



# DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

## Semiconductor-based DNA sequencing of histone modification states

The Harvard community has made this article openly available.  
[Please share](#) how this access benefits you. Your story matters.

<b>Citation</b>	Cheng, C. S., K. Rai, M. Garber, A. Hollinger, D. Robbins, S. Anderson, A. Macbeth, et al. 2013. "Semiconductor-based DNA sequencing of histone modification states." Nature communications 4 (1): 2672. doi:10.1038/ncomms3672. <a href="http://dx.doi.org/10.1038/ncomms3672">http://dx.doi.org/10.1038/ncomms3672</a> .
<b>Published Version</b>	<a href="https://doi.org/10.1038/ncomms3672">doi:10.1038/ncomms3672</a>
<b>Accessed</b>	April 17, 2018 4:50:06 PM EDT
<b>Citable Link</b>	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:12152958">http://nrs.harvard.edu/urn-3:HUL.InstRepos:12152958</a>
<b>Terms of Use</b>	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a>

*(Article begins on next page)*

Published in final edited form as:

Nat Commun. 2013 ; 4: 2672. doi:10.1038/ncomms3672.

## Semiconductor-based DNA sequencing of histone modification states

Christine S. Cheng<sup>1</sup>, Kunal Rai<sup>2</sup>, Manuel Garber<sup>1,3</sup>, Andrew Hollinger<sup>1</sup>, Dana Robbins<sup>1</sup>, Scott Anderson<sup>1</sup>, Alyssa Macbeth<sup>1</sup>, Austin Tzou<sup>1</sup>, Mauricio O. Carneiro<sup>1</sup>, Raktima Raychowdhury<sup>1</sup>, Carsten Russ<sup>1</sup>, Nir Hacohen<sup>1,4,5</sup>, Jeffrey E. Gershenwald<sup>6</sup>, Niall Lennon<sup>1</sup>, Chad Nusbaum<sup>1</sup>, Lynda Chin<sup>2,†</sup>, Aviv Regev<sup>1,7,†</sup>, and Ido Amit<sup>1,8,†</sup>

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

<sup>2</sup>Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>3</sup>University of Massachusetts Medical School, Bioinformatics and Integrative Biology, Worcester, MA 01605, USA

<sup>4</sup>Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, Massachusetts, USA

<sup>5</sup>Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

<sup>6</sup>Department of Surgical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>7</sup>Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

<sup>8</sup>Weizmann Institute, Department of Immunology, Rehovot 76100, Israel

### Abstract

The recent development of a semiconductor-based, non-optical DNA sequencing technology promises scalable, low-cost and rapid sequence data production. The technology has previously been applied mainly to genomic sequencing and targeted re-sequencing. Here we demonstrate the utility of Ion Torrent semiconductor-based sequencing for sensitive, efficient and rapid chromatin immunoprecipitation followed by sequencing (ChIP-seq) through the application of sample preparation methods that are optimized for ChIP-seq on the Ion Torrent platform. We leverage this method for epigenetic profiling of tumour tissues.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>†</sup>To whom correspondence should be addressed: [aregev@broadinstitute.org](mailto:aregev@broadinstitute.org); [ido.amit@weizmann.ac.il](mailto:ido.amit@weizmann.ac.il); [lchin@mdanderson.org](mailto:lchin@mdanderson.org).

### AUTHOR CONTRIBUTIONS

C.S.C, A.R and I.A designed the study. C.S.C. performed the experiments and conducted data analysis. N.L., A.H., D.R., S.A., A.R. and A.T. developed and conducted the Ion Torrent ChIP-seq library construction experiments. C.S.C, A.R. and I.A. wrote the manuscript, with help from C.N., C.R. and N.H. A.R. guided data analysis. I.A. guided experimental designs. L.C. provided expertise and guidance in melanoma biology and analysis. K.R. provided fixed melanoma cell pellets from tumor derived cell lines and assisted in design of melanoma related experiments. M.G. provided analysis expertise and aided in the analysis of the data. M.O.C. carried out error rate analysis and helped with writing. R.R. carried out mouse dendritic cell culturing and fixing the cell pellets. J.E.G. and L.C. collected and provided tumor samples from melanoma patients. C.N., C.R. and N.H. helped editing the manuscript. All authors edited the manuscript.

### Statement of Competing Financial Interests

The authors declare no competing financial interests.

**Accession Codes:** Sequencing data have been uploaded to the NCBI GEO database under accession number GSE49477.

## Introduction

The semiconductor-based, non-optical sequencing technology used by the Ion Torrent sequencer (Life Technologies, Carlsbad, CA) has the potential for scalable, rapid and low-cost sequence data production<sup>1</sup>. Given the current industry standard for the density of transistors on the surface of a semiconductor, the technology has not yet reached its full possible capacity<sup>2</sup> and has the potential to provide comparable sequencing data yields to conventional optical based sequencers in a fraction of the time and cost<sup>3,4</sup>. The technology has recently been applied to genomic sequencing<sup>1</sup>, microbial genotyping<sup>5</sup> and targeted re-sequencing<sup>6</sup>.

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) is a powerful tool for characterizing the epigenetic landscape and transcriptional network in the context of both normal physiology and disease<sup>7-9</sup>. However, Ion Torrent sequencing has not yet been used for ChIP-Seq due to challenges in using ChIP DNA samples for sequencing library preparation. First, chromatin immunoprecipitation yields relatively low amounts of DNA, while commercial ChIP-Seq protocols recommend at least 500ng to 1ug of starting material for the library construction process. This is an issue in particular for ChIP DNA samples from immunoprecipitation of transcription factors or from limiting samples such as rare cell types or clinical samples, which are often at the few nanograms range. While recent studies presented ChIP-Seq protocols with low input (low cell number) for the Illumina platform<sup>10,11</sup>, such protocols are not yet available for the Ion Torrent platform. Second, the Ion Torrent process works optimally with a tight size range of DNA molecules of ~280bp +/- 20bp, whereas ChIP DNA typically spans a range of sizes from 200-600bp.

Here, we demonstrate the utility of Ion Torrent sequencing for ChIP-seq samples with sub-nanogram amounts of DNA. Furthermore, we apply the method to profile epigenetic marks of tumor tissues from melanoma patients and show its potential for analyzing tumor progression.

## Results

### Development of a ChIP-seq application for Ion Torrent

Our starting point was an automated 454-library construction method we previously developed<sup>12</sup>. To overcome the low input material obtained by ChIP, we devised a low input, scalable and robust library construction protocol for ChIP DNA that increases sensitivity and minimizes operator-dependent variability by incorporating a high yielding amplification enzyme (Kapa Biosystems, Woburn, MA), which has higher yield and higher genome coverage<sup>13</sup> than the Phusion polymerase that is commonly used in the standard Illumina protocol, low microliter volume reactions, molecularly barcoded oligonucleotide adapters, and automated fluid handling protocols (Supplementary Fig. S1 and **Methods**).

To address the wide size range of ChIP DNA, we first tested a standard enzymatic DNA shearing method that is routinely used with Ion Torrent genomic libraries, but failed to generate usable ChIP-Seq libraries. To overcome this problem, we started the library construction process without shearing and then used an automated gel size-selection system (Pippin Prep, Sage Science) to select appropriately sized library molecules post adapter ligation. We note that Illumina ChIP-seq libraries are usually not sheared, as the shearing step results in significant material loss, which is of particular concern with very low input samples, such as ChIP samples. Using this process, we successfully created libraries for 32 of 36 samples attempted (88.9% pass rate; success defined as having sufficient library

material to attempt at least three sequencing reactions. An Illumina ChIP-Seq library construction following a successful ChIP is closer to 100%).

To compare results between Ion Torrent sequencing and those from Illumina sequencing for ChIP applications, we performed ChIP with antibodies to the common histone mark, Histone 3 lysine 4 tri-methyl (H3K4me3), the C terminal domain of RNA polymerase II (Pol-II) and IgG (negative control) in mouse dendritic cells stimulated with lipopolysaccharide (LPS). The resulting immunoprecipitated DNA was used as input for both our Ion Torrent and standard Illumina library construction procedures. We sequenced the libraries on the Ion Torrent 316 sequencing chips (on average, 2 million reads/library, average read length: 180 bases) and with the gold standard ChIP-Seq data production using the Illumina Hi-Seq 2000 (15 million reads/library; read length: 40 base single end)<sup>7,14,15</sup>.

Illumina has a lower percentage of unmapped bases and a significantly higher rate of well-mapped bases than Ion Torrent (Supplementary Table S1). Although the Ion Torrent reads had higher error rates for both SNPs (10 fold higher) and indels (100 fold higher), these were still below 1 in 1,000 bases (Supplementary Table S1) and thus do not impact the quality of the chromatin maps.

We found excellent agreement between the two resulting maps. The ChIP-Seq enrichment scores, defined as the ratio of observed/expected number of reads at each peak region, are highly correlated between the two samples (H3K4me3: Pearson  $R=0.893$ , polII:  $R=0.722$ , Fig. 1a, b). Despite the differences in mapped bases and error rates, Ion Torrent sequencing produced comparable enrichment peaks to Illumina<sup>5,9,10</sup> sequencing for both H3K4me3 and polII ChIP-seq (Fig. 1a, b and Supplementary Fig. S2). Saturation analysis by sub-sampling the Ion Torrent reads indicates that the Ion Torrent library was sequenced to sufficient depth at 2 million reads (Fig. 1c). To examine the possibility that the longer read length in Ion Torrent sequencing reads contributes to this phenomenon, we extended the 40 base reads from Illumina to 180 bases. The extended Illumina reads (randomly down sampled from 15 million to 2 million reads) produced enrichment peaks that were more comparable to the 2 million Ion Torrent reads, indicating that longer read lengths may be beneficial in ChIP-seq applications independent of the sequencing platform used (Supplementary Fig. S3). Similarly, when we shortened the Ion Torrent reads from 180 bases to 40 bases and then performed alignment, the enrichment peaks were reduced, similar to Illumina reads down sampled to 2 million reads (40 bases) (Supplementary Fig. S3). As expected, we did not find enriched peaks with a negative control IgG ChIP-seq from Ion Torrent sequencing (Supplementary Fig. S4).

To determine the sensitivity of Ion Torrent sequencing for ChIP-seq, we tested a titration of ChIP DNA input amounts. We used 56ng, 4ng and 0.4ng of H3K4me3 ChIP DNA from a single ChIP experiment as input for our modified library construction protocol for Ion Torrent. We obtained successful libraries from all three aliquots with comparable results (Fig. 1d-f). For example, the lowest input library (0.4ng; equivalent to H3K4me3 immunoprecipitated DNA from 20,000 cells) was comparable to the highest input library (56 ng; equivalent to  $10 \times 10^6$  cells) based on enriched peaks (Fig. 1c) and correlation in enriched scores ( $R=0.753$ , Fig. 1f). Such low ChIP DNA input amounts are an order of magnitude lower than the Ion Torrent guidelines for library production and sequencing, and are comparable to recently developed protocols<sup>10,11</sup> for low input ChIP-seq with the Illumina platform. We also tested 0.05 ng of ChIP DNA, but failed to produce a high quality library. In summary, using our protocol we created successful libraries from very low starting amounts of ChIP DNA and obtained comparable results with Ion Torrent ChIP-seq to those with our standard methods, optical-based Illumina sequencing, while requiring an order of magnitude fewer sequencing reads (albeit longer reads).

## Epigenetic signatures during cancer progression

The rapid nature and relatively low-cost of Ion Torrent sequencing make it a promising candidate for diagnostic applications. As a proof of principal, we next performed H3K4me3 ChIP-seq experiments with a matched pair of primary tumor and metastasis cell lines derived from the same melanoma patient. Although the matched pair WM115 (primary tumor) and WM266-4 (metastasis) showed global correlated enrichment of H3K4me3 in gene promoters ( $r^2R=0.83$ , Fig. 2a and Supplementary Fig. S5), a large number of the genes demonstrated increased levels of H3K4me3 on their promoters in either the primary tumor (Fig. 2a, b) or metastasis (Fig. 2a, c). To test if any biological functions were associated with differential levels of H3K4me3, we performed Gene Set Enrichment Analysis (GSEA) with the gene sets in the Molecular Signatures Database (MSigDB)<sup>16</sup>.

Interestingly, we found that decreased H3K4me3 levels in metastasis are significantly associated with genes whose expression is repressed in embryonic stem cells, including targets of the Polycomb Repressive Complex (PRC) and for genes enriched for the H3K27me3 histone mark in embryonic stem cells<sup>17</sup> (Fig. 2d). Conversely, increased H3K4me3 levels in metastasis are significantly associated with interferon response and inflammatory response genes (Fig. 2e).

H3K4me3 is mostly enriched in promoters of actively transcribed genes, while H3K27me3 is usually enriched at repressive chromatin regions. In embryonic stem cells, however, many of the Polycomb targets, and especially key developmental regulators, are marked by both marks ('bi-valent domains')<sup>18</sup>. To explore the relationship between these two marks and metastasis, we performed H3K27me3 ChIP-seq experiments in the same cell lines. Indeed, increased H3K27me3 levels in metastasis correlate with loss of H3K4me3 and such genes are enriched for polycomb target genes in embryonic stem cells (Supplementary Fig. S6a). Interestingly, genes that have decreased H3K27me3 levels in metastasis are enriched for interferon response genes (Supplementary Fig. S6b, a few of the Interferon response gene sets are enriched in the top 150 gene sets).

To gain initial insight into the clinical applicability of the technology, we performed H3K4me3 ChIP-seq of seven metastatic tumor tissues from melanoma patients. Consistent with our observation in the progressive cell lines, genes that have increased H3K4me3 levels in the metastatic tumors are enriched for Interferon, inflammatory and immune response genes and genes that have decreased H3K4me3 levels are enriched for H3K27me3/polycomb target genes (Fig. 2f–g and Supplementary Fig. S7–10).

## Discussion

The repression of developmental gene sets in melanoma metastasis by an H3K27me3 gain and H3K4me3 loss is consistent with recent findings of similar embryonic stem cell signatures in aggressive tumors from other cancers<sup>19,20</sup>. Furthermore, several reports show that several histone modifying enzymes are misregulated in human cancers<sup>21</sup>. For example, EZH2, an H3K27me3 writer, is overexpressed in various solid tumors and its expression is correlated with tumor aggressiveness and metastatic progression<sup>22,23</sup>. Consistent with this, a stem cell Polycomb repression signature is also enriched in genes that gain H3K27me3 marks in metastatic prostate cancer<sup>24</sup>. Our finding that interferon and inflammatory response genes have higher levels of the H3K4me3 mark in melanoma metastasis is consistent with recent findings that link inflammation with cancer<sup>25–27</sup>.

In summary, we have demonstrated a rapid, sensitive, scalable and cost-effective semiconductor-based ChIP-seq pipeline for characterizing epigenetic signatures of metastatic human tumors from limiting samples with comparable sensitivity to recently

developed protocols<sup>10,11</sup> for ChIP-seq with low input using the Illumina platform. The technical and analytical methods for ChIP followed by Ion Torrent sequencing provide a platform for discovery and future diagnostic applications.

## METHODS

### Cell culture and human tissues

Mouse dendritic cells were isolated from wild type female 6–8 week old C57BL/6 mice obtained from the Jackson Laboratories and cultured in RPMI medium (Invitrogen) supplemented with 10% heat inactivated FBS (Invitrogen) and GM-CSF (20 ng/ml; Peprotech, Rocky Hill, NJ)<sup>28</sup>. Cells were cultured for 9 days and stimulated for 2 hours with LPS (100 ng/ml, rough, ultra-pure E. coli K12 strain, Invitrogen). Paired primary melanoma tumor derived cell line (WM115) and metastatic melanoma tumor derived cell line (WM266-4) from the same patient were obtained from the Wistar Institute (Philadelphia, PA). Metastatic melanoma tissues were collected from the Department of Surgical Oncology, University of Texas, MD Anderson Cancer Center with informed consent of the patients and prior MIT Committee On the Use of Humans as Experimental Subjects (COUHES) approval.

### Chromatin immunoprecipitation assay (ChIP)

Chromatin immunoprecipitation assays were performed using a previously published protocol with some minor modifications<sup>28</sup>. Briefly, cells were fixed for 10 min with 1% formaldehyde, quenched with glycine. Cells were lysed for 10 min on ice with RIPA lysis buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 140 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% DOC) (for mouse dendritic cells) or RIPA lysis buffer with 0.2% SDS and without Triton X-100 (for WM115 and WM266-4 melanoma cell lines) and then sonicated using the Branson sonicator as described in Garber et al<sup>28</sup>. Frozen human melanoma tissues (50–100mg) were thawed out on ice and chopped finely with a razor blade. Then tissues were fixed for 10 min with 1% formaldehyde in PBS and quenched with glycine. Fixed tissues were pulverized with Covaris CryoPrep CP02 at setting 5 for two times in TT1ET tissue tube. Cells were lysed for 10 min on ice with 1% SDS lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.1) and then sonicated using the Branson sonicator. Immunoprecipitation was performed by incubation of the sonicated cell lysate with 75  $\mu$ l of protein G magnetic dynabeads (Invitrogen) coupled to target antibody for over night at 4 degrees. Magnetic beads were then washed 5 times with cold RIPA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 140 mM NaCl, 1% Triton X-100, 0.1% SDS), twice with high salt RIPA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 500 mM NaCl, 1% Triton X-100, 0.1% SDS), twice with LiCl buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% Na-DOC), twice with TE buffer, then eluted in 50  $\mu$ l of elution buffer (10 mM Tris-HCl, pH 8.0, 5 mM EDTA, pH 8.0, 300 mM NaCl and 0.5% SDS). The eluate was reverse crosslinked at 65 degrees for 6 hours and then treated with 2  $\mu$ l of RNase A (Roche Applied Science) for 30 minutes and 2.5  $\mu$ l of Proteinase K (Invitrogen) for 2 hours. Finally, the decrosslinked DNA samples were cleaned up with 120  $\mu$ l of solid phase reversible immobilization (SPRI) beads and eluted in 50  $\mu$ l of EB buffer (10 mM Tris-HCl, pH 8.0).

### Ion Torrent library production and sequencing

ChIP-Seq libraries for Ion Torrent sequencing were created using a modified protocol of the manufacturer's instructions. The process for sample preparation is outlined in Supplementary Fig. S1. Briefly, 45  $\mu$ l of each ChIP DNA was added to a 96-well microtiter plate. Physical barcoding of all sample receptacles was employed at each step to ensure sample tracking integrity. Fragments were enzymatically end-repaired using an enzyme and

buffer cocktail (Kapa Biosystems, Woburn, MA). Following end-repair, automated reaction cleanup was performed using a solid-phase reversible immobilization (SPRI) process with a ratio of 1.8 times beads to DNA (Ampure, Agencourt, Beckman Coulter). Samples were eluted in 30ul and added to a 50ul adapter ligation reaction (Ligase enzyme and buffer from Kapa Biosystems), which also contained 5ul of Ion Torrent compatible oligonucleotide adapters (Integrated DNA Technologies). Adapters were used at 8uM concentration (1/5<sup>th</sup> of the standard amount) to minimize adapter dimer with the low input samples. A further 1.8X SPRI was performed post adapter-ligation. All automated fluid handling steps were carried out on a Bravo Automated Liquid Handling Platform, with a 96LT Disposable Tip pipette head (Agilent Technologies, Santa Clara, CA). After adapter ligation, 10uL of loading solution was added to each sample and each sample was size selected (280 base target size) using 2% gel cartridges (SAGE Pippen prep, SAGE Science, Beverly, MA). A further SPRI (2X Bead:DNA ratio) was performed post sizing and samples were eluted in 23ul. An amplification reaction was set up in a final volume of 50ul with the following cycling profile: 98°C for 45 seconds; 72°C for 20 minutes; followed by 12 cycles of 98°C for 15s, 63°C for 30s, 72°C for 30s; finally, 72°C for 60s. Amplification enzymes and master mix were from Kapa Biosystems (Woburn, MA), primers were those provided in the Ion Torrent template preparation kit. A SPRI cleanup with a 1.5X Bead:DNA ratio was performed post amplification and final libraries were eluted in 25ul. Libraries were quantified and checked for size on an Agilent Bioanalyzer (Agilent, Santa Clara, CA).

Template preparation was conducted using the Ion PGM™ 200 Xpress™ Template Kit, following the Ion PGM™ 200 Xpress™ Template Kit protocol (version 3). Libraries were diluted and 18ul of the  $1.55 \times 10^7$  molecules/ul dilution was added to an aqueous master mix containing polymerase and Ion Sphere Particles (ISPs) at the manufacturer's specified proportions. Emulsions were created using an IKA Ultra-Turrax Tube Drive (IKA, Wilmington, NC). After emulsion PCR, DNA positive ISPs were recovered and enriched according to standard protocols. A sequencing primer was annealed to DNA positive ISPs and the sequencing polymerase bound, prior to loading of ISPs into Ion 316 sequencing chips. Sequencing of the samples was conducted according to the Ion PGM™ 200 Sequencing Kit Protocol (version 6). One or more 316 sequencing chips were loaded and run on an Ion Personal Genome Machine for each sample. Each run was programmed to include 520 nucleotide flows to deliver 200 base read lengths, on average. Libraries were sequenced on Ion Torrent Personal Genome Machine. Basecalling and alignment were performed by the Torrent Suite 2.0.1 software.

### **Illumina library production and sequencing**

ChIP-seq library for Illumina sequencing were prepared using a previously published protocol<sup>28</sup>. Briefly, enzymes from New England Biolabs were used for the following library construction processes, DNA end-repair, A-base addition, adaptor ligation and Pfu Ultra II fusion enzyme (Agilent Technologies) was used for the enrichment step. Illumina ChIP libraries were barcoded and pooled as previously described<sup>28</sup>.

### **ChIP-seq analysis**

Reads were aligned to the reference mouse genome (mm9) or human genome (hg19) using the BWA aligner version 0.5.9. Sequencing metrics were extracted using the GATK tools<sup>29</sup> to traverse the genome and qualify mapped bases as well aligned - when the reads had mapping quality greater than Q20 (phred scaled) and high quality - when the reads had mapping quality greater than Q20 and the base had base quality greater than Q20. Error rates were measured in every 100 bases. Mismatches were counted for every base that mismatched the reference sequence in the alignment. Insertions and deletions were counted as events (not by the number of bases in the events). The rate is the number of insertion or

deletion events found in every 100 bases. Read length was calculated over all reads (including unmapped). Unmapped reads are all the reads that did not find a likely mapping in the mouse reference genome mm9 (this does not include mapping quality 0 reads). ChIP-seq peak calling was performed using the contiguous segmentation algorithm as part of the Scripture package<sup>28</sup> (<http://www.broadinstitute.org/software/scripture/>). Pearson correlation coefficients in Figure 1 were calculated by performing Pearson correlation analysis of pairwise comparison of ChIP-seq peak enrichment scores (log<sub>2</sub>) over a 500 bases sliding window.

In order to simulate for longer reads, we extended the 40 bases Illumina reads to 180 bases by setting the `–extFactor` flag to 140 in `igvtools` ([http://www.broadinstitute.org/software/igv/igvtools\\_commandline](http://www.broadinstitute.org/software/igv/igvtools_commandline)). To simulate for shorter Ion Torrent reads, we selected the first 40 bases from the ~180 bases of Ion Torrent reads. The sequencing data can be downloaded from the NCBI GEO database with the following accession number GSE49477.

### GSEA analysis

GSEA v2.2 was used to test for enrichment of each of the 3,398 gene sets in the Chemical and Genetic Perturbation (CGP) Collection of the Molecular Signature Database (MSigDB v3.1). Reads were first normalized based on the total number of reads for each sample. For each of the 20,000 NCBI human Refseq genes, the accumulated H3K4me3 ChIP-seq reads over 1kb upstream and 1kb downstream of transcription start site was calculated to present the H3K4me3 ChIP-seq signal for each gene. The residuals of the natural logarithms of the accumulated reads calculated from the linear model for each gene were used as the ranked list input for the GSEAPreranked function. Three biological repeats of H3K4me3 ChIP-seq Illumina sequencing data from normal skin melanocytes were downloaded from NCBI GEO database (GSE16368)<sup>30</sup>.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

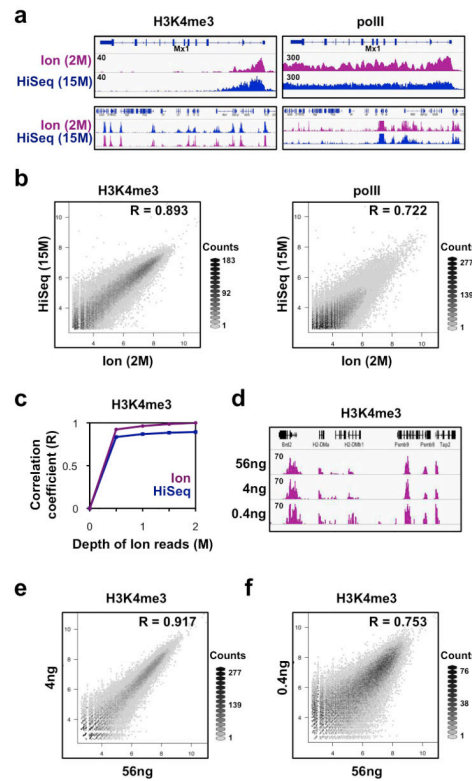
We thank R. Satija and S. Schwartz for helpful discussions. We thank J. Bochicchio for project management, A. Wysocki for the development of the data transfer pipeline and the Broad's Genomics Platform for sequencing. We also thank T.L. Calderone for preparing and sending the melanoma tumors. We thank L. Gaffney for production of the graphical abstract. This work was supported by the Human Frontiers Science Program Career Development Award (IA), the ISF Bikura Institutional Research Grant Program (IA), ERC starting grant 309788 (IA), NHGRI (1P50HG006193-01 to IA and AR), NIH Pioneer Award (DP1OD003958-01 to AR), Klarman Cell Observatory (AR), NCI (R01 CA93947 to LC), NCI (U01 CA1411508 to LC), NHGRI (U54HG003067 to CR), NIH Ruth L. Kirschstein National Research Service awards for individual Postdoctoral Fellowships (CC).

### References

1. Rothberg JM, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011; 475:348–52. [PubMed: 21776081]
2. Ferain I, Colinge CA, Colinge JP. Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors. *Nature*. 2011; 479:310–6. [PubMed: 22094690]
3. Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour*. 2011; 11:759–69. [PubMed: 21592312]
4. Merriman B, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*. 2012; 33:3397–417. [PubMed: 23208921]
5. Vogel U, et al. Ion torrent personal genome machine sequencing for genomic typing of *Neisseria meningitidis* for rapid determination of multiple layers of typing information. *Journal of clinical microbiology*. 2012; 50:1889–94. [PubMed: 22461678]

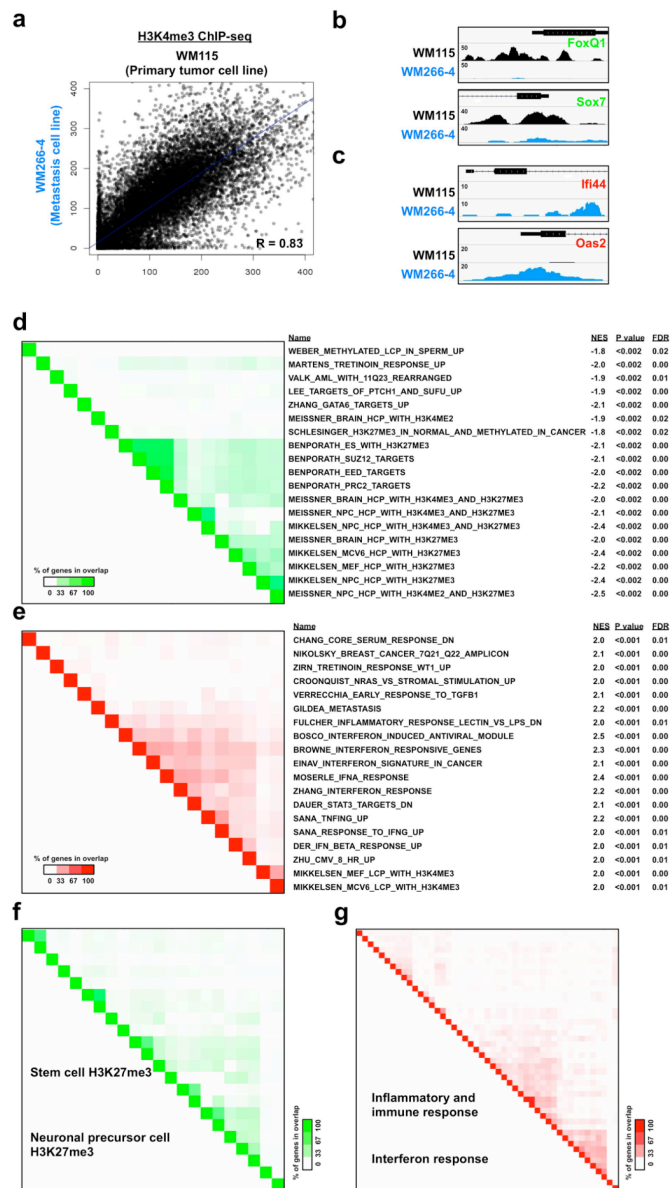


6. Elliott AM, Radecki J, Moghis B, Li X, Kammesheidt A. Rapid detection of the ACMG/ACOG-recommended 23 CFTR disease-causing mutations using ion torrent semiconductor sequencing. *Journal of biomolecular techniques: JBT*. 2012; 23:24–30. [PubMed: 22468138]
7. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011; 473:43–9. [PubMed: 21441907]
8. Gerstein MB, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature*. 2012; 489:91–100. [PubMed: 22955619]
9. Ram O, et al. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*. 2011; 147:1628–39. [PubMed: 22196736]
10. Adli M, Bernstein BE. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nature protocols*. 2011; 6:1656–68.
11. Adli M, Zhu J, Bernstein BE. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nature methods*. 2010; 7:615–8. [PubMed: 20622861]
12. Lennon NJ, et al. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome biology*. 2010; 11:R15. [PubMed: 20137071]
13. Quail MA, et al. Optimal enzymes for amplifying sequencing libraries. *Nature methods*. 2012; 9:10–1. [PubMed: 22205512]
14. Landt SG, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research*. 2012; 22:1813–31. [PubMed: 22955991]
15. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology*. 2011; 9:e1001046. [PubMed: 21526222]
16. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:15545–50. [PubMed: 16199517]
17. Lee TI, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*. 2006; 125:301–13. [PubMed: 16630818]
18. Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–26. [PubMed: 16630819]
19. Ben-Porath I, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature genetics*. 2008; 40:499–507. [PubMed: 18443585]
20. Kim J, et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell*. 2010; 143:313–24. [PubMed: 20946988]
21. Chi P, Allis CD, Wang GG. Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nature reviews. Cancer*. 2010; 10:457–69.
22. Min J, et al. An oncogene-tumor suppressor cascade drives metastatic prostate cancer by coordinately activating Ras and nuclear factor-kappaB. *Nature medicine*. 2010; 16:286–94.
23. Tsang DP, Cheng AS. Epigenetic regulation of signaling pathways in cancer: role of the histone methyltransferase EZH2. *Journal of gastroenterology and hepatology*. 2011; 26:19–27. [PubMed: 21175789]
24. Yu J, et al. A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer research*. 2007; 67:10657–63. [PubMed: 18006806]
25. Ben-Neriah Y, Karin M. Inflammation meets cancer, with NF-kappaB as the matchmaker. *Nature immunology*. 2011; 12:715–23. [PubMed: 21772280]
26. Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002; 420:860–7. [PubMed: 12490959]
27. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature*. 2008; 454:436–44. [PubMed: 18650914]
28. Garber M, et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Molecular cell*. 2012; 47:810–22. [PubMed: 22940246]
29. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*. 2011; 43:491–8. [PubMed: 21478889]
30. Bernstein BE, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology*. 2010; 28:1045–8.



**Figure 1. Ion Torrent and Illumina HiSeq ChIP-seq**

(a) Density profiles of H3K4me3 (left) and polII ChIP-seq (right) from 2 million Ion Torrent reads (purple) and 15 million Illumina HiSeq reads (blue) over the *Mx1* gene locus (top) and a larger region spanning the *TNF* locus (bottom). The scale of the density profile was set to the same level for each pair of samples, and is marked. All ChIP-seq experiments were performed with mDCs stimulated with LPS (0.1 ug/ml) for 2 hours. (b) A comparison of H3K4me3 (left) and polII (right) ChIP-seq peak enrichment scores ( $\log_2$ ) over a 500bp sliding window from 2 million Ion Torrent reads (X axis) and 15 million Illumina HiSeq reads (Y axis). The Pearson correlation coefficient is marked at the upper right corner. The grayscale represents a two-dimensional density plot, where the grid is divided into hexagonal bins and the color intensity reflects the density (counts) in the bin. An intensity bar of the counts is shown at the right bottom corner. (c) Saturation analysis to test for sufficient sequencing depth. Shown is the Pearson correlation coefficient (Y axis) calculated between decreasing sequencing reads randomly sub-sampled from the H3K4me3 ChIP-seq Ion Torrent reads (X axis) and either 15 million H3K4me3 ChIP-seq Illumina HiSeq reads (blue) or 2 million Ion Torrent reads (purple). (d) Density profiles of H3K4me3 ChIP-seq from 2 million Ion Torrent reads for libraries with 56ng, 4ng or 0.4ng of immunoprecipitated input DNA. (e, f) A comparison of H3K4me3 ChIP-seq peak enrichment scores ( $\log_2$ ) over a 500bp window from 2 million Ion Torrent reads with 56ng (Y axis) and 4ng (d, X axis) or 0.4ng (e, X axis) of immunoprecipitated input DNA. The Pearson correlation coefficient is marked at the upper right corner. The grayscale representation as in panel (b).



**Figure 2. Epigenetic signatures correspond to cancer progression**

(a) A comparison of accumulated H3K4me3 profiles in a pair of matching metastasis and primary tumor cell lines. Shown are the number of ChIP-seq reads over 1kb upstream and 1kb downstream of transcription start site of each of 20,000 human genes in a matched pair of primary melanoma tumor (WM115, black, X axis) and metastasis (WM266-4, blue, Y axis) cell lines derived from the same patient. The Pearson correlation coefficient is in the bottom right corner. (b, c) Density profiles of normalized reads from H3K4me3 ChIP-seq of primary melanoma tumor derived cell line (WM115) or metastasis derived cell line (WM266-4), at illustrative loci with either lower (b) or higher (c) levels of H3K4me3 in the metastasis derived cell line. (d–g) Gene sets enriched (by GSEA) among loci with different levels of histone modifications. In each case, shown are the names of the enriched gene sets (rows) along with the normalized enrichment score (NES), enrichment p value (P value) and False discovery rates (FDR). Each entry in the heat map indicates the percent overlap in ‘leading edge genes’ (those that contribute to the enrichment) between two enriched gene

sets. (d) Genes with lower H3K4me3 in metastasis cell line (WM266-4) relative to control primary tumor cell line (WM115). (e) Genes with higher H3K4me3 in metastasis cell line (WM266-4) relative to control primary tumor (WM115). (f) Genes with lower H3K4me3 in metastatic melanoma tumors from patients (an average of seven tumors samples) relative to control normal skin melanocytes (an average of three biological samples). (g) Genes with higher H3K4me3 in metastatic melanoma tumors from patients relative to control normal skin melanocytes.