Genome-wide determination of drug localization

Published in final edited form as:
Nat Biotechnol. 2014 January ; 32(1): 92–96. doi:10.1038/nbt.2776.

Genome-wide determination of drug localization

Lars Anders[1,6], **Matthew G. Guenther**[1,6], **Jun Qi**[2], **Zi Peng Fan**[1,3], **Jason J. Marineau**[2], **Peter B. Rahl**[1], **Jakob Lovén**[1], **Alla A. Sigova**[1], **William B. Smith**[2], **Tong Ihn Lee**[1], **James E. Bradner**[2,4], and **Richard A. Young**[1,5]

James E. Bradner: james_bradner@dfci.harvard.edu; Richard A. Young: young@wi.mit.edu

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

[2]Department of Medical Oncology, Dana-Farber Cancer Institute, Massachusetts 02115, USA

[3]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[4]Department of Medicine, Harvard Medical School, Massachusetts 02115, USA

[5]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Abstract

A vast number of small-molecule ligands, including therapeutic drugs under development and in clinical use, elicit their effects by binding specific proteins associated with the genome. An ability to map the direct interactions of a chemical entity with chromatin genome-wide could provide new and important insights into chemical perturbation of cellular function. Here we describe a method that couples ligand-affinity capture and massively parallel DNA sequencing (Chem-seq) to identify the sites bound by small chemical molecules throughout the human genome. We show how Chem-seq can be combined with ChIP-seq to gain unique insights into the interaction of drugs with their target proteins throughout the genome of tumor cells. These methods provide a powerful approach to enhance understanding of therapeutic action and characterize the specificity of chemical entities that interact with DNA or genome-associated proteins.

The ability to map the locations of proteins throughout the genome has had a profound impact on our understanding of a wide range of normal and disease biology. For example, discovery of the genome-wide location of proteins using ChIP-seq has allowed global

[6]These authors contributed equally to this work

Accession Codes: Chem-seq and ChIP-seq data, GEO: GSE44098 and GSE43743, respectively.

Author Contributions: J.J.M., P.B.R., J.E.B. and R.A.Y. conceived of the Chem-seq method, L.A. and M.G.G. developed the method, L.A. generated bio-JQ1 and bio-psoralen Chem-seq data, M.G.G. generated bio-AT7519 Chem-seq data, Z.P.F. developed computational methods and analyzed the data, J.Q. and J.J.M. synthesized biotinylated derivatives of chemical probes, J.Q. performed protein biochemistry, L.A., P.B.R. and J.L. generated ChIP-seq data for BRD2, BRD3, BRD4, RNA pol II, CDK7, CDK8 and CDK9, W.B.S. generated cellular proliferation data, A.A.S. contributed to optimize Chem-seq, T.I.L. provided advice on method development, and J.E.B. and R.A.Y. supervised the research.

Competing Financial Interests: J.E.B. and R.A.Y. are founders of Syros Pharmaceuticals. J.J.M., P.B.R., M.G.G. and J.L. are employees of Syros Pharmaceuticals.

mapping of the key transcription factors and chromatin regulators that control gene expression programs in various cells, the sites that act as origins of DNA replication, and regions of the genome that form euchromatin and heterochromatin[1-6]. Models of the transcriptional regulatory circuitry that controls normal and disease cell states have emerged from genome-wide data[7-10].

An ability to map the global interactions of a chemical entity with chromatin genome-wide could provide new insights into the mechanisms by which a small molecule influences cellular functions. Many DNA-associated processes are targeted for disease therapy, including transcription, modification, replication and repair[11-16]. Ligand-affinity methodologies have greatly contributed to our understanding of drug and ligand function at the genome, and have led to the identification of numerous gene regulatory drug targets[17-20]. There have been initial efforts to map the sites of interaction of metabolic compounds in the yeast genome[21], but it would be ideal to have a method that allows investigators to determine how small-molecule therapeutics interact with the human genome. We describe here a method based on chemical affinity capture and massively parallel DNA sequencing (Chem-seq) that allows investigators to identify genomic sites where small chemical molecules interact with their target proteins or DNA (Fig. 1a). The Chem-seq method is similar to that employed for ChIP-seq, except that Chem-seq uses retrievable synthetic derivatives of a compound of interest to identify sites of genome occupancy whereas ChIP-seq uses antibodies against specific proteins for this purpose.

We used Chem-seq to investigate the genome-wide binding of the bromodomain inhibitor JQ1 to the BET bromodomain family members BRD2, BRD3 and BRD4 in MM1.S multiple myeloma cells. JQ1 was previously been shown to bind all three co-activator proteins and to inhibit growth of MM1.S and other tumor cells[13, 22-27]. We first investigated how BRD2, BRD3 and BRD4 occupy the genome of MM1.S cells using ChIP-Seq (Supplementary Fig. 1). All three proteins were found to be associated with actively transcribed genes (Supplementary Fig. 1a). Inspection of individual gene tracks (Supplementary Fig. 1b) and analysis of global genome occupancy (Supplementary Fig. 1c) showed that most core promoter elements of active genes were co-occupied by BRD2, BRD3 and BRD4 together with RNA polymerase II, the Mediator coactivator and histone H3K27Ac. In contrast, enhancers, which are occupied by histone H3K27Ac and Mediator, were preferentially occupied by BRD4, with lower relative levels of BRD2 and BRD3.

To investigate the interaction of JQ1 with chromatin genome-wide, we used the Chem-seq technique (Fig. 1a) with a biotinylated derivative of JQ1 (bio-JQ1, Fig. 1b). Enantioretentive substitution at C-6 of the JQ1 diazepine allowed coupling of a poly-ethylene glycol spacer with appended biotin feature. The potency of bio-JQ1 binding to the first bromodomain of BRD4 was nearly equivalent to the unbiotinylated compound, as determined by both differential scanning fluorimetry and isothermal titration calorimetry (Supplementary Fig. 2). Consistent with this, bio-JQ1 had only slightly reduced bioactivity in MM1.S cells relative to JQ1 (Fig. 1c). We initially treated living cells with bio-JQ1 and cross-linked proteins to DNA with formaldehyde (*in vivo* Chem-seq, Fig. 1a, upper panel). Cells were then lysed, sonicated to shear the DNA and streptavidin beads were used to isolate biotinylated ligand and associated chromatin fragments. Massively parallel sequencing was

used to identify enriched DNA fragments, and these sequences were mapped to the genome to reveal sites bound by the small molecule probe.

In addition, we developed an *in vitro* version of this method, which allows analysis of biotinylated molecules with potentially limited cell permeability (*in vitro* Chem-seq, Fig. 1a, lower panel). To this end, MM1.S cells were fixed and the derived sonicated lysate incubated with biotinylated JQ1 to enrich for bound chromatin regions *in vitro*.

We found that both *in vivo* and *in vitro* Chem-seq produced essentially the same result: the genomic sites bound by biotinylated JQ1 are highly similar to the sites occupied by BRD2, BRD3 and BRD4 (Fig. 1d, e). This was further confirmed by inspection of data at individual genes with pivotal roles in myeloma biology, such as CCND2 (Fig. 1f). By contrast, a functionally inactive enantiomer of bio-JQ1 (bio-JQ1R, Supplementary Fig. 3a) did not produce significant Chem-seq signals (Supplementary Fig. 3b, c). These results indicate that both live-cell and cell-lysate based Chem-seq approaches (Fig. 1a) can be used to uncover the interactions of small molecules with their chromatin targets across the human genome. Of note, JQ1 is known to displace BET bromodomains from the genome, but the ability to detect the bio-JQ1/BRD complex on chromatin is likely made possible by covalent tethering of these proteins to chromatin during fixation (Supplementary Fig. 4).

We next investigated the extent to which Chem-seq and ChIP-seq signals overlap (Fig. 2). The pattern of JQ1 occupancy was best associated with the pattern of BRD4 occupancy (Fig. 2a). Pearson correlation analysis also showed that bio-JQ1 signals were most highly correlated with BRD4, somewhat less frequently with BRD2 and much less frequently with BRD3 (Fig. 2b). We then developed a generalized linear model (GLM) to identify genomic regions with differential signal between bio-JQ1 Chem-seq and each of the BRD ChIP-seq datasets. We found that bio-JQ1 co-occupied nearly all regions (>99%) with BRD4 genome-wide across triplicate datasets, bio-JQ1 and BRD2 co-occupied 96% of all genomic sites, and bio-JQ1 and BRD3 co-occupied 63% of all genomic sites (Fig. 2c). Inspection of gene tracks for regions differentially occupied by bio-JQ1 and the three BET proteins provided visual confirmation that bio-JQ1 tends to co-occupy enhancers where there are substantial BRD4 signals and lower signals for BRD2 and BRD3 (Fig. 2d). The pattern of BRD3 genome occupancy differed most from that of the other two BET proteins (Fig. 2a–c), and this was due to pronounced signals at a subset of core promoter sites (Fig 2e). Similar results were obtained with an alternative BRD3 ChIP antibody directed against a different epitope of this protein (Supplementary Fig. 5). Taken together, these results indicate that the pattern of JQ1 occupancy of chromatin is most correlated with that of BRD4 in MM1.S cells, consistent with the relative affinities of JQ1 for these BET proteins previously established *in vitro*[27].

To extend the Chem-seq method to other drug classes, we initially focused on AT7519, an inhibitor of the cyclin-dependent kinase CDK9 (ref. [28]), which is associated with the transcription apparatus at promoters. CDK9 phosphorylation of RNA polymerase II and various pause control factors stimulates active elongation[29]. We first confirmed that CDK9 co-occupies the promoters of active genes with RNA polymerase II by using ChIP-seq (Fig. 3a, b). CDK9 is a core component of the positive transcription elongation factor, p-TEFb[29],

and its inhibition would be expected to affect the levels of elongating RNA polymerase II, which is located across the body of genes, to a much greater extent than the levels of initiating RNA polymerase located at the transcription start site. Indeed, treatment of MM1.S cells with AT7519 was found to cause a reduction in the level of elongating RNA polymerase II based on examination of individual gene tracks (Fig. 3c) and on analysis of the ratio of initiating versus elongating RNA polymerase II molecules at active genes throughout the genome (Fig. 3d). We next generated a retrievable biotinylated derivative of AT7519 (Fig. 3e). The biotinylated compound was found to have reduced ability to enter cells (Fig. 3f, g), so we used the *in vitro* Chem-seq method to investigate binding of bio-AT7519 to chromatin genome-wide. The results show that bio-AT7519 Chem-seq signals occur frequently at sites occupied by CDK9 (Fig. 3h, i). The bio-AT7519 Chem-seq signals were weaker than those observed for bio-JQ1, which may reflect differences in accessibility, association constants, or ligand-receptor sensitivity to sample preparation. Nonetheless, there was a correlation between bio-AT7519 occupancy and CDK9 occupancy genome-wide (Supplementary Fig. 6a). There were also a substantial number of sites that were not co-occupied by bio-AT7519 and CDK9; it is possible that this is due to the relatively weak signals we obtained for bio-AT7519 Chem-seq or to the fact that AT7519 can inhibit other kinases[28, 30] that may occupy other genomic sites (Supplementary Fig. 6b). Notably, a comparison of the Chem-seq data for bio-AT7519 and bio-JQ1 with ChIP-seq data for various components of the transcription apparatus (CDK7, CDK8, CDK9, RNA Polymerase II, Mediator and BRD4) revealed that bio-AT7519 was most associated with CDK9, whereas bio-JQ1 was most associated with BRD4 (Supplementary Fig. 6c). These results suggest that Chem-seq can be useful for identifying the genomic binding sites of kinase inhibitors.

To further extend the Chem-seq method to other drug classes, we investigated how the DNA intercalator psoralen interacts with genomic DNA *in vivo*. Recent studies have shown that psoralen preferentially intercalates at the transcriptional start sites (TSS) of active genes[31, 32]. We used the *in vivo* Chem-seq method with biotinylated psoralen (bio-psoralen) (Fig. 3j) to explore this observation genome-wide in MM1.S cells. The results confirm that bio-psoralen preferentially binds to the TSS of active genes (Fig. 3k–m). Thus, Chem-seq can detect local enrichment of DNA intercalating agents throughout the human genome.

A broad range of drugs should generally be amenable to biotinylation and Chem-seq analysis. The design and synthesis of biotinylated probes can be informed by structural data from drug-target complexes; such X-ray structures allow the identification of suitable attachment positions that can be covalently linked to the biotin moiety and that remain freely accessible in the complex. Suitable attachment points could also be inferred from structure-activity relationship data derived from structurally related compounds. If such data are not available, several attachment sites can be selected for biotinylation, and the derived probes can be tested experimentally for their ability to retain binding to the target. Points of attachment can either be provided by functional groups already present in the drug molecule, or may be obtained through chemical modification of the compound structure, such as alkylation or addition of amide or ester linkages. Finally, as there is an expanding interest in

elucidating the mechanisms of action of many drugs, biotinylated versions of such compounds are increasingly becoming commercially available.

In summary, Chem-seq provides a method to identify the sites bound by small chemical molecules throughout the human genome. When combined with other global analysis methods such as ChIP-seq, Chem-seq provides a powerful approach to investigate the direct, genome-wide effects of therapeutic modalities. This ability to map the global interactions of a chemical entity with chromatin genome-wide should provide new insights into the mechanisms by which small molecules perturb gene expression programs.

Methods

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

Online Methods

Commercially available compounds

AT7519 and biotinylated Psoralen (EZ-Link Psoralen-PEG-Biotin) were obtained from Selleck Chemicals and Thermo Fisher Scientific, respectively.

Cell culture and treatment with unbiotinylated drugs

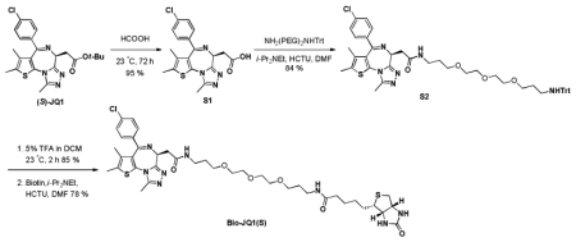
Multiple Myeloma MM1.S cells (CRL-2974, ATCC) were maintained in RPMI-1640 supplemented with 10% fetal bovine serum and 1% GlutaMAX (Invitrogen, 35050-061). For JQ1 treatment experiments, asynchronous cells were treated with varying concentrations of JQ1 or vehicle (DMSO) for 6. Alternatively, cells were treated with 2 μM AT7519 for 6h.

Genome-wide occupancy analysis of drug target proteins (ChIP-seq)

ChIP coupled with massively parallel DNA sequencing (ChIP-seq) was performed as previously described[33]. The following antibodies were used for Chromatin Immunoprecipitation (ChIP): anti-BRD4 (Bethyl Labs, A301-985A), anti-BRD2 (Cell signaling, 5848), anti-BRD3 (Bethyl Labs, A302-367A and A302-368A), anti-MED1 (Bethyl Labs, A300-793A), anti-H3K27Ac (Abcam, ab4729), anti-RNA-pol II (Santa Cruz, sc-899), anti-CTCF (Millipore, 07-729) and anti-CDK9 (sc-484). Illumina sequencing, library construction and ChIP-seq analysis methods were previously described[33].

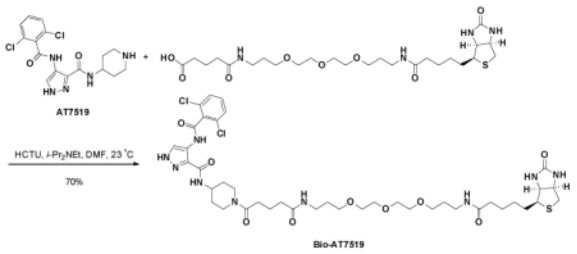
Synthesis of bio-JQ1(*S*)

The synthesis of active bio-JQ1(bio-JQ*S*) started with the (*S*)-JQ1, the active enantiomer that inhibits the bromodomain extra-terminal (BET) subfamily. As shown in the schematic below, removal of the *tert*-butyl group of the ester on (*S*)-JQ produced the acid S1. The resulting acid was then coupled with mono-protected $(PEG)_2$ linked diamine to give an amide S2. The protecting group on the terminal amine of compound S2 was removed under acidic condition to generate free amine, which was further coupled with biotin to afford the final active bio-JQ1(bio-JQ1S). The biotinylated inactive enantiomer bio-JQ1(R) was synthesized in the same synthetic route using inactive enantiomer (*R*)-JQ1.



Synthesis of bio-AT7519

As illustrated in the schematic below, biotinylated AT7519 (bio-AT7519) was directly synthesized from commercially available AT7519 (Selleck Chemicals) with Biotin-PEG2-COOH by amide coupling catalyzed by HCTU in DMF.



Expression of recombinant BRD4(1)

The first bromodomain of BRD4 was purified as a poly-histidine-tagged recombinant human protein expressed in E.coli, as previously described[24].

Differential scanning fluorimetry

Thermal melting experiments were carried out using a 7300 Real Time PCR machine (AB Applied Biosystems). Proteins were buffered in 10 mM HEPES pH 7.5, 500 mM NaCl and assayed in a 96-well plate at a final concentration of 1 μM in 20 μL volume. Compounds were added at a final concentration of 10 μM. SYPRO Orange (Molecular Probes) was added as a fluorescence probe at a dilution of 1:1000. Excitation and emission filters for the SYPRO-Orange dye were set to 465 nm and 590 nm, respectively. The temperature was raised with a step of 4 °C per minute from 25 °C to 96 °C and fluorescence readings were taken at each interval. The observed temperature shifts, ΔTm obs, were recorded as the difference between the transition midpoints of sample and reference (DMSO) wells containing protein without ligand in the same plate.

Isothermal titration calorimetry

ITC was performed using a ITC200 microcalorimeter from MicroCal™(Northampton, MA). All experiments were carried out at 25 °C while stirring at 1000 rpm, in ITC buffer (50 mM HEPES pH 7.4 at 25 °C, 150 mM NaCl). The microsyringe was loaded with a solution of the protein sample (190 μM, in ITC buffer). All titrations were conducted using an initial injection of 0.2 μl, followed by 19 identical injections of 2 μl with a duration of 5 sec (per injection) and a spacing of 90 sec between injections. The heat of dilution was determined by independent titrations (protein into buffer) and was subtracted from the experimental

data. The collected data were implicated in the MicroCal™ Origin software supplied with the instrument to yield enthalpies of binding (ΔH) and binding constants. A single binding site model was employed. Dissociation constants and thermodynamic parameters are presented in Supplementary Fig. 2c.

In vivo genome-wide occupancy analysis of biotinylated JQ1 (In vivo Chem-seq)

Exponentially growing MM1.S cells (2×10^8 cells per sample) were treated simultaneously with either 5 uM biotinylated JQ1 (Bio-JQ1) or DMSO (vehicle) and 1% Formaldehyde for 20 min in cell culture medium. Chemical crosslinking was terminated by addition of TRIS buffer, pH 7.5, to a final concentration of 300mM TRIS. Cells were harvested using a silicon scraper, centrifuged and the derived pellets washed three times with PBS. Cell nuclei were prepared as follows: cells were lysed in 50 mM HEPES, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100 plus protease inhibitor cocktail 'complete' (Roche), and cell nuclei were washed once with 10 mM Tris-HCL, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors. Nuclei were resuspended and sonicated in 50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS (sonication buffer) and protease inhibitor cocktail at 18 W for 10 cycles (30 s each) on ice with 30 s intervals between cycles. Sonicated lysates were cleared by centrifugation and incubated for 16 – 20 h at 4 °C with magnetic Streptavidin Dynabeads (MyOne Streptavidin T1, Invitrogen) (beads were blocked in PBS containing 0.5% BSA before this incubation step). Following incubation in nuclear sonicated lysate, beads were washed twice in sonication buffer, once in sonication buffer containing 500 mM NaCl, once in LiCl buffer (20 mM Tris-HCL, pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate), and once in 10 mM TRIS, pH 7.5, 0.1 mM EDTA. Bound protein-DNA complexes were subsequently eluted in 50 mM Tris-HCL, pH 8.0, 10 mM EDTA, 10% SDS at 65 °C for 15 min, and crosslinks were reversed by overnight incubation of the eluate at 65 °C. Contaminating RNA and protein were digested by addition of RNase and Proteinase K, respectively, and the DNA purified as previously described[34]. Finally, purified DNA fragments were massively parallel sequenced and the sequencing data analyzed as described[33].

In vitro genome-wide occupancy analysis of biotinylated JQ1 (In vitro Chem-seq)

Exponentially growing, untreated MM1.S cells were fixed with 1% Formaldehyde for 20 min in cell culture medium. Chemical crosslinking was terminated, cell nuclei prepared and sonicated nuclear lysate obtained as described above. Unlike in the in vivo protocol, however, Streptavidin Dynabeads were pre-incubated in PBS containing 0.5% BSA and either 200 μM biotinylated drug or vehicle (DMSO) for 6 h. Drug-bound beads were subsequently washed four times in PBS/0.5%BSA to remove unbound drug, and incubated in nuclear sonicated lysate for 16 – 20 h at 4 °C. All following steps are identical to those described above (in vivo Chem-seq method).

In vitro genome-wide occupancy analysis using biotinylated AT7519 (in vitro Chem-seq)

Exponentially growing, untreated MM1.S cells were fixed with 0.5% Formaldehyde for 5 min in cell culture medium. Chemical crosslinking was terminated by addition of TRIS buffer, pH 7.5, to a final concentration of 300mM TRIS. Cells were washed 3× in PBS and

cell nuclei prepared as follows: Cell nuclei were lysed in 50 mM HEPES, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100 plus protease inhibitor cocktail 'complete' (Roche), and cell nuclei were washed once with 10 mM Tris-HCL, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors. Nuclei were resuspended and sonicated in 50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.5% NP-40, 0.5% Triton-X (sonication buffer). Pellets were sonicated at 9-12 W for 4 cycles (30 s each) in a Misonix sonicator on ice with 1 min rest intervals between cycles. Drug-bound beads were added to the cleared sonicate and the precipitation allowed to proceed for 12-18 hours. Drug-bound beads were subsequently washed four times in sonication buffer, proteins eluted in 1% SDS, and crosslinks were reversed by overnight incubation of the eluate at 65 °C in 1% SDS. Contaminating RNA and protein were digested by sequential incubation with RNase A and Proteinase K, and the DNA purified as previously described[34]. Purified DNA fragments were subjected to massively parallel sequencing (Illumina) and the sequencing data analyzed as described[33].

In vitro genome-wide occupancy analysis of biotinylated psoralen by Chem-seq

Cell nuclei were prepared from exponentially growing MM.S cells using the Nuclei EZ prep kit (SIGMA). Nuclei were then resuspended in ice-cold PBS and directly incubated with 5 μM biotinylated psoralen or vehicle (DMSO) for 30 min at 4 °C. Nuclei were washed once in PBS and immediately irradiated at 360 nm for 30 min (Stratalinker) on ice. Nuclei were resuspended and sonicated in 50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS (sonication buffer) and protease inhibitor cocktail at 18 W for 10 cycles (30 s each) on ice with 30 s intervals between cycles. Sonicated lysates were cleared by centrifugation and incubated for 16 – 20 h at 4 °C with magnetic Streptavidin Dynabeads (MyOne Streptavidin T1, Invitrogen) (beads were blocked in PBS containing 0.5% BSA before this incubation step). Following incubation in nuclear sonicated lysate, beads were washed twice in sonication buffer, once in sonication buffer containing 500 mM NaCl, once in LiCl buffer (20 mM Tris-HCL, pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate), and once in 10 mM TRIS, pH 7.5, 0.1 mM EDTA. Bound protein-DNA complexes were subsequently eluted in 50 mM Tris-HCL, pH 8.0, 10 mM EDTA, 10% SDS and 10 mM Biotin, and the eluate incubate do/n at 65 °C. Contaminating RNA and protein were digested by addition of RNase and Proteinase K, respectively, and the DNA purified as previously described[34]. Finally, purified DNA samples were irradiated at 254 nm for 5 min (Stratalinker) to reverse psoralen-DNA crosslinks, followed by library preparation, massively parallel DNA sequencing and analysis of sequencing data[33].

Chem-seq and ChIP-seq data analysis

All ChIP-seq and Chem-seq datasets were aligned using Bowtie (version 0.12.2)[35] to build version NCBI36/HG18 of the human genome. We used the MACS version 1.4.1 (Model based analysis of ChIP-seq)[36] peak finding algorithm to identify regions of ChIP-seq enrichment over background. A p-value threshold of enrichment of 1e-9 was used for all datasets except for Chem-seq bioAT7519 (1e-6). To obtain the normalized read density of ChIP-seq datasets in any region, ChIP-seq reads aligning to each region were extended automatically by MACs, and the density of reads per basepair (bp) was calculated. The

density of reads in each region was normalized to the total number of million mapped reads producing read density in units of reads per million mapped reads per bp (rpm/bp).

Definition of transcribed genes

A gene was defined as transcribed if an enriched region for either H3K4me3 or RNA Pol II was located within ±5kb of the TSS. H3K4me3 is a histone modification associated with transcription initiation[37].

Definition of active enhancers

Active enhancers were defined as regions of enrichment for H3K27Ac outside of promoters (greater than 2.5kb away from any TSS). H3K27Ac is a histone modification associated with active enhancers[38, 39]. Active enhancers form loops with promoters that are facilitated by the Mediator complex[40]. Thus, we validated H3K27Ac definitions of enhancers using ChIP-Seq data for the mediator subunit Med1. Enriched regions from Med1 had >90% overlap with H3K27Ac regions in all datasets.

Determination of RNA Pol II traveling ratio

We determined the ratio of RNA Pol II ChIP-seq levels in initiating to elongating regions, a measure known as the traveling ratio (TR) (Fig. 3c and 3d)[41]. We defined the initiating region as +/-300bp around the TSS. We defined the elongating region as +300bp from the TSS to +3,000bp after the gene end. In order to make higher confidence comparisons, we limited our analysis to genes with detectable signal above noise in the initiating and elongating regions across all samples. The statistical significance of changes in the distribution of traveling ratios was determined using two-tailed *t* test.

Heatmap representation of read density profiles

The enriched regions, the merged regions, or the annotated transcription start sites of all refseq genes were aligned at the center in the composite view of signal density profile. The average ChIP-Seq or Chem-seq read density (rpm/bp) around ±5kb centered on the centers in 50 bp bin was calculated. The enriched regions of BRD2, BRD3, and BRD4 ChIP-seq and bio-JQ1 Chem-seq were merged together if overlapping by 1bp, resulting a total of 25,450 merged regions. For bio-psorlen Chem-seq analysis, the annotated transcription start sites of all refseq genes were used.

Pairwise comparison between Chem-seq and ChIP-seq

The set of genomic regions that were enriched for Chem-seq or ChIP-seq signal used in pairwise comparisons were merged together if overlapping by 1bp. The average ChIP-Seq or Chem-seq read density (RPM/PM) was calculated for each of the merged regions. The pair-wise comparisons by Pearson correlation were performed on all datasets using the average read density at the merged regions. The average linkage hierarchical clustering of the Pearson correlation was shown in the heatmap (Supplementary Fig. 2b, 5a, and 6c). For Fig. 2b, the enriched regions of 3 replicate JQ1 *in vitro* Chem-seq datasets, 3 replicate BRD2 ChIP-seq datasets, 3 replicate BRD3 ChIP-seq datasets, and 3 replicate BRD4 ChIP-seq datasets were merged together, resulting a total of 29,693 merged regions. For

Supplementary Fig. 5a, the enriched regions of bio-JQ1 (*in vitro* Chem-seq), BRD2, BRD3 from two different antibodies, and BRD4 (ChIP-seq) were merged together, resulting a total of 25,693 merged regions. For Supplementary Fig. 6c, the enriched regions of CDK7, CDK8, CDK8, BRD4, MED1, RNA pol II, H3K20me3, and H3K27me3 were merged together, resulting a total of 50638 merged regions.

Specificity analysis based on overlap between Chem-seq and ChIP-seq data

We analyzed Chem-seq and ChIP-seq data to identify genomic regions with substantial JQ1 Chem-seq signal but no significant BRD2, BRD3, or BRD4 ChIP-seq signal and vice versa (Fig. 2b). We adopted a generalized linear model (GLM method) to identify regions with differential signal between JQ1 and BET proteins[42, 43]. We first identified the set of genomic regions that were enriched for JQ1 Chem-seq or BET protein ChIP-seq signal in any one of the twelve datasets being considered (3 replicate JQ1 vitro Chem-seq datasets, 3 replicate BRD2 ChIP-seq datasets, 3 replicate BRD3 ChIP-seq datasets, and 3 replicate BRD4 ChIP-seq datasets). Regions from a dataset that overlapped with regions from another data set by 1bp were merged together to form a representative region that spans the combined genomic region. A total of 29,693 regions were identified. The read density in each region was calculated in units of reads per million mapped reads per bp (rpm/bp) for each dataset. The *edgeR* package was used to model technical variation due to noise among triplicate datasets and the biological variation due to differences in signal between JQ1 Chem-seq and BET protein ChIP-seq datasets[42]. Sequencing depth and upper-quantile techniques were used to normalize all twelve datasets together before common and tagwise dispersions were estimated. The statistical significance of differences between JQ1 Chem-seq signals and each of BET protein ChIP-seq signals was next calculated using an exact test and resulting P values were subjected to Benjamini–Hochberg multiple testing correction (FDR). For robustness, only regions where all triplicates showed significant enrichment of either ChIP-seq or Chem-seq signal (signal in units of rpm/bp above bottom 5 percentile of enriched regions) were used for differential signal analyses. This resulted in only 16266, 17556, and 17802 regions being further considered in the pair-wise comparisons of BRD2 vs. JQS, BRD3 vs. JQS, and BRD4 vs. JQS respectively. We only detect only 1 region showing substantial more JQ1 Chem-seq signal than BRD4 ChIP-seq signal. However, this region did have enrichment of BRD4 ChIP-seq. These data indicate that, under these experimental conditions, there is no identifiable off-target interaction that results in JQ1 Chem-seq signal at regions of the genome without BRD4.

Cyclin T1-CDK9 *in vitro* kinase assay

Life Technologies SelectScreen Profiling service was used to obtain IC50 values for inhibition by AT7519 versus biotinylated AT7519.

Composite biotinylated psoralen (bio-psoralen) occupancy profile

The transcription start sites of active and inactive genes were aligned at the center in the composite view of bio-psoralen enrichment. The average Chem-Seq read density of bio-psoralen and DMSO control around ±1kb centered on the TSS in 50bp bin was calculated in rpm/bp. The log2 ratio of mean signal of bio-psoralen and DMSO in 50bp bins is plotted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Tom Volkert, Jennifer Love, Jeong-Ah Kwon and Sumeet Gupta at the Whitehead Genome Technology Core for Solexa sequencing, Evan Cohick for help with cell culture, and Alexander Federation and Rudolph St. Pierre at the Dana-Farber Cancer Institute for biochemistry support. This work was supported by National Institutes of Health grants HG002668 (R.A.Y.) and CA146445 (R.A.Y., T.I.L.), Swedish Research Council Postdoctoral Fellowship VR-B0086301 (J.L.), American Cancer Society Postdoctoral Fellowship PF-11-042- 01-DMC (P.B.R.), the Leukemia & Lymphoma Society (J.Q. and J.E.B.) and the Damon-Runyon Cancer Research Foundation (J.E.B).

References

1. Bell O, Tiwari VK, Thoma NH, Schubeler D. Determinants and dynamics of genome accessibility. Nat Rev Genet. 2011; 12:554–564. [PubMed: 21747402]
2. Gilbert DM. Evaluating genome-scale approaches to eukaryotic DNA replication. Nat Rev Genet. 2010; 11:673–684. [PubMed: 20811343]
3. Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. Nat Rev Genet. 2010; 11:476–486. [PubMed: 20531367]
4. MacQuarrie KL, Fong AP, Morse RH, Tapscott SJ. Genome-wide transcription factor binding: beyond direct target regulation. Trends Genet. 2011; 27:141–148. [PubMed: 21295369]
5. Neph S, et al. Circuitry and dynamics of human transcription factor regulatory networks. Cell. 2012; 150:1274–1286. [PubMed: 22959076]
6. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet. 2011; 12:7–18. [PubMed: 21116306]
7. Ng HH, Surani MA. The transcriptional and signalling networks of pluripotency. Nat Cell Biol. 2011; 13:490–496. [PubMed: 21540844]
8. Northrup DL, Zhao K. Application of ChIP-Seq and related techniques to the study of immune function. Immunity. 2011; 34:830–842. [PubMed: 21703538]
9. Orkin SH, Hochedlinger K. Chromatin connections to pluripotency and cellular reprogramming. Cell. 2011; 145:835–850. [PubMed: 21663790]
10. Young RA. Control of the embryonic stem cell state. Cell. 2011; 144:940–954. [PubMed: 21414485]
11. Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov. 2012; 11:384–400. [PubMed: 22498752]
12. Copeland RA, Solomon ME, Richon VM. Protein methyltransferases as a target class for drug discovery. Nat Rev Drug Discov. 2009; 8:724–732. [PubMed: 19721445]
13. Dawson MA, et al. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. Nature. 2011; 478:529–533. [PubMed: 21964340]
14. Deshpande AJ, Bradner J, Armstrong SA. Chromatin modifications as therapeutic targets in MLL-rearranged leukemia. Trends Immunol. 2012; 33:563–570. [PubMed: 22867873]
15. Sandoval J, Esteller M. Cancer epigenomics: beyond genomics. Curr Opin Genet Dev. 2012; 22:50–55. [PubMed: 22402447]
16. You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell. 2012; 22:9–20. [PubMed: 22789535]
17. Raida M. Drug target deconvolution by chemical proteomics. Curr Opin Chem Biol. 2011; 15:570–575. [PubMed: 21763176]
18. Bantscheff M, et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. Nature biotechnology. 2011; 29:255–265.
19. Ito T, et al. Identification of a primary target of thalidomide teratogenicity. Science. 2010; 327:1345–1350. [PubMed: 20223979]

20. Taunton J, Hassig CA, Schreiber SL. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science. 1996; 272:408–411. [PubMed: 8602529]
21. Tung SY, Hong JY, Walz T, Moazed D, Liou GG. Chromatin affinity-precipitation using a small metabolic molecule: its application to analysis of O-acetyl-ADP-ribose. Cell Mol Life Sci. 2012; 69:641–650. [PubMed: 21796450]
22. Chung CW, et al. Discovery and characterization of small molecule inhibitors of the BET family bromodomains. J Med Chem. 2011; 54:3827–3838. [PubMed: 21568322]
23. Delmore JE, et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell. 2011; 146:904–917. [PubMed: 21889194]
24. Filippakopoulos P, et al. Selective inhibition of BET bromodomains. Nature. 2010; 468:1067–1073. [PubMed: 20871596]
25. Loven J, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013; 153:320–334. [PubMed: 23582323]
26. Nicodeme E, et al. Suppression of inflammation by a synthetic histone mimic. Nature. 2010; 468:1119–1123. [PubMed: 21068722]
27. Zuber J, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature. 2011; 478:524–528. [PubMed: 21814200]
28. Santo L, et al. AT7519, A novel small molecule multi-cyclin-dependent kinase inhibitor, induces apoptosis in multiple myeloma via GSK-3beta activation and RNA polymerase II inhibition. Oncogene. 2010; 29:2325–2336. [PubMed: 20101221]
29. Zhou Q, Li T, Price DH. RNA polymerase II elongation control. Annu Rev Biochem. 2012; 81:119–143. [PubMed: 22404626]
30. Squires MS, et al. AT7519, a cyclin-dependent kinase inhibitor, exerts its effects by transcriptional inhibition in leukemia cell lines and patient samples. Mol Cancer Ther. 2010; 9:920–928. [PubMed: 20354122]
31. Kouzine F, et al. Transcription-dependent dynamic supercoiling is a short-range genomic force. Nat Struct Mol Biol. 2013; 20:396–403. [PubMed: 23416947]
32. Naughton C, et al. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat Struct Mol Biol. 2013; 20:387–395. [PubMed: 23416946]
33. Marson A, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell. 2008; 134:521–533. [PubMed: 18692474]
34. Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc. 2006; 1:729–748. [PubMed: 17406303]
35. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]
36. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]
37. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007; 130:77–88. [PubMed: 17632057]
38. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. [PubMed: 21160473]
39. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:21931–21936. [PubMed: 21106759]
40. Kagey MH, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010; 467:430–435. [PubMed: 20720539]
41. Rahl PB, et al. c-Myc regulates transcriptional pause release. Cell. 2010; 141:432–445. [PubMed: 20434984]
42. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]

43. Ross-Innes CS, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012; 481:389–393. [PubMed: 22217937]

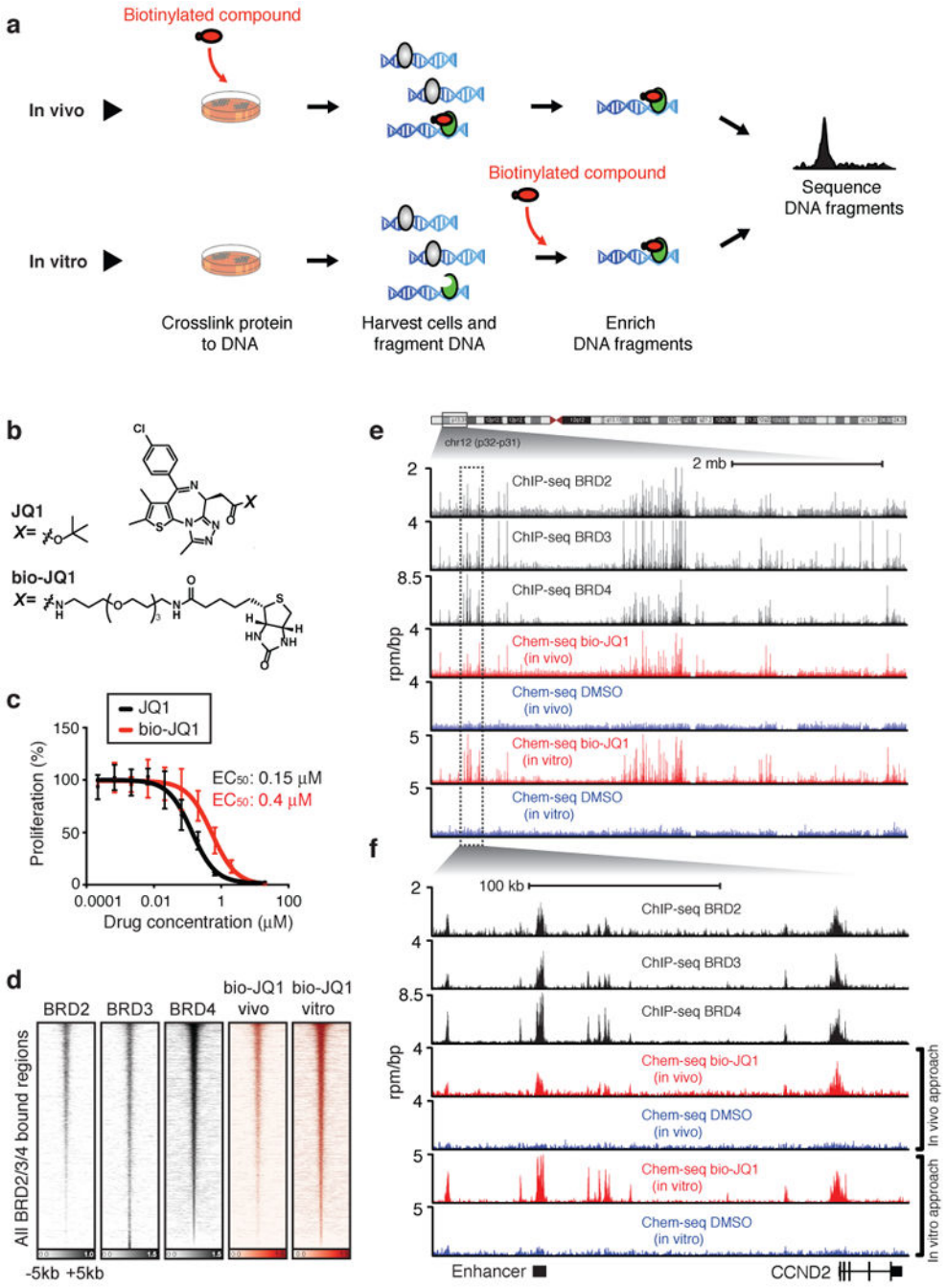


Figure 1.

Chem-seq from intact cells or cellular lysates reveals genomic sites bound by the BET bromodomain-targeting drug JQ1.

(a) Features of the Chem-seq method in living cells (*in vivo*, top) and cell lysates (*in vitro*, bottom). Top (*in vivo*): cells are treated with a biotinylated drug to allow drug-target binding to take place in the cellular context. Formaldehyde treatment cross-links chromatin-associated proteins to DNA, including drug-target complexes associated with chromatin. Following cell lysis and sonication, DNA fragments bound to the drug-target complex are

enriched using streptavidin beads. Sequencing of the enriched DNA fragments permits genome-wide identification of the loci to which the drug target binds. Bottom (*in vitro*): the biotinylated drug is added to the cell extract, where it binds protein-DNA complexes. Enrichment of DNA fragments and sequencing is carried out as in the *in vivo* method.
(b) Chemical structures of JQ1 and its biotinylated version, bio-JQ1.
(c) Effect of JQ1 (black) and bio-JQ1 (red) on MM1.S cell proliferation. Cells were treated with varying concentrations of drug for 72 h.
(d) Heatmap representation of binding of the individual BET proteins (ChIP-seq, black) and bio-JQ1 (*in vivo* and *in vitro* Chem-seq, red) to the union of all 25,450 regions occupied by BRD2, BRD3, BRD4 and bio-JQ1. Read density surrounds the center (±5kb) of all occupied regions, rank ordered from highest to lowest BRD4 occupancy.
(e) Gene tracks showing BRD2, 3, 4 and bio-JQ1 occupancy of a region of chromosome 12. ChIP-seq reads for BRD2, 3 and 4 (black), Chem-seq reads for biotinylated JQ1 (bio-JQ1, red) or DMSO vehicle control (blue) are shown. The genome-wide data is plotted in reads per million per base pair (rpm/bp).
(f) Close-up view of gene tracks showing BRD2, 3 and 4 occupancy (ChIP-seq) and bio-JQ1 occupancy (Chem-seq) across the CCND2 gene locus.

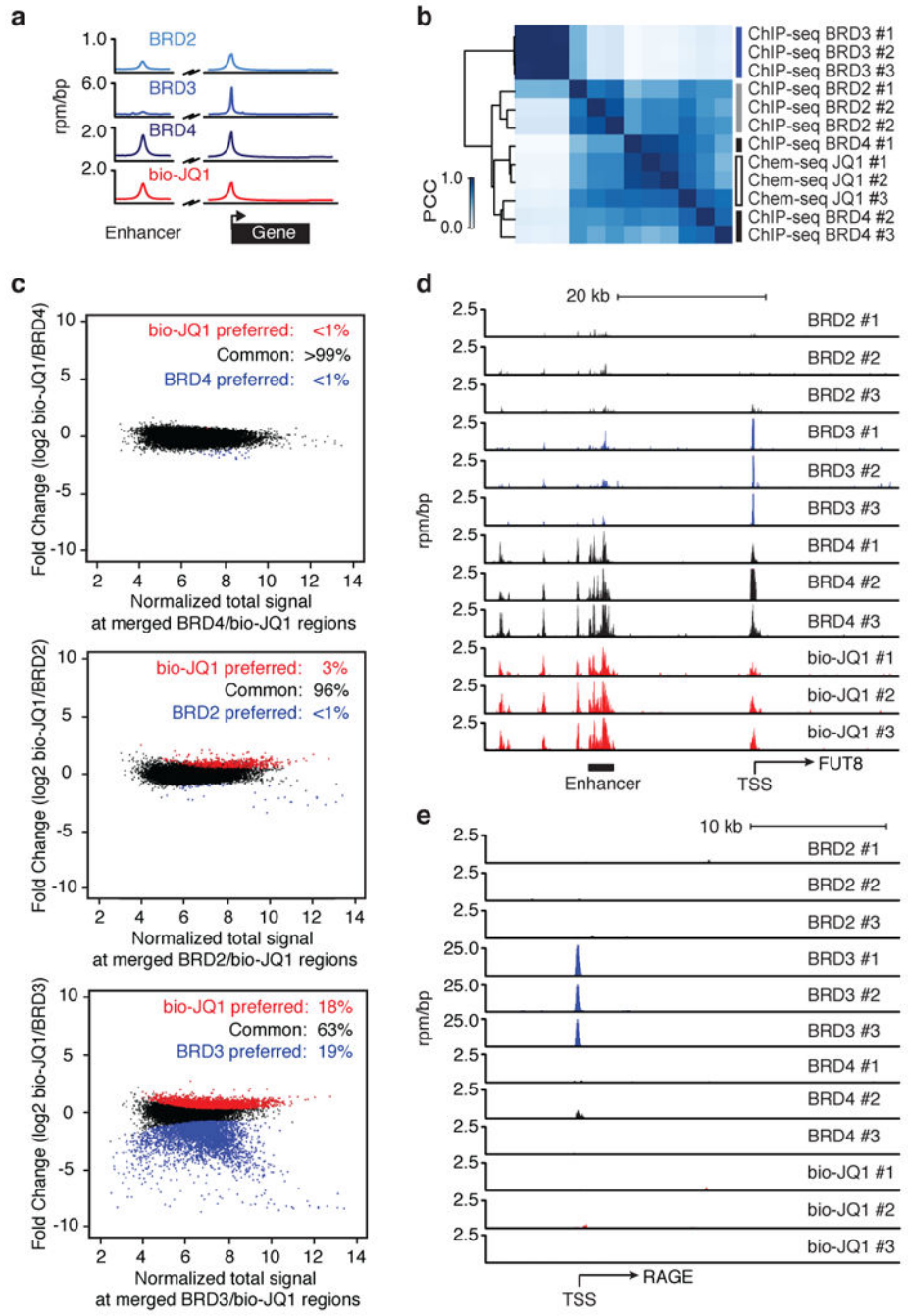


Figure 2.

Genome-wide drug target analysis.

(a) Genome-wide binding averages of BRD2, BRD3, BRD4 (ChIP-seq) and bio-JQ1 (*in vitro* Chem-seq) on active enhancers, active promoters and gene bodies in MM1.S cells.

(b) Heatmap showing the similarity of signal distribution between bio-JQ1 (*in vitro* Chem-seq) and BET bromodomain proteins BRD2, 3 and 4 (ChIP-seq) by Pearson correlation at 25,693 genomic regions bound by BET proteins and bio-JQ1. Blue reflects high similarity of signal between of each pair of factors. Factors are arranged and clustered along both axes

based on the distance calculated from Pearson correlation. ChIP-seq and Chem-seq data for each factor were generated from three independent experiments.

(c) Differential occupancy analysis of bioJQ1 (*in vitro* Chem-seq) and either BET bromodomain protein (ChIP-seq). The log ratios of normalized bio-JQ1 *in vitro* Chem-seq signal to BRD ChIP-seq signal are plotted for genomic regions identified as enriched for the presence of bio-JQ1 or either BRD protein. Triplicate ChIP-seq or Chem-seq datasets were used for each calculation.

(d) Gene tracks showing the FUT8 enhancer, a site identified as bio-JQ1 high/BRD3 low region identified in (b, lower panel, red). Triplicate datasets were generated for each factor.

(e) Gene tracks surrounding the transcriptional start site (TSS) of the PAGE gene, a ‘BRD3 preferred region’ (BRD3 high/bio-JQ1 low) identified in (b, lower panel, blue).

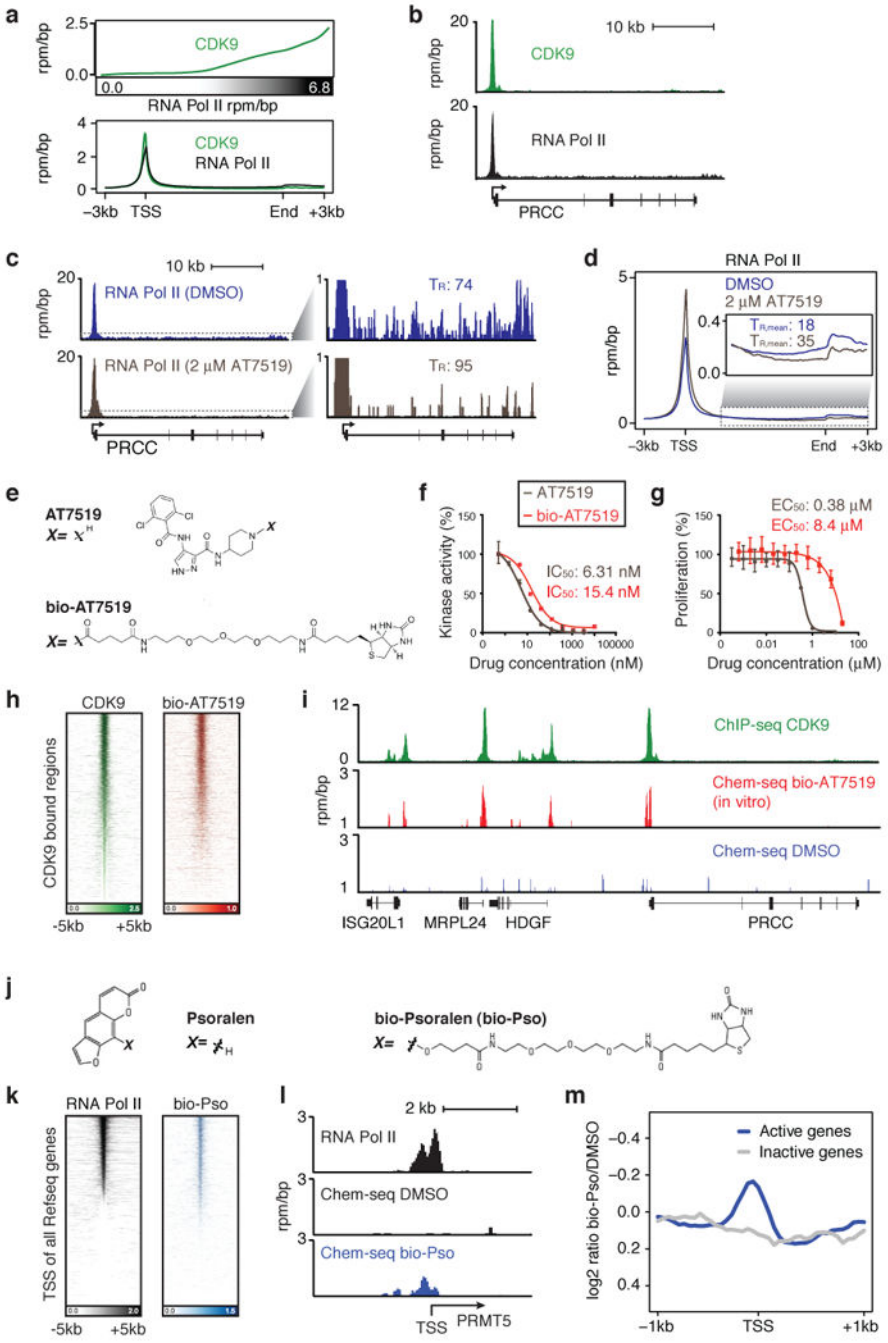


Figure 3.

Chem-seq reveals genomic occupancy of a protein kinase inhibitor and a DNA-intercalating drug.

(a) Upper panel: CDK9 occupancy is correlated with the RNA pol II at promoters in MM1.S cells. Median CDK9 signal at promoters is ranked by increasing RNA pol II occupancy. Signals are shown in units of reads per million mapped reads per base pair (rpm/bp). Promoters were binned (50/bin) and a smoothing function was applied to median signals.

Lower panel: genome-wide binding averages of CDK9 and RNA pol II on active promoters and gene bodies in MM1.S cells as determined by ChIP-seq analysis.

(b) Gene tracks showing occupancy of the PRCC gene by CDK9 and RNA pol II based on ChIP-seq data.

(c) Effect of AT7519 treatment on RNA pol II occupancy at the PRCC gene. MM1.S cells were treated with either DMSO vehicle (blue) or 2 μM AT7519 (brown) for 6 h, followed by RNA pol II ChIP-seq analysis. Twenty-fold magnifications of the rpm/bp scale of these gene tracks are shown in the right panel to show the difference in reads for elongating RNA pol II. T_R, RNA pol II traveling ratio.

(d) Genome-wide binding average RNA pol II (ChIP-seq) on active promoters and gene bodies following treatment of MM1.S cells with DMSO vehicle (blue) or 2 μM of AT7519 (brown) for 6 h. Magnification of the rpm/bp scale at gene bodies is shown in the inset. The inset includes RNA polymerase II traveling ratio distributions ($T_{R,\ mean}$) derived from MM1.S cells treated with DMSO (blue) or 2 μM AT7519 (red).

(e) Chemical structures of the pan-CDK inhibitor AT7519 and its biotinylated counterpart bio-AT7519.

(f) In vitro kinase assays with recombinant cyclin T-CDK9 complex in the presence of increasing concentrations of AT7519 or bio-AT7519. The derived IC_{50} values for each compound are shown.

(g) Effect of AT7519 and bio-AT7519 on MM1.S cell proliferation. Cells were treated with varying concentrations of drug for 72 h as indicated. The derived EC_{50} values for each compound are shown.

(h) Heatmap representation of CDK9 (ChIP-seq, green) and bio-AT7519 binding (*in vitro* Chem-seq, red) to all CDK9 occupied regions, rank ordered from highest to lowest CDK9 occupancy. Read density surrounds the center (± 5kb) of all occupied regions.

(i) Gene tracks showing occupancy of the PRCC gene locus by bio-AT7519 (red) and DMSO (vehicle, blue) as assessed by *in vitro* Chem-seq analysis, and by CDK9 (ChIP-seq, green).

(j) Chemical structures of psoralen and biotinylated psoralen.

(k) Heatmap representation of RNA pol II (ChIP-seq, black) and bio-psoralen binding (*in vivo* Chem-seq, light blue) to all human Refseq genes, rank ordered from highest to lowest RNA pol II occupancy. Read density surrounds the center (± 5kb) of occupied regions.

(l) Gene tracks centered at the TSS of the PRMT5 gene, showing occupancy of bio-psoralen (middle panel, light blue) versus DMSO (upper panel) as revealed by *in vivo* Chem-seq analysis, together with RNA pol II ChIP-seq data (lower panel, black).

(m) Metagene representation of bio-psoralen *in vivo* Chem-seq data at ±1kb around the TSS of active (light blue) and inactive (grey) genes. Log2 ratio of the mean bio-psoralen Chem-seq signal to mean DMSO signal in 50bp bins is plotted at the x-axis.