Title:

ChIP-seq of plasma cell-free nucleosomes identifies cell-of-origin gene expression programs

Authors: Ronen Sadeh[1,2], Gavriel Fialkoff[1,2,#], Israa Sharkia[1,2,#], Ayelet Rahat[2], Mor Nitzan[1], Ilana Fox-Fisher[3], Daniel Neiman[3], Guy Meler[1,2], Zahala Kamari[1,2], Dayana Yaish[4], Samir Abu-Gazala[5], Tommy Kaplan[1], Ruth Shemer[3], David Planer[6], Aviad Zick[7], Eithan Galun[4], Benjamin Glaser[8], Yuval Dor[3], Nir Friedman [1,2,*]

Affiliations:

[1] School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel.

[2] Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel.

[3] Institute for Medical Research Israel-Canada, The Hebrew University-Hadassah Medical School, Jerusalem 9112001, Israel.

[4] The Goldyne Savad Institute for Gene Therapy, Hadassah-Hebrew University Medical Center, Jerusalem 9112001, Israel.

[5] Department of Surgery, Hadassah-Hebrew University Medical Center, Jerusalem 9112001, Israel.

[6] Department of Cardiology, Hadassah-Hebrew University Medical Center, Jerusalem 9112001, Israel.

[7] Department of Oncology, Hadassah-Hebrew University Medical Center, Jerusalem 9112001, Israel.

[8] Dept of Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical Center, Jerusalem 9112001, Israel.

[#] equal contribution

[*] contact author: nir.friedman@mail.huji.ac.il

Abstract:

Blood cell-free DNA (cfDNA) is derived from fragmented chromatin in dying cells. As such, it remains associated with histones that may retain the covalent modifications present in the cell of origin. Until now this rich epigenetic information carried by cell-free nucleosomes has not been explored at the genome level. Here, we perform ChIP-seq of cell free nucleosomes (cfChIP-seq) directly from human blood plasma to sequence DNA fragments from nucleosomes carrying specific chromatin marks. We assay a cohort of healthy subjects and patients and use cfChIP-seq to generate rich sequencing libraries from low volumes of blood. We find that cfChIP-seq of chromatin marks associated with active transcription recapitulates ChIP-seq profiles of the same marks in the tissue of origin, and reflects gene activity in these cells of origin. We demonstrate that cfChIP-seq detects changes in expression programs in patients with

1

heart and liver injury or cancer. cfChIP-seq opens a new window into normal and pathologic tissue dynamics with far-reaching implications for biology and medicine.

One Sentence Summary:

Chromatin immunoprecipitation and sequencing of histone modifications on blood-circulating nucleosomes (cfChIP-seq) provides detailed information about gene expression programs in different human organs.

Main Text:

Cell death throughout the human body results in the release of short nucleosome-size DNA fragments (cfDNA) into the circulatory blood system (Mandel and Metais 1948). The plasma of healthy individuals contains the equivalent of ~1000 genomes per ml, with a marked increase in the amount of cfDNA in many pathologies (e.g. cancer) (Lu and Liang 2016) and physiological conditions (e.g. exercise) (Haller et al. 2017). cfDNA fragments are short-lived with an estimated half-life of less than one hour (De Vlaminck et al. 2014; Jiang and Lo 2016; Schwarzenbach, Hoon, and Pantel 2011; Lo et al. 1999), making them ideal biomarkers for noninvasive monitoring of active physiological and pathological processes. Indeed, genetic variation in cfDNA is used to detect fetal chromosomal aberrations in maternal plasma, graft rejection, mutations, and for monitoring tumor dynamics (Schwarzenbach, Hoon, and Pantel 2011; Sun et al. 2015; Lu and Liang 2016; Wan et al. 2017). Importantly, cfDNA contains information beyond genetic variation. For example, the precise genomic location of specific cfDNA sequences reflects nucleosome positions in the source tissue and may thus suggest the cfDNA's cellular origins (Snyder et al. 2016); the underrepresentation of specific promoter sequences in cfDNA may reflect nucleosome-free regions associated with genes expressed in the source tissue (Ulz et al. 2016); and, cfDNA methylation patterns can be used to determine its tissue of origin (Guo et al. 2017; Zemmour et al. 2018; Lehmann-Werman et al. 2016, 2018; Kang et al. 2017; Li et al. 2018; Xu et al. 2017; Shen et al. 2018; Moss et al. 2018). Indeed, cfDNA methylation analysis demonstrated that most of the cfDNA pool of healthy individual originates from leukocytes, specifically neutrophils, and monocytes (Moss et al. 2018; Sun et al. 2015).

Genomic DNA, the source of cfDNA, is packaged into nucleosomes complexes made of 150bp DNA and histone proteins which are heavily post-translationally modified. Upon cell death, the genome is fragmented to nucleosome sized DNA fragments, most still associated with the histone proteins, and released to the blood as cell-free nucleosomes (cf-nucleosomes) (Holdenrieder et al. 2001, 2005; Rumore

2

and Steinman 1990), and there is evidence that these nucleosomes carry histone modifications (Gezer et al. 2013; Bauden et al. 2015; Deligezer et al. 2011).

In cells a plethora of histone modifications mark specific regions of the genome, such as enhancers and promoters, and their level correlates with gene activity (Gates, Foulds, and O'Malley 2017; Guenther et al. 2007; Barski et al. 2007; Roadmap Epigenomics Consortium et al. 2015). For example, tri-methylation of Lysine 4 on histone H3 (H3K4me3) mark active and paused promoters (Barth and Imhof 2010; Guenther et al. 2007). Additional histone modification mark accessible enhancers (H3K4me1/2) or elongation by RNA Pol II at gene bodies (H3K36me3).

We reasoned that if active chromatin marks are retained on circulating cf-nucleosomes, capturing and sequencing marked nucleosomes may inform on transcriptional activity within cells contributing to the cf-nucleosome pool (Figure 1A). Here, we develop a method to perform chromatin immunoprecipitation of modified nucleosomes directly from plasma followed by sequencing (cfChIP-seq). We show that cfChIP-seq can specifically capture nucleosomes with different active chromatin marks, and that these recapitulate the original genomic distribution of the modifications, and detect changes in gene expression programs in the cells of origin.

Results

ChIP-seq of cf-nucleosomes from plasma

We devised a simple protocol for cf-nucleosome ChIP-seq (cfChIP-seq) from 2ml of plasma from healthy subjects and <0.5ml from patients with increased levels of cfDNA (Figure 1B, Methods). Briefly, to overcome the extremely low concentration of cf-nucleosomes and the high concentration of native antibodies in plasma we incorporated two modifications to standard ChIP-seq. First, we covalently immobilized the ChIP antibodies to paramagnetic beads (Figure 1B), which can be incubated directly in plasma without interferences with native antibodies. Second, we minimize material loss by using an on bead ligation (Lara-Astiaso et al. 2014; Gutin et al. 2018; Singh et al. 2014; Rhee and Pugh 2011), where barcoded sequencing DNA adaptors are ligated directly to chromatin fragments prior to the isolation of DNA. The resulting protocol allows us to simply and efficiently enrich and sequence targeted chromatin fragments from low volumes of plasma.

We performed cfChIP-seq on multiple plasma samples from healthy individuals with antibodies targeting marks of accessible/active promoters (H3K4me3), enhancers (H3K4me2, or H3K4me1), and gene body of actively transcribed genes (H3K36me3) with reproducible yields (Figure 1C, Table S1). Several lines

3

of arguments demonstrate the specificity of cfChIP-seq: (a) cfChIP-seq signal is consistent with reference ChIP-seq against the same modification in tissues (Roadmap Epigenomics Consortium et al. 2015). This is seen in the remarkable agreement of peaks in genome browser (Figures 1C, S1A), in the average pattern around promoters and enhancers (Figures 1D, S1B), and in quantitative comparison of the signal across multiple genomic locations, such as all promoters, ($R > 0.8$ Figures 1E, S1C). Essentially all promoters that are ubiquitously marked (housekeeping) by H3K4me3 in reference ChIP-seq are significantly enriched for this mark in cfChIP-seq (9,795/10,505 promoters 93%, $p < 10^{-1000}$). Focusing on remaining marked promoters in cfChIP-seq, there is a significant overlap (1,324/2,311 promoters 57%, p $<10^{-288}$) with promoters from monocytes and neutrophils that are the major contributors to the cfDNA pool (Moss et al. 2018; Sun et al. 2015) (Figures 1F). (b) performing cfChIP-seq with a mock antibody resulted in dramatically lower yield, with no observable signal as seen for histone modifications (Table S1). (c) We estimated the rate of non-specific events in each sample (Methods), and used this background noise model to evaluate the expected amount of signal originating from non-specific source in each assay (Table S1). These results show that for H3K4me3 the levels of non-specific reads is comparable to or better than reference ChIP (Figure S1D) while for other antibodies such as H3K36me3 the performance is lower, but still informative (below).

A potential concern is contamination by chromatin released from in-tube lysis of white blood cells however, this is highly unlikely for several reasons. (a) Fragment size distributions of cfChIP-seq correspond to DNA wrapped around mono- and di-nucleosomes (Figure 1F), consistent with apoptotic or necrotic cell death, but not with cell lysis, which results in much larger fragments (Mizuta et al. 2013). (b) We identified hundreds of promoters carrying H3K4me3 that are absent in ChIP-seq from white blood cells (leukocytes, peripheral blood mononuclear cells; Figures 1F), these include promoters of genes that are expressed specifically in megakaryocytes, which reside in the bone marrow (Figures 1H and S1E). (c) We are able to detect disease-related chromatin from remote tissues from patients (below).

Together, these results strongly suggest that cf-nucleosomes preserve the endogenous patterns of active histone methylation marks in the cells of origin and can be assayed with cfChIP-seq.

cfChIP-seq of H3K4me3 correlates with gene expression

Having established that cfChIP-seq captures active histone modifications, we next asked whether these reflect gene expression patterns in cells. We decided to focus first on H3K4me3 for several reasons: (a) H3K4me3 ChIP signal is concentrated as narrow peaks at promoters of active and poised genes. This

allowed us to differentiate signal from noise. (b) H3K4me3 antibodies are well established in ChIP-seq assays in terms of specificity and sensitivity. (c) The localization of the mark at the promoter simplified connecting ChIP-seq signal and the expression of the associated genes. (d) The level of H3K4me3 at promoters from tissue samples correlates with the level of transcription (Karlić et al. 2010; Weiner et al. 2015; Liu et al. 2005; Pokholok et al. 2005) and is predictive of gene expression levels (Karlić et al. 2010; Weiner et al. 2015; Liu et al. 2005) (Figure S2A). Consistently, we find that H3K4me3 cfChIP-seq signal of healthy subjects is in agreement with gene expression levels in leukocytes (Figure 2A).

We next estimated the specific H3K4me3 signal at each promoter as the total number of cfChIP-seq reads mapped to it above the estimated level of background (Figure 2B, Methods, and Supplemental Note). Comparison of the specific signal at individual promoters shows a good agreement with expression levels of cells contributing to cfDNA (R = 0.560, Figure 2C), similarly to a comparison of ChIP-seq data and matched expression data from tissues (Roadmap Epigenomics Consortium et al. 2015) (0.600 < R < 0.675 Figure S2B), but not with gene expression levels of an irrelevant tissue (Figure S2C).

Together, these results strongly suggest that H3K4me3 cfChIP-seq signal is informative of gene-specific expression levels in tissue of origin.

cfChIP-seq detects expression changes

Can cfChIP-seq profiles capture changes in gene expression that reflect the underlying physiology? To better understand the variation of cfChIP-seq signal among subjects and in different physiological conditions, we performed H3K4me3 cfChIP-seq on samples from a diverse cohort of subjects (clinical details summarized in Table S2). These include: 15 healthy subjects (ages 23 -52); Four patients admitted to the emergency room with acute myocardial infarction (AMI); Nine patients with gastrointestinal (GI) tract adenocarcinoma; and two patients that underwent a partial hepatectomy (PHx). Some of these subjects were sampled multiple times at different intervals (e.g., before and after medical procedure). In some of these subjects there are expected changes in the cfDNA content: Cancer subjects are known to have large amount of tumor cfDNA (Swarup and Rajeswari 2007; Leon et al. 1977) ; AMI patients after percutaneous coronary intervention (PCI) have increased amount of cfDNA from cardiomyocytes (Zemmour et al. 2018). In other cases, such as hepatectomy, we expect damage to the tissue following the procedure.

To get a bird's eye view of the differences in cfChIP-seq signal among samples, we performed hierarchical clustering of 17,750 promoters of RefSeq genes that have a signal in at least one sample

5

(Methods). This clustering includes cfChIP-seq samples from our cohort and representative reference ChIP-seq (Roadmap Epigenomics Consortium et al. 2015) processed with the same pipeline (Methods). The clustering shows several trends (Figure 3A).

A large group of 9,376 genes show relatively small differences among samples. The genes in this cluster tend to be highly expressed, housekeeping genes with CpG-island at their promoters (Figure S3).

The remaining 8,374 genes display a rich tapestry of patterns. To investigated these patterns, we trimmed the cluster hierarchy at 35 clusters, and summarized their average profile across samples (Figure 3B). To provide orthogonal view, we also summarized the average expression levels of genes in each cluster in two expression databases (Methods): The GTEx compendium (GTEx Consortium 2015) contains samples from multiple tissues, and the BLUEPRINT Epigenome Project (Stunnenberg, International Human Epigenome Consortium, and Hirst 2016) focuses on specific hematopoietic cell types sorted from blood, cord blood, and bone marrow samples. Combining these patterns with enrichments analysis (Kuleshov et al. 2016) allowed us to assign putative names to clusters (Figure 3 and Table S3).

Several clusters show moderate to high signal in cfChIP-seq from healthy subjects (e.g clusters 16-26). Some of these have high signal in cells that contribute significantly to the cfDNA pool in healthy individuals such as neutrophils and monocytes (clusters 20 and 26) (Moss et al. 2018; Sun et al. 2015). Cluster 23 has strong cfChIP-seq signal and very low ChIP signal in all tissues. This cluster is enriched for genes such as GP6 and GP9 that are expressed in megakaryocytes and their protein products function in platelets, suggesting large contribution of bone marrow-residing megakaryocytes, or their immediate progenitors, to the cfDNA pool. We also identified clusters with strong patient-specific cfChIP-seq signals. These clusters are enriched for genes that are expressed in GI tract (clusters 1 and 4) with corresponding strong cfChIP-seq signal in GI cancers. Genes expressed in heart (clusters 3 and 5) with corresponding strong cfChIP-seq signal in AMI patients. And, genes expressed in liver (clusters 9 and 30) with corresponding strong cfChIP-seq signal in PHx and some AMI patients that also experienced liver injury (see below). In contrast, genes specific to T-lymphocytes that have long half life and hence are not major contributors to the cfDNA pool are found in cluster 8 that indeed show low cfChIP-seq signal.

Multiple sources for differences in cfChIP-seq signal

What drives the observed differences between cfDNA profiles? (Figure 3) cfChIP-seq measures the cf-nucleosome pool that results from combined contribution from multiple cells. Thus, we need to consider two scenarios for generating the observed differences (Figure 4A). (Scenario 1) Differences in

6

the proportion of cells that contributed to the cf-nucleosome pool. Such changes lead to coordinated changes in the cfChIP-seq signal of genes that are constitutive in the particular cell type (e.g., ALB, complement, and cytochrome-C genes in hepatocytes). (Scenario 2) Changes in the expression of a pathway (e.g., glycolysis) or a broader program (e.g., proliferation) in a subset of cells will also result in coordinated changes in the signal of multiple genes. More generally the changes we observe would be superimposition of both types of changes.

The observations above suggest that some of the differences between samples are due to differences in contributions from cell types. For example, cluster 30 is highly enriched for liver-expressed genes. However, other differences potentially reflect changes in transcriptional programs. For example, type I interferon response in cluster 31, and histone 1 gene cluster in cluster 18 represent programs that can be activated in various cell types.

cfChip detects cell-type expression programs

To better understand the contribution of cell types to the patterns of cfChIP-seq signal, we set out to define cell-type/tissue - specific signatures. Briefly, using the reference ChIP-seq compendium (Roadmap Epigenomics Consortium et al. 2015), we searched for genomic locations (e.g., promoters) that have high signal in the cell type in question and low, or non-existent, signal in all other cell types (Methods, Table S4). We then estimated the cumulative signal of each cell type-specific signature in each cfChIP-seq sample (Figure S4A, Methods). In healthy subjects most signal is from neutrophils and monocytes, and lower but significant signal from liver, in agreement with cfDNA methylation analysis (Moss et al. 2018) (Figure 4B). Testing significance against the null hypothesis of non-specific (background) signal, shows that liver-specific signature although much weaker than those of monocytes and neutrophils, is significantly higher than background. This is consistent with the estimates of 1-2% liver contribution to the cfDNA pool (Moss et al. 2018). In contrast, the heart-specific and brain-specific signatures are not observed in a significant manner above background (Figure 4B), consistent with cfDNA methylation analysis.

In patients we see contribution of additional cell-types/tissues. These changes are consistent with predictions. For example, we detected heart-specific signature (Figure S4B) in samples from AMI patient undergoing percutaneous coronary intervention (PCI) (Figure 4C). We see good agreement between the strength of the cfChIP-seq heart signature, the levels of troponin measured in the blood, and the estimate of heart cfDNA based on heart-specific differentially methylated CpGs (Zemmour et al. 2018) (Figure 4C). When examining the changes in heart signature from admittance to the emergency room to post-PCI checkup (Figure 4D), we see an increase in heart signature immediately following the procedure, as

7

previously reported by assaying cfDNA methylation (Zemmour et al. 2018).

Another example involves a patient undergoing partial hepatectomy (PHx) . We expected that during and following the surgical procedure we will see increase in liver cell death. Indeed, we observe dramatic changes during the operation which persist for a few days and slowly decay to original levels (Figure 4F). These changes are strikingly consistent with measurements of the classic marker for liver damage, the enzyme ALT. A noticeable difference is that the cfChIP-seq liver signature dropped back to normal levels about 2 days earlier than ALT, likely reflecting the shorter half life of cfDNA (<2 hours) compared to ALT (~47 hours) in the circulation (Giannini, Testa, and Savarino 2005).

For an unbiased view of contributions of different cell types to cfDNA, we evaluated our panel of cell-type specific signatures across cfChIP-seq samples (Figure 4G and Table S5). This analysis shows that in all samples we can detect signatures of a range of cell types from the blood (e.g., monocytes and neutrophils), and organs (e.g., liver). Of note, there is a decrease in blood cell types in samples with increased cfDNA load, consistent with smaller proportion of cfDNA from these cells. These samples include the the cancers C001 and C002 (cfDNA: 46.7, 84, 122ng/ml for C001.1, C002.1, and C002.2, respectively, compared to 4.5, and 11.75 ng/ml for healthy subjects H012.1, and H013.1, respectively) and AMI patient M004 (cfDNA: 21.13ng/ml and 35% of cfDNA originating from heart based on analysis of DNA methylation markers).

This unbiased approach reveals a more complex picture in AMI patients. In addition to the heart signature discussed above, in some AMI patients both before and shortly after PCI, we observe a significant increase in liver cell signature. This signature includes clear signal at liver-specific genes, such as Albumin and complement genes (Figure S4C). This increase is presumably a result of the well-known phenomenon of liver injury in AMI patients secondary to low organ perfusion and liver hypoxia (Ebert 2006). To confirm our cfChIP-seq observations we analysed the cfDNA methylation status for liver-specific DNA methylation regions indicative of liver cell death (Lehmann-Werman et al. 2018). Indeed, we observe excellent agreement between liver cfChIP-seq signature levels and liver cfDNA estimates (R^2=0.96, Figure 4H).

In samples from GI tract cancer patients we observe signal from tissues that are not observed in healthy subjects. Most evidently we observe signal originating from GI tissue and GI smooth muscle, which is in agreement with the primary sites of the tumors. A weaker but significant GI signature was evident even when the primary tumor was removed by surgery (patients C003-C007).

Together, these results demonstrate that differences in the tissues of origin contribute to the differences in

8

cfChIP-seq signal among subjects. In particular, in patients the differences correspond to the tissue where ongoing pathological processes take place, such as heart, liver, and gastrointestinal tissue.

cfChIP-seq signal reflects gene programs activity patterns

Since cfChIP-seq signal correlates with the gene expression programs in the cell of origin, we proceeded to inquire whether cfChIP-seq can reveal more dynamic transcriptional programs beyond the information on tissue of origin. To test this hypothesis, we evaluated the H3K4me3 cfChIP-seq signal in gene signatures representing different cellular processes, protein complexes and transcriptional responses based on gene expression studies (Liberzon et al. 2015; Drew et al. 2017; Giurgiu et al. 2019; Kamburov et al. 2013) (Methods, Table S6).

This analysis uncovered multiple signatures that differ from expected signal --- that is, the amount of H3K4me3 cfChIP-seq signal captured for a signature in a subject is significantly different from the observed signature in an averaged reference composed of a large cohort of healthy subjects (33 samples) (Methods, Table S7). For example, a strong signature of Heme Metabolism is observed in M002 who suffered from hypoxia (Figure 5A). The blood counts of M002 indeed show high RDW and low RBC and HGB, indicating higher production rate of red blood cells. During red blood cell production, erythroblasts lose their nucleus to become erythrocytes, presumably releasing their nucleosomes to the bloodstream (Lam et al. 2017; Moss et al. 2018). In C001 and C002 where the majority of the cfDNA originated from the tumors, we observe a sharp decrease in this signature consistent with overall reduction in the contribution of the non-tumor cells (Figure 4G). Thus, this signature is indicative of a specific hematopoietic cellular differentiation process.

Other signatures, such as glycolysis, unfolded protein response (UPR), and ribosomal protein genes reflect processes that can take place in multiple cell types, but are known to be upregulated in cancers (Hanahan and Weinberg 2011; Cubillos-Ruiz, Bettigole, and Glimcher 2017; Bhat et al. 2015). Indeed, we see upregulation of these pathways in cancer patients (Figure 5A). In C001 and C002 we see large increase in glycolysis signature, in agreement with the metabolic reprogramming, known as the Warburg effect, that is considered a hallmark of advanced cancers (Hanahan and Weinberg 2011). Interestingly, we observe increased UPR and ribosomal proteins signatures in C002 but not in C001, suggesting that these tumors are molecularly different. In other cancer samples, the amount of tumor derived cfDNA is probably too low to pass a significance test by our current analysis. This could be improved in the future by correction to the fraction of tumor derived cfDNA from the total cfDNA.

9

Another example is Interferon-alpha response that is normally induced due to the presence of pathogens such as viruses and bacteria. We observe dramatic increase in interferon signature in M004, who seems to undergo a more severe heart damage compared to other AMI patients in terms of troponin levels and cfChIP-seq heart markers (Figure 4C). Induction of interferon response was recently shown to promote a fatal response to AMI (King et al. 2017)

Looking at protein complexes signatures, we find a dramatic downregulation in the signature of the SWI/SNF (BAF) tumor suppressor, and chromatin remodeling complex in C001 and C002. Genes encoding for this complex are collectively mutated in ~20% of human cancers including GI-tract cancers (Kadoch et al. 2013). This decrease is SWI/SNF signature likely reflects tumor-specific transcriptional programs and not merely tissue of origin (i.e GI tract) since the signature is lower in liver tissue compared with GI tract tissues (Figure S4D), yet the significant reduction of the SWI/SNF signature is observed only in the cancer patients cfChIP-seq, but not in patients with liver injury.

Together, these observations demonstrate that cfChIP-seq reflects detailed changes in gene expression programs beyond cell type-specific programs.

cfChIP-seq allows to dissect patient-specific molecular phenotypes

A hallmark of cancer cells is genetic alterations that lead to dysregulated gene expression programs (Hanahan and Weinberg 2011). Identification of such cancer-specific transcriptional programs can assist diagnosis and treatment choice (Bradner, Hnisz, and Young 2017). We tested each sample for genes whose signal was elevated compared to "reference" healthy samples. As a control, unrelated healthy samples outside the reference set were in high correlation with healthy reference, with few genes (usually less than 100) showing significantly elevated signal (Figure 5B). In contrast, samples from patients revealed hundreds to thousands of genes with significantly elevated signal (Figure 5B, Table S8). Examining these genes for enrichment in annotated gene lists (Kuleshov et al. 2016) recapitulated some of the results discussed above (Figure 4F, Table S5). For example, genes with abnormally high H3K4me3 mark in C001 were enriched for gene sets of GI tract and Brain consistent with the pathology of this patient (Table S5).

To test for cancer-specific signatures in the H3K4me3 cfChIP-seq signal we analysed expression profiles from The Cancer Genome Atlas and GTEx projects (GTEx Consortium 2015; Cancer Genome Atlas Research Network et al. 2013). For each tumor type we identified a set of genes whose expression is significantly higher in the tumor compared to normal tissues (Methods, Table S9). We then tested for significant overlaps between the set of genes with higher H3K4me3 signal in a cfChIP-seq sample and the

10

set of genes over-expressed in a tumor type (Methods). For example, the set of genes with high signal in C002.1 has significant overlap ($q < 10^{-60}$) with GI-tract adenocarcinoma genes, but only negligible overlap with non GI cancer such as diffuse large b-cell lymphoma (DLBC) (Figure 5C). The analysis of all samples against all tumor types (Figure 5C and Table S10) shows that only samples from cancer patients have significant enrichment of tumor related gene expression, while healthy and MI patients do not. Importantly, the enrichment for cancers of the GI tract is in line with the diagnosed pathology.

Focusing on specific genes that are known to be upregulated in GI tract cancers (Nissan et al. 2012; Rodia et al. 2016; Wu et al. 2018) we observe a clear increase of the H3K4me3 cfChIP-seq signal in the patients compared to healthy reference (Figure 5D). Among these genes we find the carcinoma markers CEACAM5 and CEACAM6. The protein products of these genes are used in an antibody-based assay for clinical cancer diagnosis (Duffy 2001). A second colorectal cancer marker, the long non-coding RNA CCAT1 (colorectal cancer associated transcript 1)(Ozawa et al. 2017) shows strong signal in one of the cancer patients (C002) but not in healthy subjects. Another example is the non-coding antisense RNA EGFR-AS1 that mediates cancer addiction to EGFR and when highly expressed can render tumors insensitive to anti EGFR antibodies (Tan et al. 2017). While cfChIP-seq signal for EGFR is detected in all cancers, EGFR-AS1 is detected only in C002. This finding, which would not be detected by cfDNA mutation analysis, raises the exciting possibility that cfChIP-seq can be informative for treatment choice beyond genomic mutations.

Analysis of enhancer and gene body marks

Our analysis of the active promoter mark H3K4me3 provided rich information regarding transcriptional programs in tissue of origin. Can we gain information from additional chromatin marks that are associated with enhancer and gene activity? Mono and di-methylation of H3 lysine 4 (H3K4me1 and H3K4me2, respectively) are found in two types of genomic regions: 1) promoter flanking regions at the boundaries of regions marked with H3K4me3. 2) poised/active enhancers, where the H3K4me3 is barely detected (Visel et al. 2009). cfChIP-seq of these marks recapitulates the expected distribution (Figures 1C, 1D). Around active promoters, the H3K4me2 and H3K4me1 flank the main H3K4me3 peak and correlate with H3K4me3 (Figure S5A). Additionally, in non genic region adjacent to the IFNB1 locus, we clearly see H3K4me1 and H3K4me2 signal with little or no H3K4me3 coinciding with regions annotated as enhancers by ChromHMM (Roadmap Epigenomics Consortium et al. 2015) (Figure 6A). As expected, not all the blood enhancers in this region have cfChIP signal since not all blood cells contribute to the

11

cfDNA pool (e.g B cells).

We chose to focus on H3K4me2 cfChIP-seq for enhancer analysis since this antibody had low background and high reproducibility (Figure S5B). Comparing cfChIP-seq of the same blood sample with the H3K4me2 and H3K4me3 antibodies shows expected differences (Figures 1C, 1D, 6A). While we see H3K4me3 cfChIP-seq signal almost exclusively at promoters, a large fraction of the reads from H3K4me2 cfChIP-seq are mapped to putative enhancer regions based on ChromHMM (Figure S5C).

We next examined enhancers' tissue-specificity. We assigned each putative enhancer region to a cell type or combination of cell types. To ensure that we are not biased by TSS H3K4me2 signal or noise, we focused on enhancers of size larger than 600bp that are at least 5Kb from the nearest TSS and do not overlap a gene body. This smaller set of regions (48,525/2,345,831 regions) can be safely assumed to be enhancers. Using the predictions from Roadmap Epigenomics compendium, we assigned for each cell-type a set of distal enhancers that are annotated only in that cell-type. Examining the number of reads in these groups (Figure S5D) in healthy samples recapitulated the observations for H3K4me3 with high coverage for neutrophils and monocytes, less in liver and T-cells, and essentially none in heart and brain.

Comparing H3K4me2 signal in two samples from a colorectal cancer patient (C002.1 and C002.2) against healthy samples, we again recapitulate the observations made above with H3K4me3 signal - the cancer samples have lower neutrophils and monocytes signal, and higher signal in colon-specific enhancers, which are barely present in healthy samples (Figure 6B).

We next examined the coordination between promoter marking with H3K4me3 and enhancer marking with H3K4me2. As expected, we find examples of coordinated promoter:enhancer activation (Figure 6C, CDX1). The intestine-specific transcription factor CDX1 is expressed weakly in colon but not in leukocytes and its expression is dramatically increased in GI cancers. Indeed, we observe strong promoter H3K4me3 signal only in cancer patients but not in healthy subjects (Figure S5E). The strong H3K4me3 signal in C002 promoter is accompanied by a pronounced H3K4me2 signal over large GI-specific enhancer regions in the vicinity of the gene only in C002 (Figure 6C), suggesting that CDX1 is activated through these enhancers in colon tissue. We also find examples for enhancer swapping (Figure 6D). The transcription factor TCF3 has strong promoter H3K4me3 signal in every subject tested including C002 (Figure S5E), and is expressed in many tissues including leukocytes. However, the H3K4me2 signal in the vicinity of the gene is strikingly different, with only cancer samples showing clear H3K4me2 signal in regions that correspond to putative colon enhancers (Figure 6D, regions E2 and E4). Interestingly, we only observe cancer-specific increase in H3K4me2 when the signal correspond to putative fetal colon enhancers (compare E1 with E2, E4, and E5), consistent with de-represion of fetal oncogenes. Altogether,

12

these results suggest that CDX1 is activated through different enhancers in different tissues and these differences can be captured by cfChIP-seq and add important information on tissue of origin beyond the promoter signal.

Tri-methylation of H3 lysine 36 (H3K36me3) is found at the body of transcribed genes. Unlike H3K4me3, which marks transcription start sites at both poised and active genes, H3K36me3 requires active transcription elongation to be deposited, and is hence more indicative of gene activity (Guenther et al. 2007). Despite the high background of H3K36me3 in cfChIP (Figure S1D), we do observe the typical enrichment at gene bodies (Figures 1D and S6A) and the signal is in correlation to leukocyte H3K36me3 and RNA-seq (Figures S6B and S6C). Comparing the H3K36me3 signal from a healthy subject to that of a colorectal adenocarcinoma patient, we see 3580 genes that have H3K36me3 that are significantly increased by at least 4 fold in the cancer sample (Figure S6D).

Genes with H3K36me3 mark in this cancer sample can be assigned to three main classes. Class I includes 3404 genes that are marked by both marks in healthy and cancer samples (e.g., DHX9, Figure 6E). The other two classes involve H3K36me3 marks observed only in the cancer sample. 1290 Class II genes are marked with H3K4me3 in both healthy and cancer samples (e.g., SAP18 and SKA1, Figure 6E) and provide new information beyond H3K4me3. In contrast, 163 Class III genes are not marked with either signal in healthy samples (e.g., VWA2, Figure 6E).

Contrasting the set of highly expressed genes in colorectal adenocarcinoma (COAD) from analysis of the TCGA expression profiles (Methods), with these three classes, we observe that each of them captures different parts of these sets (Figure 6F). Specifically, there are 24 COAD genes that were not detected by H3K4me3 and are detected by H3K36me3. Moreover, there are 48 COAD genes (10 in Class II and 38 in Class III) for which the change in H3K4me3 signal is further corroborated by H3K36me3 signal.

Altogether, these results demonstrate how cfChIP-seq can probe the state of various genomic functionalities including promoters, enhancers, and gene bodies. Each is highly informative on transcriptional activity in cells of origin, but in a complementary manner, suggesting increased information that can be obtained by combining data from different histone marks.

Discussion

With advances in sequencing technologies there is a growing interest in using cfDNA as a non-invasive assay for monitoring human physiology. Here we introduce cfChIP-seq, a new method, to infer the transcriptional programs of dying cells from plasma cell-free nucleosomes. We have established

13

capability for genome-wide mapping of plasma cf-nucleosomes carrying histone marks associated with active transcription. We demonstrated that cfChIP-seq signal is informative about the cf-nucleosomes tissue of origin, and used this information to detect various pathophysiological states including heart and liver injury, and cancer. Notably, our results show that cfChIP-seq detects activation of genes that are normally not expressed in healthy subjects, allowing us to identify abnormal transcriptional processes. The assay requires a modest amount of plasma (2ml or less), low sequencing depth, is inexpensive, fully automatable, and provides a sensitive and robust signal. Moreover, the assay leaves most of the original sample intact, allowing reuse for multiple assays, such as genomic sequencing, methylation analysis, or cfChIP-seq with additional antibodies, which is important in situations where blood volume is a limiting factor.

Most current cfDNA-based methods rely on detecting genomic alterations in cfDNA to quantify the contribution of cfDNA from cells with altered genomic sequence, such as fetus, a transplant, or mutated genes in tumors (Schwarzenbach, Hoon, and Pantel 2011; Sun et al. 2015; Lu and Liang 2016; Wan et al. 2017). Thus, these methods are biased towards a set of pre-selected genes, and are blind to events that involve turnover and death of cells whose genome is identical to the host genome. More recent approaches leverage epigenetic information in cell free DNA. Extremely deep sequencing of total cfDNA to identify nucleosomes and transcription factors positions (Snyder et al. 2016) and occupancy (Ulz et al. 2016) reflect tissue of origin and gene expression. However, they rely on detecting changes in coverage over target regions, with a signal of source tissue imposed on the background of normal cells (e.g., detection of an event causing nucleosome depletion in 10% of the cells requires 90% occupancy to be distinguished from 100% occupancy). Thus, such methods avoid sampling noise by using extremely deep sequencing coverage (100s of million reads per sample). Even with such sequencing depth, there is a prohibitive harsh detection limit for events in rare subsets of cells (Snyder et al. 2016). A promising alternative is assaying DNA CpG methylation along the sequence to identify cell of origin (Guo et al. 2017; Zemmour et al. 2018; Lehmann-Werman et al. 2016, 2018; Kang et al. 2017; Li et al. 2018; Xu et al. 2017; Shen et al. 2018). DNA methylation serves as a stable epigenetic memory and is largely unchanged upon dynamic cellular responses. As such, it is highly informative regarding cell lineage, but much less about transient changes in expression. Moreover, unbiased analysis of DNA methylation requires high sequencing depth since most CpGs are methylated.

cfChIP-seq has the potential to circumvent some of these limitations. Targeted enrichment of active marks results in dramatic reduced representation of the genome such that fewer sequencing reads (~two orders of magnitude less) are required to obtain informative signal. Since we target marks associated with active

14

transcription, we are assaying a positive signal, where few reads are indicative of the presence of a particular cell type or expression program. This is in contrast to methods such as occupancy or DNA methylation that either measure negative signal (lack of nucleosome occupancy) or both negative and positive signals (e.g., %methylated).

Intensive research during the last two decades established the connection between specific histone marks and chromatin-templated processes including transcription, replication, and damage repair. Leveraging this rich and complex information to circulating cfDNA analysis has the potential to unravel physiological processes in remote organs, such as cell proliferation, hypoxia, inflammation, metabolic changes, and cancerous transformation, with minimal invasiveness. All of these processes involve activation of large transcriptional programs, which leave unique imprint on chromatin.

Assaying modified cf-nucleosomes, either used alone or in combination with existing biomarkers, has multiple potential medical applications, such as disease detection (e.g., detecting unknown tumors), improved diagnosis (e.g., replacing tissue biopsy with liquid biopsy), and non-invasive monitoring of disease progression and treatment efficacy. Moreover, the use of minimally invasive, easy to collect assay opens up a wide range of opportunities for studying basic questions in human physiology that have not been accessible until now.

References

Barski, Artem, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. 2007. "High-Resolution Profiling of Histone Methylations in the Human Genome." *Cell* 129 (4): 823–37.

Barth, Teresa K., and Axel Imhof. 2010. "Fast Signals and Slow Marks: The Dynamics of Histone Modifications." *Trends in Biochemical Sciences* 35 (11): 618–26.

Bauden, Monika, Dorian Pamart, Daniel Ansari, Marielle Herzog, Mark Eccleston, Jake Micallef, Bodil Andersson, and Roland Andersson. 2015. "Circulating Nucleosomes as Epigenetic Biomarkers in Pancreatic Cancer." *Clinical Epigenetics* 7 (October): 106.

Bhat, Mamatha, Nathaniel Robichaud, Laura Hulea, Nahum Sonenberg, Jerry Pelletier, and Ivan Topisirovic. 2015. "Targeting the Translation Machinery in Cancer." *Nature Reviews. Drug Discovery* 14 (4): 261–78.

Bradner, James E., Denes Hnisz, and Richard A. Young. 2017. "Transcriptional Addiction in Cancer." *Cell* 168 (4): 629–43.

Cancer Genome Atlas Research Network, John N. Weinstein, Eric A. Collisson, Gordon B. Mills, Kenna R. Mills Shaw, Brad A. Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M. Stuart. 2013. "The Cancer Genome Atlas Pan-Cancer Analysis Project." *Nature Genetics* 45 (10): 1113–20.

Cubillos-Ruiz, Juan R., Sarah E. Bettigole, and Laurie H. Glimcher. 2017. "Tumorigenic and Immunosuppressive Effects of Endoplasmic Reticulum Stress in Cancer." *Cell* 168 (4): 692–706.

15

Deligezer, Ugur, Elif Z. Akisik, Ebru E. Akisik, Müge Kovancilar, Dursun Bugra, Nilgün Erten, Stefan Holdenrieder, and Nejat Dalay. 2011. "H3K9me3/H4K20me3 Ratio in Circulating Nucleosomes as Potential Biomarker for Colorectal Cancer." In *Circulating Nucleic Acids in Plasma and Serum*, 97–103. Springer Netherlands.

De Vlaminck, Iwijn, Hannah A. Valantine, Thomas M. Snyder, Calvin Strehl, Garrett Cohen, Helen Luikart, Norma F. Neff, et al. 2014. "Circulating Cell-Free DNA Enables Noninvasive Diagnosis of Heart Transplant Rejection." *Science Translational Medicine* 6 (241): 241ra77.

Drew, Kevin, Chanjae Lee, Ryan L. Huizar, Fan Tu, Blake Borgeson, Claire D. McWhite, Yun Ma, John B. Wallingford, and Edward M. Marcotte. 2017. "Integration of over 9,000 Mass Spectrometry Experiments Builds a Global Map of Human Protein Complexes." *Molecular Systems Biology* 13 (6): 932.

Duffy, M. J. 2001. "Carcinoembryonic Antigen as a Marker for Colorectal Cancer: Is It Clinically Useful?" *Clinical Chemistry* 47 (4): 624–30.

Ebert, Ellen C. 2006. "Hypoxic Liver Injury." *Mayo Clinic Proceedings. Mayo Clinic* 81 (9): 1232–36.

Gates, Leah A., Charles E. Foulds, and Bert W. O'Malley. 2017. "Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle." *Trends in Biochemical Sciences* 42 (12): 977–89.

Gezer, Ugur, Duran Ustek, Ebru E. Yörüker, Aris Cakiris, Neslihan Abaci, Gloria Leszinski, Nejat Dalay, and Stefan Holdenrieder. 2013. "Characterization of H3K9me3- and H4K20me3-Associated Circulating Nucleosomal DNA by High-Throughput Sequencing in Colorectal Cancer." *Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine* 34 (1): 329–36.

Giannini, Edoardo G., Roberto Testa, and Vincenzo Savarino. 2005. "Liver Enzyme Alteration: A Guide for Clinicians." *CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne* 172 (3): 367–79.

Giurgiu, Madalina, Julian Reinhard, Barbara Brauner, Irmtraud Dunger-Kaltenbach, Gisela Fobo, Goar Frishman, Corinna Montrone, and Andreas Ruepp. 2019. "CORUM: The Comprehensive Resource of Mammalian Protein Complexes-2019." *Nucleic Acids Research* 47 (D1): D559–63.

GTEx Consortium. 2015. "Human Genomics. The Genotype-Tissue Expression (GTEx) Pilot Analysis: Multitissue Gene Regulation in Humans." *Science* 348 (6235): 648–60.

Guenther, Matthew G., Stuart S. Levine, Laurie A. Boyer, Rudolf Jaenisch, and Richard A. Young. 2007. "A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells." *Cell* 130 (1): 77–88.

Guo, Shicheng, Dinh Diep, Nongluk Plongthongkum, Ho-Lim Fung, Kang Zhang, and Kun Zhang. 2017. "Identification of Methylation Haplotype Blocks Aids in Deconvolution of Heterogeneous Tissue Samples and Tumor Tissue-of-Origin Mapping from Plasma DNA." *Nature Genetics* 49 (4): 635–42.

Gutin, Jenia, Ronen Sadeh, Nitzan Bodenheimer, Daphna Joseph-Strauss, Avital Klein-Brill, Adi Alajem, Oren Ram, and Nir Friedman. 2018. "Fine-Resolution Mapping of TF Binding and Chromatin Interactions." *Cell Reports* 22 (10): 2797–2807.

Haller, Nils, Suzan Tug, Sarah Breitbach, Arne Jörgensen, and Perikles Simon. 2017. "Increases in Circulating Cell-Free DNA During Aerobic Running Depend on Intensity and Duration." *International Journal of Sports Physiology and Performance* 12 (4): 455–62.

Hanahan, Douglas, and Robert A. Weinberg. 2011. "Hallmarks of Cancer: The next Generation." *Cell* 144 (5): 646–74.

Holdenrieder, Stefan, P. Stieber, H. Bodenmüller, M. Busch, G. Fertig, H. Fürst, A. Schalhorn, N. Schmeller, M. Untch, and D. Seidel. 2001. "Nucleosomes in Serum of Patients with Benign and Malignant Diseases." *International Journal of Cancer. Journal International Du Cancer* 95 (2): 114–20.

16

Holdenrieder, Stefan, Petra Stieber, Lisa Y. S. Chan, Sandra Geiger, Andreas Kremer, Dorothea Nagel, and Y. M. Dennis Lo. 2005. "Cell-Free DNA in Serum and Plasma: Comparison of ELISA and Quantitative PCR." *Clinical Chemistry* 51 (8): 1544–46.

Jiang, Peiyong, and Y. M. Dennis Lo. 2016. "The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics." *Trends in Genetics: TIG* 32 (6): 360–71.

Kadoch, Cigall, Diana C. Hargreaves, Courtney Hodges, Laura Elias, Lena Ho, Jeff Ranish, and Gerald R. Crabtree. 2013. "Proteomic and Bioinformatic Analysis of Mammalian SWI/SNF Complexes Identifies Extensive Roles in Human Malignancy." *Nature Genetics* 45 (6): 592–601.

Kamburov, Atanas, Ulrich Stelzl, Hans Lehrach, and Ralf Herwig. 2013. "The ConsensusPathDB Interaction Database: 2013 Update." *Nucleic Acids Research* 41 (Database issue): D793–800.

Kang, Shuli, Qingjiao Li, Quan Chen, Yonggang Zhou, Stacy Park, Gina Lee, Brandon Grimes, et al. 2017. "CancerLocator: Non-Invasive Cancer Diagnosis and Tissue-of-Origin Prediction Using Methylation Profiles of Cell-Free DNA." *Genome Biology* 18 (1): 53.

Karlić, Rosa, Ho-Ryun Chung, Julia Lasserre, Kristian Vlahovicek, and Martin Vingron. 2010. "Histone Modification Levels Are Predictive for Gene Expression." *Proceedings of the National Academy of Sciences of the United States of America* 107 (7): 2926–31.

King, Kevin R., Aaron D. Aguirre, Yu-Xiang Ye, Yuan Sun, Jason D. Roh, Richard P. Ng Jr, Rainer H. Kohler, et al. 2017. "IRF3 and Type I Interferons Fuel a Fatal Response to Myocardial Infarction." *Nature Medicine* 23 (12): 1481–87.

Kuleshov, Maxim V., Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, et al. 2016. "Enrichr: A Comprehensive Gene Set Enrichment Analysis Web Server 2016 Update." *Nucleic Acids Research* 44 (W1): W90–97.

Lam, W. K. Jacky, Wanxia Gai, Kun Sun, Raymond S. M. Wong, Rebecca W. Y. Chan, Peiyong Jiang, Natalie P. H. Chan, et al. 2017. "DNA of Erythroid Origin Is Present in Human Plasma and Informs the Types of Anemia." *Clinical Chemistry* 63 (10): 1614–23.

Lara-Astiaso, David, Assaf Weiner, Erika Lorenzo-Vivas, Irina Zaretsky, Diego Adhemar Jaitin, Eyal David, Hadas Keren-Shaul, et al. 2014. "Immunogenetics. Chromatin State Dynamics during Blood Formation." *Science* 345 (6199): 943–49.

Lehmann-Werman, Roni, Judith Magenheim, Joshua Moss, Daniel Neiman, Ofri Abraham, Sheina Piyanzin, Hai Zemmour, et al. 2018. "Monitoring Liver Damage Using Hepatocyte-Specific Methylation Markers in Cell-Free Circulating DNA." *JCI Insight* 3 (12). https://doi.org/10.1172/jci.insight.120687.

Lehmann-Werman, Roni, Daniel Neiman, Hai Zemmour, Joshua Moss, Judith Magenheim, Adi Vaknin-Dembinsky, Sten Rubertsson, et al. 2016. "Identification of Tissue-Specific Cell Death Using Methylation Patterns of Circulating DNA." *Proceedings of the National Academy of Sciences of the United States of America* 113 (13): E1826–34.

Leon, S. A., B. Shapiro, D. M. Sklaroff, and M. J. Yaros. 1977. "Free DNA in the Serum of Cancer Patients and the Effect of Therapy." *Cancer Research* 37 (3): 646–50.

Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. 2015. "The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection." *Cell Systems* 1 (6): 417–25.

Liu, Chih Long, Tommy Kaplan, Minkyu Kim, Stephen Buratowski, Stuart L. Schreiber, Nir Friedman, and Oliver J. Rando. 2005. "Single-Nucleosome Mapping of Histone Modifications in S. Cerevisiae." *PLoS Biology* 3 (10): e328.

Li, Wenyuan, Qingjiao Li, Shuli Kang, Mary Same, Yonggang Zhou, Carol Sun, Chun-Chi Liu, et al. 2018. "CancerDetector: Ultrasensitive and Non-Invasive Cancer Detection at the Resolution of Individual Reads Using Cell-Free DNA Methylation Sequencing Data." *Nucleic Acids Research* 46 (15): e89.

17

Lo, Y. M., J. Zhang, T. N. Leung, T. K. Lau, A. M. Chang, and N. M. Hjelm. 1999. "Rapid Clearance of Fetal DNA from Maternal Plasma." *American Journal of Human Genetics* 64 (1): 218–24.

Lu, Jun-Liang, and Zhi-Yong Liang. 2016. "Circulating Free DNA in the Era of Precision Oncology: Pre- and Post-Analytical Concerns." *Chronic Diseases and Translational Medicine* 2 (4): 223–30.

Mandel, P., and P. Metais. 1948. "Les Acides Nucleiques Du Plasma Sanguin Chez 1 Homme." *CR Seances Soc Biol Fil* 142: 241–43.

Mizuta, Ryushin, Shinsuke Araki, Makoto Furukawa, Yuki Furukawa, Syota Ebara, Daisuke Shiokawa, Katsuhiko Hayashi, Sei-Ichi Tanuma, and Daisuke Kitamura. 2013. "DNase γ Is the Effector Endonuclease for Internucleosomal DNA Fragmentation in Necrosis." *PloS One* 8 (12): e80223.

Moss, Joshua, Judith Magenheim, Daniel Neiman, Hai Zemmour, Netanel Loyfer, Amit Korach, Yaacov Samet, et al. 2018. "Comprehensive Human Cell-Type Methylation Atlas Reveals Origins of Circulating Cell-Free DNA in Health and Disease." *Nature Communications*, October, 448142.

Nissan, Aviram, Alexander Stojadinovic, Stella Mitrani-Rosenbaum, David Halle, Ronit Grinbaum, Marina Roistacher, Andrea Bochem, et al. 2012. "Colon Cancer Associated Transcript-1: A Novel RNA Expressed in Malignant and Pre-Malignant Human Tissues." *International Journal of Cancer. Journal International Du Cancer* 130 (7): 1598–1606.

Ozawa, T., T. Matsuyama, Y. Toiyama, N. Takahashi, T. Ishikawa, H. Uetake, Y. Yamada, M. Kusunoki, G. Calin, and A. Goel. 2017. "CCAT1 and CCAT2 Long Noncoding RNAs, Located within the 8q.24.21 'Gene Desert', Serve as Important Prognostic Biomarkers in Colorectal Cancer." *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 28 (8): 1882–88.

Pokholok, Dmitry K., Christopher T. Harbison, Stuart Levine, Megan Cole, Nancy M. Hannett, Tong Ihn Lee, George W. Bell, et al. 2005. "Genome-Wide Map of Nucleosome Acetylation and Methylation in Yeast." *Cell* 122 (4): 517–27.

Rhee, Ho Sung, and B. Franklin Pugh. 2011. "Comprehensive Genome-Wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution." *Cell* 147 (6): 1408–19.

Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, et al. 2015. "Integrative Analysis of 111 Reference Human Epigenomes." *Nature* 518 (7539): 317–30.

Rodia, Maria Teresa, Giampaolo Ugolini, Gabriella Mattei, Isacco Montroni, Davide Zattoni, Federico Ghignone, Giacomo Veronese, et al. 2016. "Systematic Large-Scale Meta-Analysis Identifies a Panel of Two mRNAs as Blood Biomarkers for Colorectal Cancer Detection." *Oncotarget* 7 (21): 30295–306.

Rumore, P. M., and C. R. Steinman. 1990. "Endogenous Circulating DNA in Systemic Lupus Erythematosus. Occurrence as Multimeric Complexes Bound to Histone." *The Journal of Clinical Investigation* 86 (1): 69–74.

Schwarzenbach, Heidi, Dave S. B. Hoon, and Klaus Pantel. 2011. "Cell-Free Nucleic Acids as Biomarkers in Cancer Patients." *Nature Reviews. Cancer* 11 (6): 426–37.

Shen, Shu Yi, Rajat Singhania, Gordon Fehringer, Ankur Chakravarthy, Michael H. A. Roehrl, Dianne Chadwick, Philip C. Zuzarte, et al. 2018. "Sensitive Tumour Detection and Classification Using Plasma Cell-Free DNA Methylomes." *Nature* 563 (7732): 579–83.

Singh, Shivani S., Navjot Singh, Richard P. Bonocora, Devon M. Fitzgerald, Joseph T. Wade, and David C. Grainger. 2014. "Widespread Suppression of Intragenic Transcription Initiation by H-NS." *Genes & Development* 28 (3): 214–19.

Snyder, Matthew W., Martin Kircher, Andrew J. Hill, Riza M. Daza, and Jay Shendure. 2016. "Cell-Free DNA Comprises an In Vivo Nucleosome Footprint That Informs Its Tissues-Of-Origin." *Cell* 164 (1-2): 57–68.

Stunnenberg, Hendrik G., International Human Epigenome Consortium, and Martin Hirst. 2016. "The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and

18

Discovery." *Cell* 167 (7): 1897.

Sun, Kun, Peiyong Jiang, K. C. Allen Chan, John Wong, Yvonne K. Y. Cheng, Raymond H. S. Liang, Wai-Kong Chan, et al. 2015. "Plasma DNA Tissue Mapping by Genome-Wide Methylation Sequencing for Noninvasive Prenatal, Cancer, and Transplantation Assessments." *Proceedings of the National Academy of Sciences of the United States of America* 112 (40): E5503–12.

Swarup, V., and M. R. Rajeswari. 2007. "Circulating (cell‑free) Nucleic acids–A Promising, Non‑invasive Tool for Early Detection of Several Human Diseases." *FEBS Letters*. https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/j.febslet.2007.01.051.

Tan, Daniel S. W., Fui Teen Chong, Hui Sun Leong, Shen Yon Toh, Dawn P. Lau, Xue Lin Kwang, Xiaoqian Zhang, et al. 2017. "Long Noncoding RNA EGFR-AS1 Mediates Epidermal Growth Factor Receptor Addiction and Modulates Treatment Response in Squamous Cell Carcinoma." *Nature Medicine* 23 (10): 1167–75.

Ulz, Peter, Gerhard G. Thallinger, Martina Auer, Ricarda Graf, Karl Kashofer, Stephan W. Jahn, Luca Abete, et al. 2016. "Inferring Expressed Genes by Whole-Genome Sequencing of Plasma DNA." *Nature Genetics* 48 (10): 1273–78.

Visel, Axel, Matthew J. Blow, Zirong Li, Tao Zhang, Jennifer A. Akiyama, Amy Holt, Ingrid Plajzer-Frick, et al. 2009. "ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers." *Nature* 457 (7231): 854–58.

Vivian, John, Arjun Arkal Rao, Frank Austin Nothaft, Christopher Ketchum, Joel Armstrong, Adam Novak, Jacob Pfeil, et al. 2017. "Toil Enables Reproducible, Open Source, Big Biomedical Data Analyses." *Nature Biotechnology* 35 (4): 314–16.

Wan, Jonathan C. M., Charles Massie, Javier Garcia-Corbacho, Florent Mouliere, James D. Brenton, Carlos Caldas, Simon Pacey, Richard Baird, and Nitzan Rosenfeld. 2017. "Liquid Biopsies Come of Age: Towards Implementation of Circulating Tumour DNA." *Nature Reviews. Cancer* 17 (4): 223–38.

Weiner, Assaf, Tsung-Han S. Hsieh, Alon Appleboim, Hsiuyi V. Chen, Ayelet Rahat, Ido Amit, Oliver J. Rando, and Nir Friedman. 2015. "High-Resolution Chromatin Dynamics during a Yeast Stress Response." *Molecular Cell* 58 (2): 371–86.

Wu, Jing, Wen Qin, Ying Wang, Arsil Sadik, Jilan Liu, Yangyang Wang, Ping Song, et al. 2018. "SPDEF Is Overexpressed in Gastric Cancer and Triggers Cell Proliferation by Forming a Positive Regulation Loop with FoxM1." *Journal of Cellular Biochemistry* 119 (11): 9042–54.

Xu, Rui-Hua, Wei Wei, Michal Krawczyk, Wenqiu Wang, Huiyan Luo, Ken Flagg, Shaohua Yi, et al. 2017. "Circulating Tumour DNA Methylation Markers for Diagnosis and Prognosis of Hepatocellular Carcinoma." *Nature Materials* 16 (11): 1155–61.

Zemmour, Hai, David Planer, Judith Magenheim, Joshua Moss, Daniel Neiman, Dan Gilon, Amit Korach, et al. 2018. "Non-Invasive Detection of Human Cardiomyocyte Death Using Methylation Patterns of Circulating DNA." *Nature Communications* 9 (1): 1443.

Acknowledgements

Materials and Methods

Patients

All clinical studies were approved by the relevant local ethics committees. The study was approved by the Ethics Committees of the Hebrew University-Hadassah Medical Center of Jerusalem. Informed consent was obtained from all subjects or from their legal guardians before blood sampling.

Sample collection

Blood samples were collected in VACUETTE® K3 EDTA tubes, transferred immediately to ice and 1X protease inhibitor cocktail (Roche) and 10mM EDTA were added. The blood was centrifuged (10 minutes, $1500 \times$ g, 4ºC), the supernatant was transferred to fresh 14ml tubes, centrifuged again (10 minutes, $3000 \times$ g, 4ºC), and the supernatant was used as plasma for ChIP experiments. The plasma was used fresh or flash frozen and stored at -80ºC for long storage.

cfChIP-seq

Bead preparation

50μg of antibody were conjugated to 5mg of epoxy M270 Dynabeads (Invitrogen) according to manufacturer instructions. The antibody-beads complexes were kept at 4ºC in PBS, 0.02% azide solution.

AB	Company	Catalog Number
IgG	Cell signalling	2729S

20

H3K4Me1	Diagenode	C15410194
H3K4Me2	Diagenode	C15410035
H3K4Me3	Diagenode	C15410003
H3K36Me3	Diagenode	C15410192

Immunoprecipitation, NGS library preparation, and sequencing

0.2mg of conjugated beads (~2μg of antibody) were used per cfChIP-seq sample. The antibody-beads complexes were added directly into the plasma (1-2 ml of plasma) and allowed to bind to cf-nucleosomes by rotating overnight at 4ºC. The beads were magnetized and washed 8 times with blood wash buffer (BWB 50mM Tris-HCl , 150mM NaCl, 1% Triton X-100, 0.1% Sodium DeoxyCholate, 2mM EDTA , 1X protease inhibitors cocktail), and three times with 10mM Tris pH 7.4. All washes were done with 150ul buffer on ice by shifting the beads from side to side on a magnet. Do not use vacuum to remove supernatant during washes in buffers that do not contain detergents.

On-beads chromatin barcoding and library amplification was done as previously described (Lara-Astiaso et al. 2014; Gutin et al. 2018) except for the DNA elution and cleanup step where the beads were incubated for 1 hour at 55ºC in 50μl of chromatin elution buffer (10mM Tris pH 8.0, 5mM EDTA, 300mM NaCl, 0.6% SDS) supplemented with 50 units of proteinase K (Epicenter), and the DNA was purified by 0.9 X SPRI cleanup (Ampure xp, agencourt). The purified DNA is eluted in 25 μl EB (10mM tris pH 8.0) and 23 μl of the eluted DNA were used for PCR amplification with Kapa hotstart polymerase (16 cycles). The amplified DNA was purified by 0.8 X SPRI cleanup and eluted in 12 μl EB. The eluted DNA concentration was measured by Qubit and the fragments size was analysed by tapestation visualization. Note: If too much adapter dimers were still visible by tapestation post library amplification, we recommend pooling samples and performing additional X 0.8 SPRI DNA cleanup, or separating the pooled samples on a 4% agarose gel (E-Gel® EX Agarose Gels, 4%, Invitrogen), and gel purification of fragments larger than adapter dimers (>150bp). DNA libraries were paired end sequenced by Illumina NextSeq 500.

Sequence analysis

Reads were aligned to the human genome (hg19) using bowtie2 with 'no-mixed' and 'no-discordant'

21

flags. We discarded reads with low alignment scores and duplicate fragments. See Table S1 for read number, alignment statistics, and numbers of unique fragments for each sample.

Roadmap Epigenome atlas

We downloaded aligned read data from the Roadmap Epigenome Consortium database (Table S11). For our analysis we discarded pre-natal, ESC, and cell-line samples, resulting with 64 tissues and cell types (Table S12). The aligned read files were then processed with the same pipeline as cfChIP-seq samples.

Tumor-type Gene Signatures

We downloaded RNA-seq data from the TCGA and GTEx projects as analysed by the Xena project (Vivian et al. 2017) (Table S11). We defined the set of genes that are over-expressed in a tumor type to satisfy three requirements: 1) Significantly higher expression in tumor samples compared to the corresponding tissue samples (t-test, $q < 0.001$ after FDR correction); 2) Significantly higher expression compared to all healthy samples (t-test, $q < 0.001$ after FDR correction); and 3) Median expression in the tumor is higher than median expression in each of the healthy samples.

Expected healthy expression level

To best emulate expression profiles, we performed *in silico* mix of the four cells types that contribute the most to cfDNA (Moss et al. 2018): neutrophils, 32%; monocytes 32%; erythrocyte progenitors 20%; and NK cells 5%. The gene expression for these cell types was downloaded from BLUEPRINT consortium website (Table S11).

TSS/Enhancer location catalogue

We downloaded the Roadmap Epigenome Consortium ChromHMM annotation of all consolidated tissues (Table S11). Using these annotations we constructed a catalogue of potential functional sites (enhancers, TSSs, and genes). We extended the catalogue to include 3kb regions centered on TSS of annotated transcripts in the UCSC gene database and ENSEMBL transcript database (Table S11). We used the combined catalogue to define regions along the genome. We used different version of the catalogue for analysis of each antibody, to match the mark. For H3K4me3 analysis we used only TSSs, for H3K36me3 analysis we used only gene bodies, and for H3K4me2 we had annotations of TSSs and enhancers. In each

22

version of the catalogue, the remaining mappable genome regions were assigned to background, and tiled at 5kb windows. See Supplemental Note for more detailed procedures.

We quantified the number of reads covering each region in the catalogue in each of our samples and atlas samples. We estimated locally adaptive model of non-specific reads along the genome for each of the samples, and extracted counts that represent specific ChIP signal in the catalogue for each sample (Supplemental Note). These were then normalized (Supplemental Note) and scaled to 1M reads in the reference healthy samples.

Tissue Signatures

To define tissue specific signatures of a specific modification, we examined binned representation of the atlas according to our catalogue. For each tissue we defined a signature of unique windows with signal in one of the samples of the target tissue and without coverage in all others (Supplemental Note).

Gene level analysis

For each gene we defined the set of windows that match the gene (TSS in H3K4me3/2 and gene body in H3K36me3). The signal for a gene is the aggregate signal-background over windows associated with it (Supplemental Note).

Statistical analysis

We test whether a signature is present in the analysis of Figure 4. Formally, we examined whether we can reject the null hypothesis that the number of reads in signature windows will be Poisson distribution according to background rate (Supplemental Note). We compute p-value of the actual number of observed reads in signature windows as the probability of having this number or higher according to the null hypothesis. Rejection of the null hypothesis for a specific signature is an indication that some of the windows in the signature carry the modification in question in a subpopulation of cells contributing to the cf-nucleosome pool.

The second test is whether a gene presents a high signal with respect to its level in healthy baseline subjects (Figure 5B). We use average signal from 5 healthy samples to define the average number of reads in each window. We then estimate two sample-specific parameters: 1) background rate (discussed above) and 2) a scaling factor that rescales average expectations to the sequencing depth of the specific

23

sample (Supplemental Note). Together, these define the expected coverage of each gene-associated group of windows under the null-hypothesis that the subject is from the healthy population. We compute p-value of the actual number of observed reads in the gene windows as the probability of having this number or higher according to the null hypothesis.

Pathways and complexes

We downloaded a large collection of gene expression signatures representing different cellular processes, protein complexes, and transcriptional responses (Liberzon et al. 2015; Drew et al. 2017; Giurgiu et al. 2019; Kamburov et al. 2013) (Table S11). Each such pathway (or complex) is represented by a list of genes participating in it. Not all of the pathways represented genes with coherent co-expression, and so we filtered the compendium to include only pathways that behave coherently across the 64 Roadmap Epigenomics reference samples (Table S6; Supplemental Note).

The score assigned to a pathway on a sample is the sum of normalized signal for the genes in the pathway. We evaluated each pathway on 33 healthy samples, to estimate distribution of scores in healthy reference. We then evaluate scores in other samples using the Z-score of this distribution. The significance of scores was determined by two-tailed test assuming Z-scores from a normal distribution.
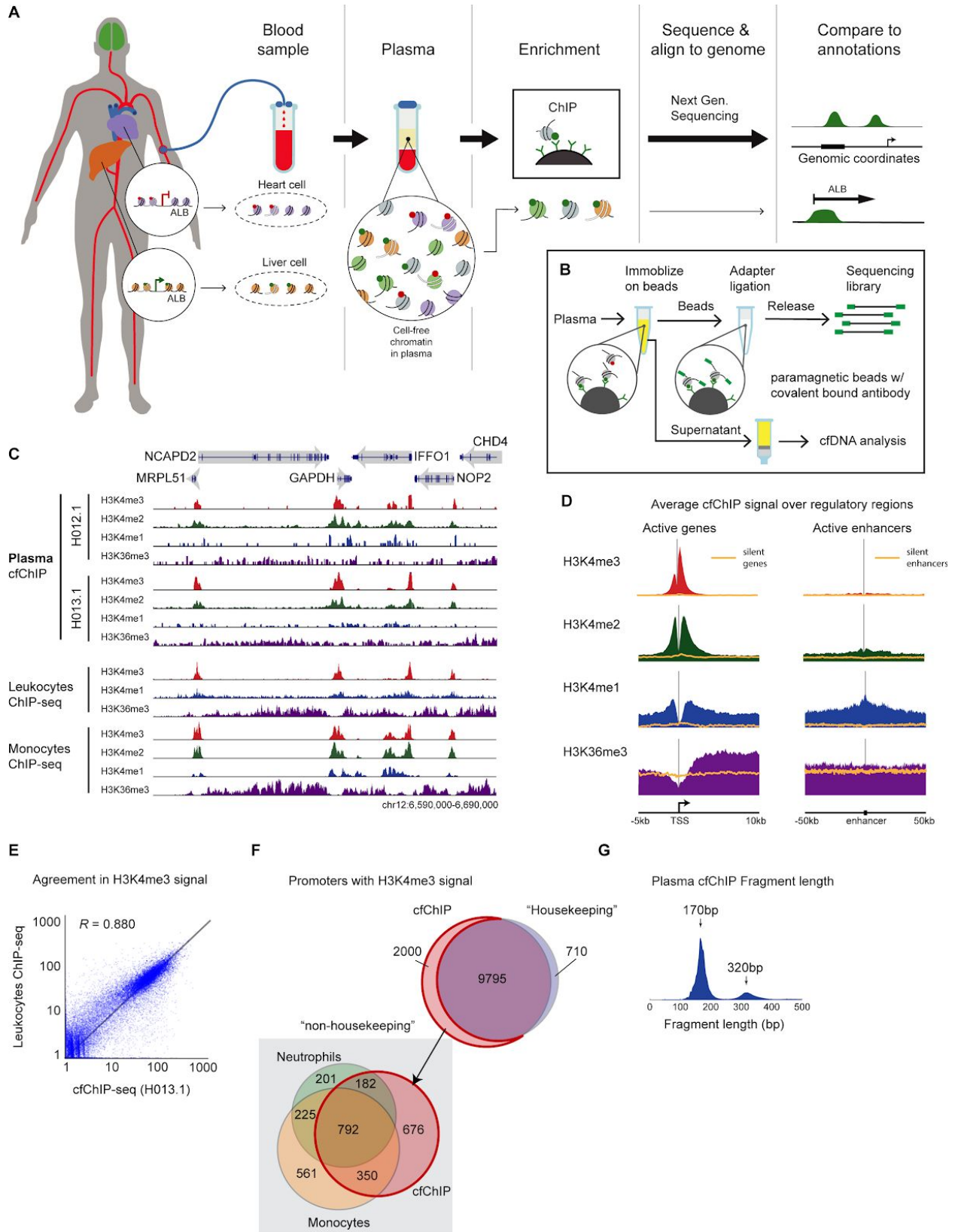
25

Figure 1: Immunoprecipitation of chromatin from plasma

A. cfChIP-seq method outline. Chromatin fragments from different cells are released to the bloodstream. These fragments are immunoprecipitated, and sequenced. Interpretation of the resulting sequences informs of gene activity programs in the tissue of origin. As an example, death of liver cells releases to the blood nucleosomes from the ALB promoter marked with H3K4me3. These are mixed with other circulating nucleosomes from other cells with the ALB promoter not marked by H3K4me3. After immunoprecipitation of H3K4me3 cf-nucleosomes and sequencing, we can detect fragments of DNA aligned to ALB promoters. These are indicative of death of liver cells since ALB promoter is not marked by H3K4me3 in other cell types.

B. cfChIP-seq protocol, using antibody covalently bound to paramagnetic beads. Target fragments are immunoprecipitated directly from plasma. After washing the plasma, on-bead-ligation is performed to add sequencing adapters (with indexing barcodes) to the fragments. Reverse crosslinking and PCR amplification results in sequencing-ready libraries.

C. Example of cfChIP-seq signal on a segment of chromosome 12. Top tracks are cfChIP-seq signals from two healthy subjects. The lower tracks are published ChIP-seq results from human white blood cells (leukocytes) (Roadmap Epigenomics Consortium et al. 2015). In each group we show four tracks corresponding to cfChIP-seq (or ChIP) of four marks -- H3K4me3 (red), H3K4me2 (green), H3K4me1 (blue), and H3K36me3 (purple).

D. Meta analysis of cfChIP-seq signal over active promoters and enhancers. In each plot the orange line denotes average of corresponding control regions (inactive genes and enhancers), providing an estimate of the background.

E. Comparison of H3K4me3 cfChIP-seq from a healthy subject against ChIP-seq from leukocytes (Roadmap Epigenomics Consortium et al. 2015). Each dot corresponds to a gene with the coordinates showing its coverage (after normalization) in the two samples: x-axis healthy cfChIP-seq sample, y-axis leukocytes cfChIP-seq.

F. Analysis of promoters of RefSeq genes with a significant cfChIP-seq signal (methods) in healthy subjects. cfChIP-seq captures most housekeeping promoters (ones that are marked in most samples in the reference compendium). The remaining 2000 non-housekeeping genes in cfChIP-seq show large overlaps with non-housekeeping promoters marked in neutrophils and monocytes, the two blood cell types that contribute most to cfDNA in healthy subjects.

G. Size distribution of sequenced cfChIP-seq fragments shows clear pattern of mono- and
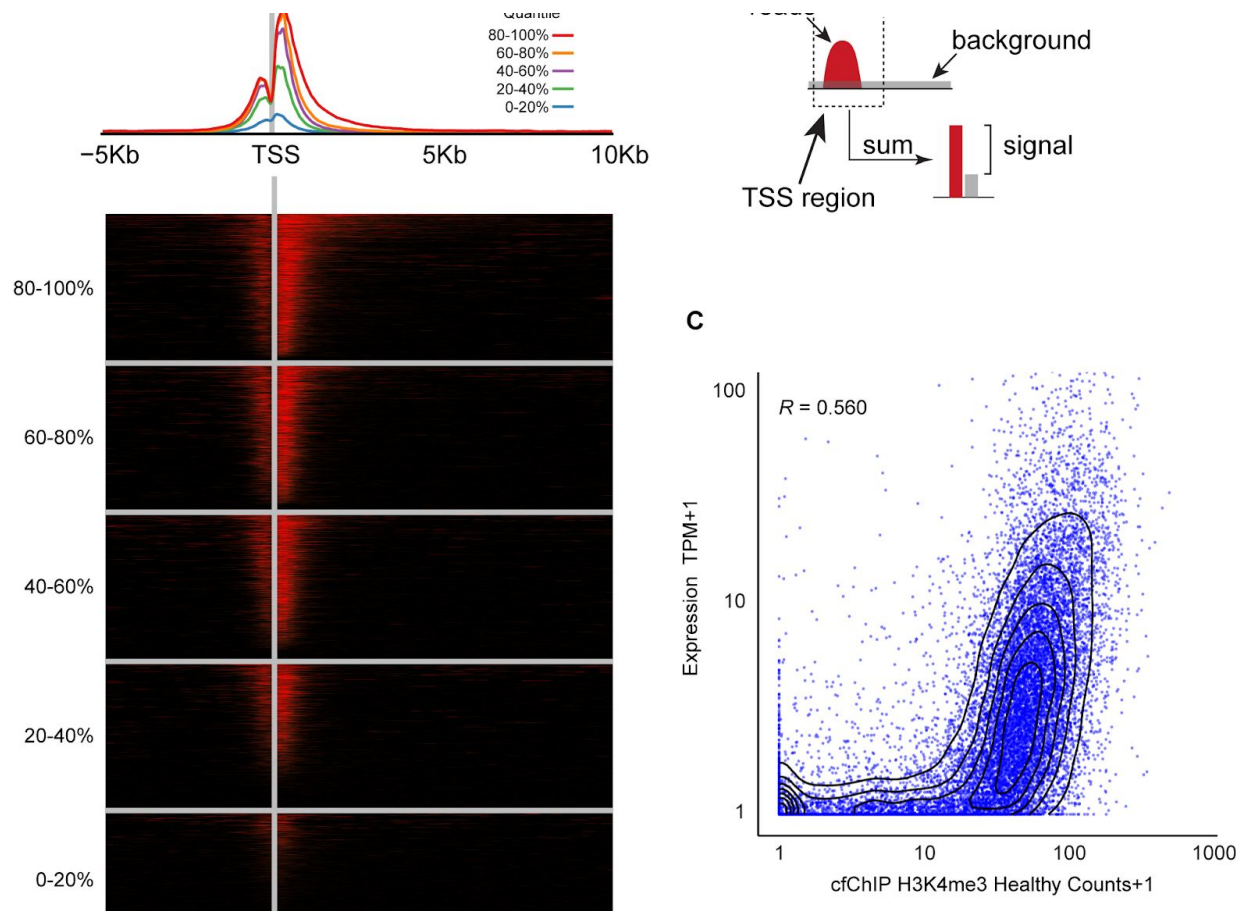
26

di-nucleosome fragment sizes.

Figure 2: H3K4me3 cfChIP-seq signal is correlated with expression levels

  A. Meta-plot and heat map for genes at different gene expression levels.

  B. Schematic of how we estimate specific-signal at each promoter (Methods, Supplemental Note).

  C. Comparison of H3K4me3 cfChIP-seq signal from a healthy subject (H012.1) with expected gene expression levels (Methods). Each dot is a gene. x-axis: normalized number of H3K4me3 reads in gene promoter. y-axis: expected expression in number of transcripts/million (TPM).

29

Figure 3: cfChIP-seq detects a rich tapestry of expression profiles

A. Heatmap showing patterns of the relative H3K4me3 cfChIP-seq coverage on 17,750 RefSeq genes that are marked in at least one of the samples. The normalized coverage on the gene promoter (Methods) was log-transformed (log2(1+coverage)) and then adjusted to zero mean for each gene across the samples. The samples include cfChIP-seq samples from a compendium that includes healthy subjects, cancer patients, acute myocardial infarction (AMI) patients, and patients treated with partial hepatectomy (see text).

B. Average profiles for 35 clusters from the heatmap of (A). Top panel: number of genes in the cluster. Middle panel: top half is average of relative H3K4me3 signal from cfChIP-seq samples and Roadmap epigenomics (Roadmap Epigenomics Consortium et al. 2015) ChIP-seq samples, bottom half is the same for RNA expression from GTEx tissues (GTEx Consortium 2015) and BLUEPRINT blood types (Stunnenberg, International Human Epigenome Consortium, and Hirst 2016). Specific samples are highlighted on the right, and specific clusters are highlighted below.

C. Elaboration of specific rows in B (see guidelines) against the highlighted clusters. Bars show change from mean for the cluster in the specific sample/tissue/cell-type.
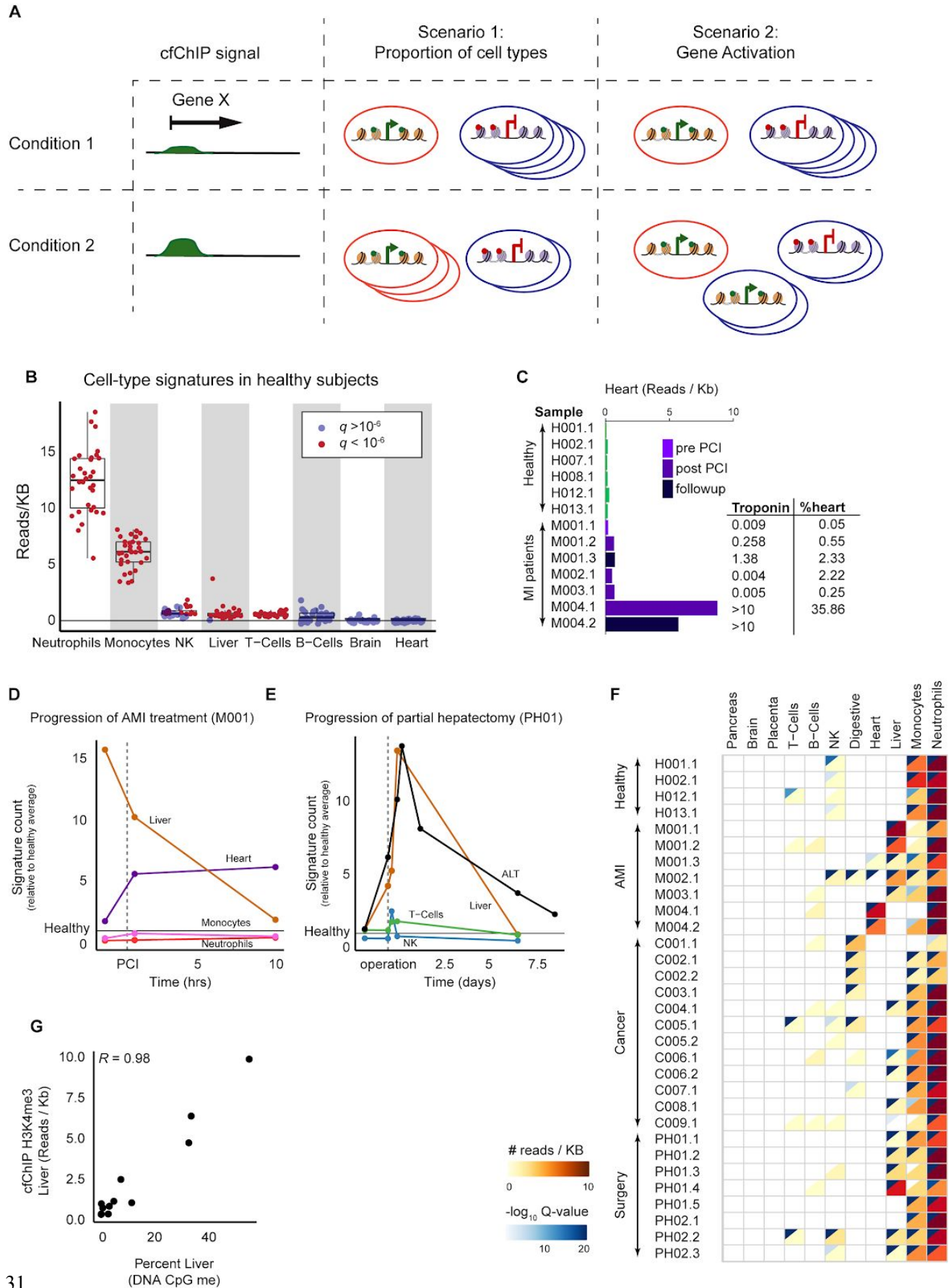
31

Figure 4: cfChIP-seq identifies cell-type specific expression programs

A. Schematic representation of two mechanisms leading to change in cfChIP-seq signal. Scenario 1 represents change in the proportion of cell types that contribute to the cfDNA pool. Scenario 2 depicts change in expression of specific genes and programs in a specific cell-type.

B. Evaluation of average signal (normalized reads/Kb) for signatures of different cell types in 33 healthy samples from 15 donors. Each dot is a sample, and boxplot summarize distribution of values in the group. Dots marked in red are ones where the signatures are significantly different from background signal (Methods). We observe evidence for H3K4me3 signal in promoters specific to neutrophils, monocytes, NK cells, liver cells, and T-cells. We observe mild signal, but not significantly above background, from B-cells, and no signal from heart or brain. These proportions are consistent with independent estimates based on DNA CpG methylation marks (Moss et al. 2018).

C. H3K4me3 cfChIP-seq signal in heart-specific windows in representative samples of healthy subjects and acute myocardial infarction (AMI) patients (see Table S2 for details). **Inset**: measured troponin levels and percent cfDNA from cardiomyocyte as measured using DNA CpG methylation markers (Zemmour et al. 2018) from the same blood draws.

D. Changes in signature strength in an AMI (M001) patient before/after PCI. Signatures strength are normalized to healthy subjects. Note the increase in heart and decrease in liver signatures post PCI.

E. Changes in cfChIP-seq liver signature (brown bars) and ALT levels (liver damage biomarker, black line) from blood samples of a patient treated with partial hepatectomy (PH01).

F. Heatmap showing significance of selected cell-type signatures in healthy subjects and patients. Each cell in the map is divided in half, the top left half represents statistical significance (FDR corrected q-value) and the bottom half density of reads in the signature (normalized reads per kb) (see Table S5, Methods).

G. Comparison to external indication of hepatocyte death: x-axis: percent cfDNA from hepatocyte as measured using DNA CpG methylation markers (Lehmann-Werman et al. 2018). y-axis: strength of cfChIP-seq signal at liver-specific signature (relative to healthy subjects).
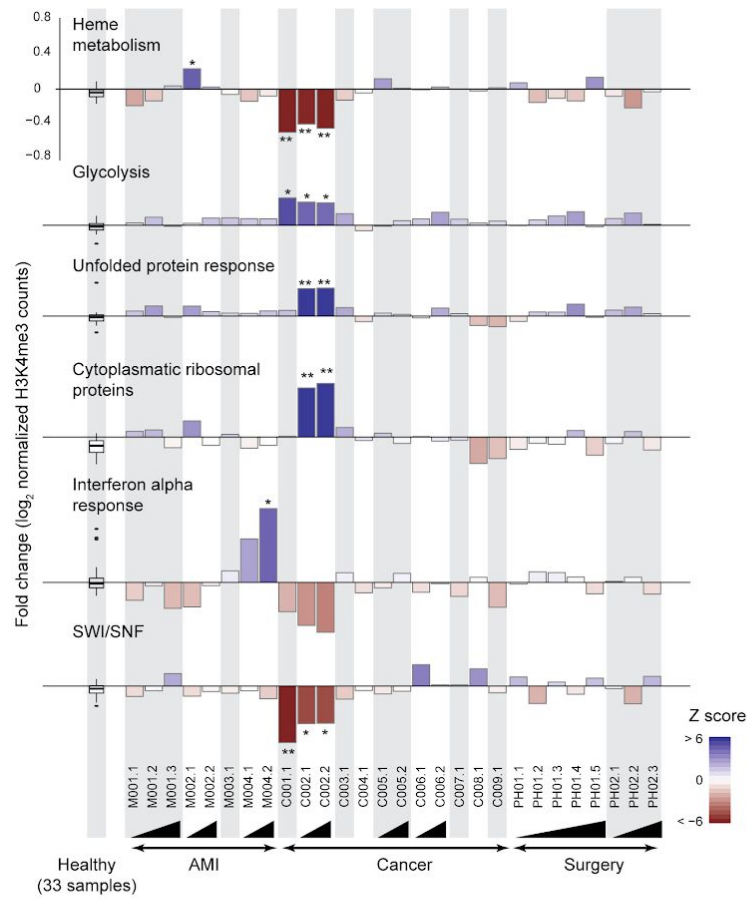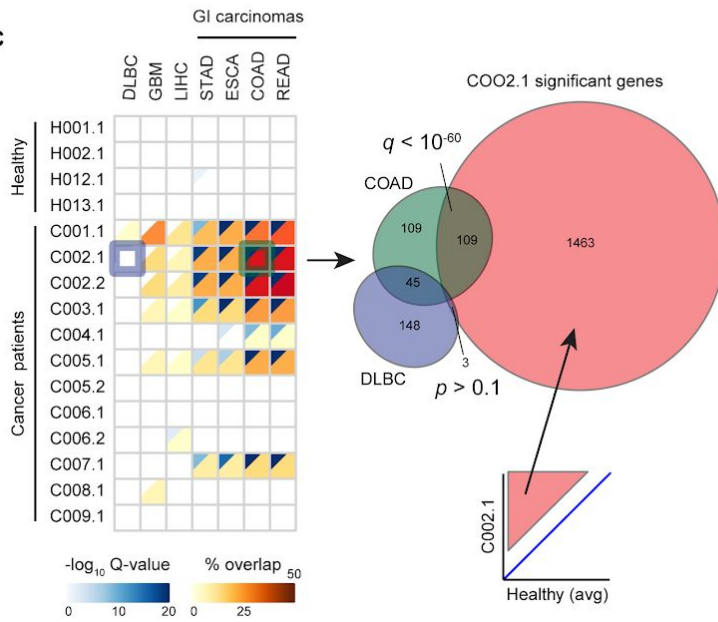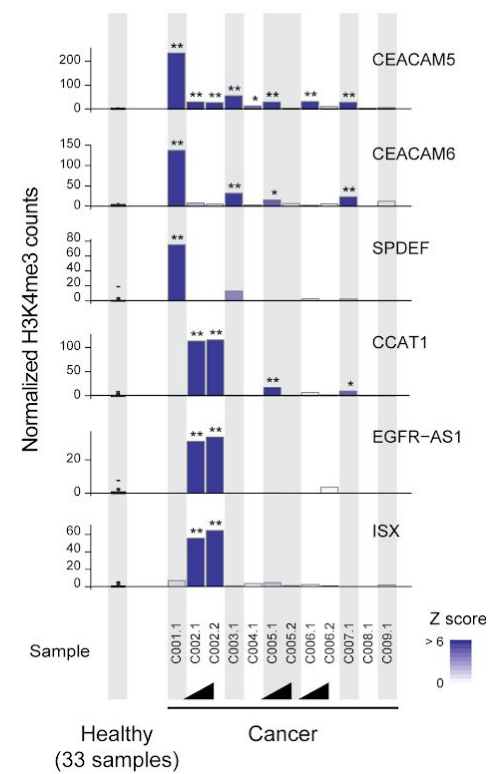
32
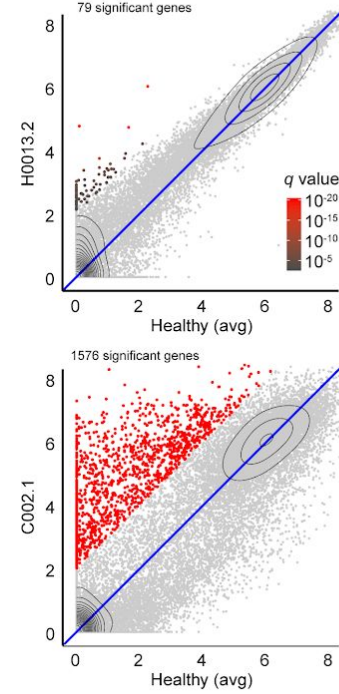
Figure 5: cfChIP-seq identifies changes in expression programs

A. Examples for changes in cfChIP-seq signal strength of specific gene programs across patient samples (Methods). To evaluate a specific program, we sum the normalized coverage of genes in the program in each sample. We compute the distribution of values over reference cohort (33 samples from 15 healthy subjects), and use these to score deviation in patients and healthy samples not included in the reference cohort (color: Z-score according to healthy samples; q-value: p-value assuming gaussian distribution with FDR correction; * $q < 10^{-3}$, ** $q < 10^{-6}$). Black triangles represent multiple samples (sorted according to collection time) from the same subject (Table S2).

B. Detection of significant high coverage genes in a sample. Top: a sample of a healthy subject, bottom: sample from GI-tract cancer patient (Table S2). For each gene we compare mean normalized coverage in a reference cohort (x-axis) against the normalized coverage in the sample (y-axis). Both counts are log-transformed (log2(1+coverage). For a gene to be significant it has to be above the y=x+2 line (e.g., 4-fold difference) and pass a test whether the observed number of reads is significantly higher than we would expect based on healthy value (corrected q-value shown by color scale; Methods).

C. Heatmap showing enrichment of tumor-specific signatures (Methods) in the group of significantly high coverage genes (as in 4B). Each cell is divided in half, the top left half represents statistical significance (FDR corrected q-value) and the bottom half overlap with the signature (% number of genes in signature). Inset: illustration of the overlap for two signatures (DLBC: diffuse large B-cell lymphoma; COAD: colorectal adenocarcinoma) against high coverage genes in C002.1.

D. Examples of cancer-associated genes and their signal in different samples. Similar to 4A, except that units are normalized counts for the gene. The values in samples from healthy subject are close to 0.
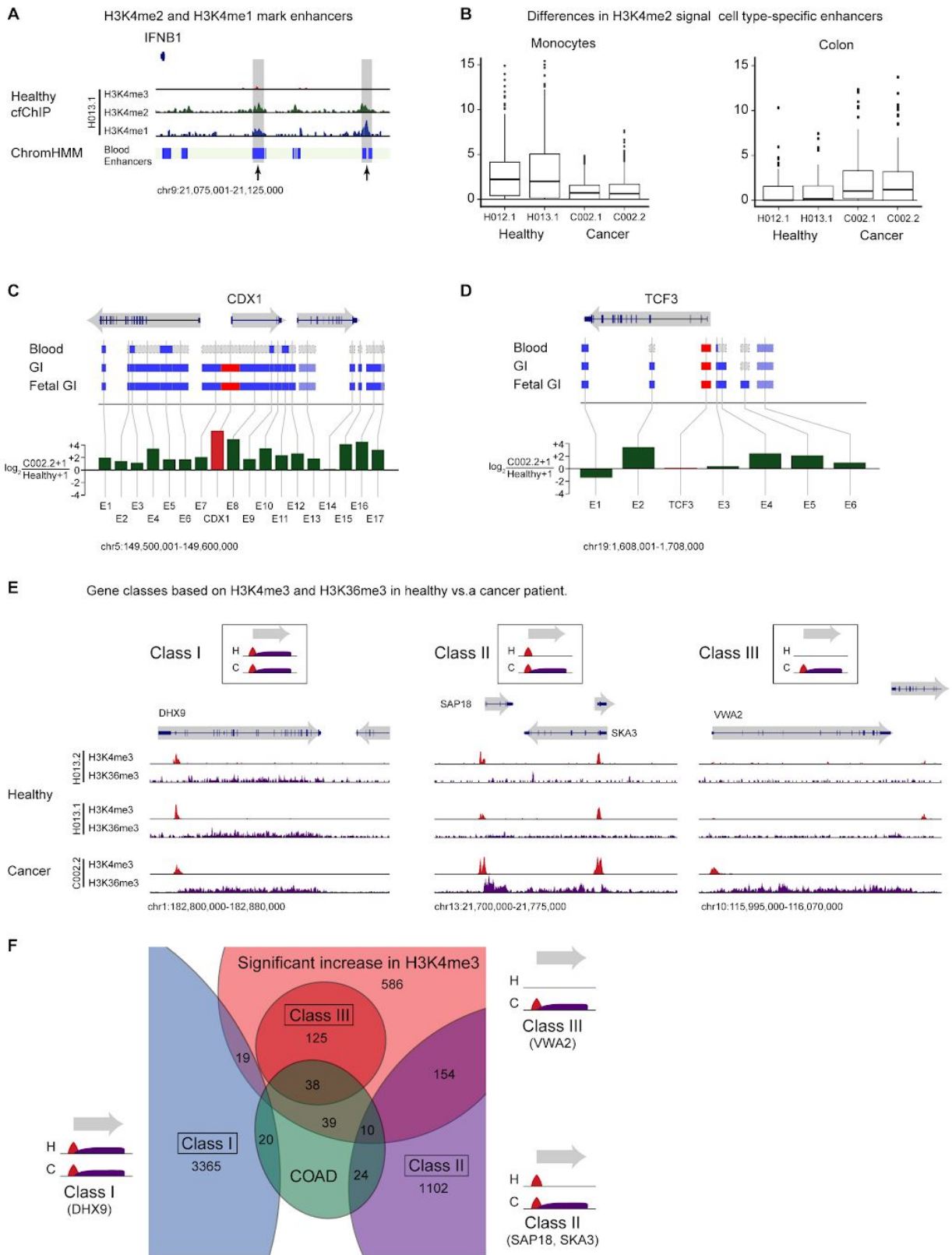
35

Figure 6: cfChIP-seq of additional marks is informative on enhancer usage and gene expression

A. Browser view of cfChIP-seq tracks from a healthy subject in the locus adjacent to IFNB1. ChromHMM tracks show prediction of enhancers according to combination of histone modification and chromatin accessibility assays (Roadmap Epigenomics Consortium et al. 2015). Arrows mark regions with enriched di- and mono-methylation.

B. Normalized H3K4me2 coverage over a group of tissue specific enhancers (Methods) highlights differences between healthy subjects and two samples from a cancer patient with GI-tract adenocarcinoma (Table S3). Shown are the distribution of normalized coverage/enhancer for monocyte-specific (left) and colon-specific (right) enhancers.

C. Browser track around the TCF3 gene. Shown are H3K4me3 and H3K4me2 cfChIP-seq normalized signal (top tracks) and tissue-specific chromHMM annotations (bottom panel). Specific enhancers are labeled (see text).

D. Same as in Figure 6C around the CDX1 gene.

E. Same as in Figure 6C around the CEACAM5 gene.

F. Browser views at genes that demonstrate different H3K4me3 and H3K36me3 classes. Class I: genes are marked with both H3K4me3 and H3K36me3 in healthy and cancer patient samples. Class II: genes are marked by H3K4me3 in both, but with H3K36me3 only in the cancer patient samples (gain of H3K36me3). Class III: genes are not marked in healthy subject samples and are marked by both marks in the cancer patient samples (gain of both marks).

G. Venn diagram (zoom in view) showing the relations of genes from the three classes (F) with the set of genes that show increased H3K4me3 and the set of colorectal adenocarcinoma (COAD) genes.

36