. The distributions were then normalized to the percent maximal within each state, and enrichment was computed relative to the genome-wide set of fragment sizes.

Nucleosome positioning. To generate the nucleosome-position data track, we chose to split reads into various bins. Reads below 100 bp were considered nucleosome free, reads between 180 and 247 bp were considered to be mononucleosomes, reads between 315 and 473 bp were considered to be dinucleosomes, and reads between 558 and 615 bp were considered to be trinucleosomes (for determining cutoffs, see **Supplementary Fig. 7**). Dinucleosome reads were split into two reads, and trinucleosome reads were split into three reads. Reads were analyzed using Danpos and Dantools[22] using the parameters -p 1, -a 1, -d 20, -clonalcut 0. The background used was nucleosome-free reads (reads less than 100 bp), allowing an effective negative weighting of these reads.

This analysis allows calling multiple overlapping nucleosomes. Although generating nucleosome tracks using simple insert size cutoffs may yield false positives due to other nucleosome-sized features, i.e., enhanceosomes, we observed that we faithfully recapitulated global features on nucleosome position genome wide (**Fig. 3c**,**d**).

ChIP-seq peak-calling and clustering. ChIP-seq data were downloaded from the UCSC ENCODE repository; for a complete list of data used, see **Supplementary Table 2**. Peaks where called using GEM[37], and the parameters used where -k_min 6 -k_max 20. Inputs were used as a control for peak-calling. Binding events were annotated by distance to the nearest dyad in bins of 10 bps. Factors were then hierarchically clustered using Euclidean distance and normalized by gene and centered by mean[38].

Footprinting using CENTIPEDE. The genome-wide set of motifs was obtained from the ENCODE motif repository (http://compbio.mit.edu/encode-motifs/). The input for Centipede (ref. 27) included the PWM score, conservation (PhyloP) and ATAC-seq counts within ±100 bp of each genomic region matching a motif. ChIP-seq data were obtained from the UCSC ENCODE repository.

Comparison of transcription factor regulatory networks. Transcription factor regulatory networks were constructed by comparing the GENCODE v.14 genes with the genome-wide set of posterior probabilities estimated by Centipede for the respective cell types. The extent to which a transcription factor regulated each gene was determined by taking the sum of the weighed posterior probabilities for a given transcription factor mapping to the same chromosome. For each mapped motif, the posterior probability was weighted on the basis of the distance to the transcription start site for each gene. Comparison of transcription factor regulatory networks was computed as the correlation of each transcription factor in a given cell type with all transcription factors in the other cell type. The resulting correlation matrix was hierarchically clustered using the Pearson correlation coefficient and complete linkage[38].

Candidate *IL2* enhancer analysis. We inspected ENCODE data on the UCSC genome browser to identify putative *IL2* enhancers in one or more cell types that may be responsive to FDA-approved immunomodulatory drugs. We scanned the intergenic region upstream of *IL2* in hg19 for (i) enhancer-associated histone marks (H3K4me1 and H3K27ac), (ii) binding by one or more TFs as confirmed by ChIP-seq and (iii) TF pathways targetable by a human therapeutic. This analysis identified IRF4- and STAT3-binding sites in addition to the known NFAT-responsive elements (ref. 28).

36. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
37. Guo, Y., Mahony, S. & Gifford, D.K. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput. Biol.* **8**, e1002638 (2012).
38. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).

doi:10.1038/nmeth.2688