Subtype Heterogeneity and Epigenetic Convergence in Neuroendocrine Prostate Cancer

Paloma Cejas[1,2,3], Yingtian Xie[1,2], Alba Font-Tello[1,2], Klothilda Lim[1,2], Sudeepa Syamala[1,2], Xintao Qiu[1,2], Alok K. Tewari[1,2,4], Neel Shah[1,2], Holly M Nguyen[5], Radhika A. Patel[6], Lisha Brown[5], Ilsa Coleman[6], Wenzel M. Hackeng[7], Lodewijk Brosens[7], Koen M.A. Dreijerink[8], Leigh Ellis[4,9], Sarah Abou Alaiwi[1], Ji-Heui Seo[1], Mark Pomerantz[1], Alessandra Dall'Agnese[10], Jett Crowdis[1,4], Eliezer M. Van Allen[1,4], Joaquim Bellmunt[11], Colm Morrisey[5], Peter S. Nelson[6], James DeCaprio[1], Anna Farago[12], Nicholas Dyson[12], Benjamin Drapkin[13,14,15], X. Shirley Liu[2,16], Matthew Freedman[1,2], Michael C. Haffner[6,17,18], Eva Corey[5], Myles Brown[1,2*] and Henry W. Long[1,2*]

1. Department of Medical Oncology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts, USA.

2. Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts, USA.

3. Translational Oncology Laboratory, Hospital La Paz Institute for Health Research (IdiPAZ) and CIBERONC, La Paz University Hospital, Madrid, Spain.

4. Broad Institute of MIT and Harvard, Cambridge, MA, USA

5. Department of Urology, University of Washington, Seattle, WA, USA

6. Divisions of Human Biology and Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

7. Department of Pathology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

8. Department of Endocrinology, Amsterdam UMC, Amsterdam, The Netherlands.

9. Department of Oncologic Pathology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA

10. Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, MA 02142, USA.

11. Beth Israel Deaconess Medical Center and PSMAR-IMIM Lab. Harvard Medical School, Boston, Massachusetts, USA.

12. Massachusetts General Hospital Cancer Center, Boston, Massachusetts, USA.

13. Nancy B. and Jake L. Hamon Center for Therapeutic Oncology Research, Dallas, TX, USA.

14. Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX, USA

15. Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA.

16. Department of Data Science, Dana-Farber Cancer Institute, Harvard T.H. Chan School of Public Health, Boston, MA, USA

17. Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

18. Department of Pathology, University of Washington, Seattle, WA, USA.

* Corresponding authors.

Abstract

Neuroendocrine carcinomas (NEC) are tumors expressing markers of neuronal differentiation that can arise at different anatomic sites but have strong histological and clinical similarities. Here we report the chromatin landscapes of a range of human NECs and show convergence to the activation of a common epigenetic program. With a particular focus on treatment emergent neuroendocrine prostate cancer (NEPC), we analyzed cell lines, patient-derived xenograft (PDX) models and human clinical samples to show the existence of two distinct NEPC subtypes based on the expression of the neuronal transcription factors ASCL1 and NEUROD1. While in cell lines and PDX models these subtypes are mutually exclusive, single cell analysis of human clinical samples exhibit a more complex tumor structure with subtypes coexisting as separate sub-populations within the same tumor. These tumor sub-populations differ genetically and epigenetically contributing to intra- and inter-tumoral heterogeneity in human metastases. Overall our results provide a deeper understanding of the shared clinicopathological characteristics shown by NECs. Furthermore, the intratumoral heterogeneity of human NEPCs suggests the requirement of simultaneous targeting of coexisting tumor populations as a therapeutic strategy.

Introduction

Neuroendocrine carcinomas (NEC) are high grade tumors, the most common being small-cell lung cancer (SCLC), that can also arise in the colon, prostate or the bladder among other anatomic sites. NECs are characterized by aggressive clinical behavior and a poor prognosis[1]. Histomorphologically, NEC comprise a heterogenous group of tumors which can have features of small cell carcinoma, and show expression of neuroendocrine markers including SYP, CHGA and INSM1[2]. Given these common characteristics, NECs constitute an unique clinicopathological entity despite their distinct anatomical origins[3]. From a genetic standpoint, NECs are often characterized by genomic aberrations in *RB1* and *TP53*[4]. The association of these genetic alterations with NEC etiology is exemplified in Merkel Cell Carcinoma (MCC). This aggressive NEC of the skin is caused in at least 60% of all MCC by clonal integration of Merkel cell polyomavirus DNA into the tumor genome with persistent expression of viral T antigens (MCCP; MCC Polyomavirus-positive). These viral proteins produce tumorigenesis by interfering with cellular tumor-suppressor proteins including RB1[5]. Alternatively, MCC can be caused by ultraviolet damage leading to highly mutated genomes causing a non-viral form of MCC (MCCN;

MCC non-viral) that almost invariably carry mutations in *TP53* and *RB1*[6]. Despite these distinct etiologies, both forms of MCC have similar presentations, prognosis, and response to therapy. Gastrointestinal NEC (GINEC) is another relevant category of NECs that typically harbor *TP53* and *RB1* alterations and are clinically aggressive and highly proliferative. In that respect, the GINECs differ from the well differentiated GI carcinoids that, while also showing a NE phenotype, are typically clinically indolent[1] and not associated with *TP53* and *RB1* alterations.

NEC can emerge either *de novo* or as a result of therapeutic pressure[7–10]. SCLC most often emerges *de novo* but can emerge after treatment of *EGFR* mutant NSCLC[11]. SCLC has been sub-classified based on the differential expression of the basic Helix-Loop-Helix (bHLH) transcription factors (TFs) ASCL1 and NEUROD1[12]. These neuronal lineage TFs have been implicated in lung development and in the maturation of resident neuroendocrine cells of the lung[13,14,15]. They are also involved in the carcinogenic process as shown in mouse models of SCLC where ASCL1 is required for tumor formation[16].

Neuroendocrine prostate cancers (NEPC), in contrast, arises most frequently as a treatment emergent phenotype from prostatic adenocarcinomas after treatment designed to repress AR pathway activity[17] and only rarely arises *de novo*. The improved potency and specificity of anti-AR therapy has led to an increased prevalence of NEPC[18]. NEPC has poor prognosis, very limited therapeutic options and is currently treated as a homogeneous disease. A better understanding of the molecular basis of NEPC is required in order to design more efficient targeted therapeutic options.

Here we utilize epigenetic profiling of a range of NECs to understand how the common phenotype is maintained across tumor types with particular focus on NEPC. We find that multiple bHLH transcription factors are implicated in a common epigenetic program across all NECs. Furthermore, epigenetic profiling reveals the existence of subtypes in treatment emergent NEPC in line with what has been described in *de novo* SCLC, despite arising from treated prostate adenocarcinoma. The subtypes coexist as separate sub-populations with distinct chromatin states within the same human NEPC specimens exhibiting heterogeneity that have clinical implications.

Results

Neuroendocrine carcinomas share a common landscape of DNA accessible regions.

Histomorphologically, NECs show similarities that could result from activation of common transcriptional regulators[19]. To investigate the impact of chromatin accessibility in determining the NEC phenotype we profiled the epigenetic landscape of NECs arising in various anatomic locations using ATAC-seq and RNA-seq applied to PDX models of NEPC[20], SCLC[20,21], and Merkel cell carcinoma (MCC) as well as GINEC clinical samples (Table 1). As NEC can emerge from a preexisting adenocarcinoma (AD), as typified by NEPC[22] and occasionally by SCLC[18,23], we hypothesized that those histologies are extremes of a spectrum of tumor progression. To determine how the chromatin state differs between neuroendocrine (NE) and adenocarcinoma (AD) by ATAC-seq analysis, we also generated data from metastatic prostate adenocarcinoma PDX models and used TCGA data for primary Prostate ADs (PRAD) and non-small cell lung adenocarcinomas (LUAD)[24]. We obtained high quality data with the fraction of reads in peaks (FrIP) scores in the range of 10-35 and peak numbers in the range of 25-75k (Table 1). Replicate profiling of two samples showed high concordance (Suppl. Figure 1a). Unsupervised principal-component analysis (PCA) performed on the ATAC-seq data revealed that the NECs cluster together indicating a convergent chromatin state, in contrast to the ADs that are segregated by anatomic site (Figure 1a). The sample-sample correlation of the ATAC-seq peaks also supports the result that NECs are more similar to each other than to their AD counterparts from the same tissue (Figure 1b), although the MCCs are particularly homogeneous (one MCCN is tightly clustered with five MCCPs). These analyses also clustered prostate PDXs and primary human tumors together emphasizing, in terms of the chromatin state, the value of the prostate LuCaP[25,26] PDXs to model human prostate cancer, as previously validated by histological and molecular characterization[25,26].

To investigate epigenetic drivers involved in the NE chromatin state, we performed a supervised analysis of the DNA accessibility between ADs and NECs and found a high number (n:16517, Padj<0.001, log2(FC) >2) of NE-specific accessible sites shared across all NE tumor types (Figure 1c). This represents a major difference in chromatin organization that we further investigated by GREAT analysis[27] that associates genomic regions with nearby genes and then examines the enrichment of Gene Ontology (GO) pathways. Genes near NE-specific DNA accessible regions showed a significant enrichment in pathways for neural differentiation,

development, morphology and axogenesis (Figure 1d). Next, we used HOMER[28] to investigate NE-specific sites for enrichment of TF DNA-binding motifs. This analysis revealed significant enrichment for motifs of the basic Helix-Loop-Helix (bHLH) TF family, specifically for ATOH1, ASCL1 and NEUROD1 as well as motifs for NFIB, SOX2 and NKX2-1 (Figure 1e and Table 2). ATOH1 has been implicated as a lineage transcription factor in MCC[29,30], while ASCL1 and NEUROD1 have been suggested to have a corresponding role in SCLC[16,31]. NFIB is a TF previously implicated in rewiring the chromatin structure in SCLC[32], while SOX2 and NKX2-1 are also known to be associated with SCLC[33,34]. Examining what motifs co-occur in the ATAC-seq peaks we observed that the module of the main three motifs (ASCL1, ATOH1 and NEUROD1) occurs very frequently combined with either SOX2 or NFIB or with both of them simultaneously (Suppl. Figure 1b).

By comparing the NE-specific sites to published ChIP-seq profiles compiled in CistromeDB[35], we identified TFs whose published binding sites have the highest overlap with the NE-specific ATAC-seq peaks as quantified by GIGGLE score[36]. Consistent with the observed shared epigenetic program amongst NECs of different tissue origins, the top overlapping ChIP-seq datasets were generated from SCLC, MCC or neural lineages (Figure 1f). In particular, binding profiles of ASCL1 (SCLC), NEUROD1 (SCLC and medulloblastoma), and MAX (MCC)[5] had the highest overlap scores with NE-specific accessible chromatin(Figure 1f). We next analyzed the expression of these TFs and other NEC-associated factors within our study samples. As expected, we observed a strong commonality in the expression of NE markers (*SYP, CHGA and INSM1*) and the stemness TF, *SOX2* (Figure 1g) across NECs. We also observed a more mutually exclusive expression pattern of bHLH TFs including *ASCL1* or *NEUROD1* in both NEPC and SCLC, *ASCL1* and *ASCL2* expression in GI-NECs and *ATOH1* expression in MCC (Figure 1g). This suggests tumor and organ specific TF drivers of this common neuroendocrine epigenetic state. Overall our results across a diverse panel of NECs show a common landscape of DNA accessible regions with expression of varying TFs within the same family.

Treatment emergent NEPC can be subclassified based on the expression of ASCL1 and NEUROD1

To explore heterogeneity in the TF regulation of the NEC epigenetic state, we performed an unsupervised analysis of the ATAC-seq data restricted to the NECs. Regardless of tissue of origin, NECs expressing *ASCL1* and/or *ASCL2* were tightly clustered together and were separate

from NECs expressing *ATOH1* or *NEUROD1* (Suppl. Fig 2a). Furthermore, the similarity in terms of the DNA accessibility shown by SCLC and NEPC depends on the status of *ASCL1* or *NEUROD1* expression but not on the tumor type (Suppl. Fig. 2b). An unsupervised analysis of the DNA accessibility in just prostate samples (Figure 2a) showed clear grouping associated with expression of *AR* (all adenocarcinomas), *ASCL1* or *NEUROD1* with the same clustering being apparent by analysis of RNA-seq data of those same prostate samples (Suppl. Fig. 2c). Although neuroendocrine subtypes based on the expression of those TFs have been previously described in SCLC[12,16], the existence of these subtypes in treatment emergent NEPC was unanticipated as ASCL1 and NEUROD1 have been specifically associated with lung neuroendocrine cells[13,14,15], the putative cell of origin of the *de novo* SCLC .

Next we aimed to identify the differential DNA accessibility associated with the ASCL1 and the NEUROD1 NEPC subtypes. Supervised analysis comparing *ASCL1* and *NEUROD1* expressing NEPC samples identified 2863 ASCL1- and 4873 NEUROD1-specific accessible regions (FDR<0.01, log2(FC) > 2) (Figure 2b). We next interrogated the NEPC subtype-specific sites in SCLC and observed a similar patterns of chromatin accessibility at these TF-specific genomic regions in SCLC samples that, in addition, displayed an association between the chromatin state and the differential expression of *ASCL1* and *NEUROD1* (Suppl. Fig. 2d). Notably, SCLC cases that coexpress *ASCL1* and *NEUROD1* showed combined accessibility at the two sets of regions (Suppl. Fig. 2d). This result underlines the striking similarity in the chromatin state of the tumor subtypes both in SCLC and NEPC. It is important to note that despite the clear differences in accessibility associated to the subtypes, still the large majority of open chromatin sites are shared between these two subtypes as expected given the NE characteristics in common for both subtypes (Suppl. Fig. 2e).

To further characterize the chromatin differences between the subtypes we investigated the relationship between TF-subtype specific chromatin accessibility and the ASCL1 and NEUROD1 genomic binding. To that aim, we performed ChIP-seq analysis for each of the two TFs in NEPC models that expressed *ASCL1* or *NEUROD1* (Table 3). This analysis identified thousands of highly conserved binding sites with both overlapping and differential sites for each of the two TFs (Suppl. Fig. 2f) that were enriched for the expected consensus motifs for ASCL1 and NEUROD1 respectively (Suppl. Fig 2g). Importantly, TF-specific regions of differential chromatin accessibility were bound by the corresponding TF, but not the other (Figure 2b). This result is consistent with a role for ASCL1 and NEUROD1 in maintaining the chromatin state in their respective subtypes.

We next sought to confirm that the enhancers specific to the ASCL1 and NEUROD1 subtypes are associated with expression of nearby genes by using the expression data generated from the same samples (Figure 2c). As expected, differential expression analysis showed that *ASCL1* was one of the most upregulated genes in the ASCL1 set, while conversely, the NEUROD1 set showed upregulation of several NEUROD family members including *NEUROD1/2/4/6* (Figure 2c). Consistent with these differentially accessible regions being functional, we observed a substantial association between differential DNA accessibility and differential gene expression (Figure 2c). Gene Set Enrichment Analysis (GSEA) to identify pathways differentially over-represented in each of the two subtypes showed that ASCL1 associated gene expression was enriched in GO pathways of response to cytokines[37] while the NEUROD1 associated expression was enriched in brain development pathways (Figure 2d).

The binding of ASCL1 and NEUROD1 TFs to their own promoters and nearby enhancers suggests they are working as lineage TFs (LTFs) in NEPC. LTFs are known to auto-activate their own expression by binding to super-enhancers (SE) establishing a positive feedback loop. In addition, LTFs form circuits of core TFs driven by the activation of super-enhancers (SE) promoting the transcriptional program required to maintain the lineage[38]. Both ASCL1 and NEUROD1 are known to be lineage transcriptional factors in neuronal systems[39,40]. To investigate a potential LTF behavior of both TFs in NEPC, we performed SE analysis by H3K27ac profiling of ASCL1 and NEUROD1 NEPCs and found that all models showed SE activation in common at a number of TFs (*INSM1* and *NFIB*) regardless of the tumor subtype. In addition, we found differential SEs at either *ASCL1* or *NEUROD1* (and other family members) in accordance with their expression status (Figure 2e). Based on those characteristics both ASCL1 and NEUROD1 can be considered as LTFs in NEPC with binding to SEs and activation of their own expression (Figure 2e, 2f). We next identified the core circuit of TFs associated with each of the two subtypes applying a previously described method to identify interconnected auto-regulated loops[38]. We identified distinct but highly overlapping sets of TF circuits in these two subtypes (Figure 2g).

Taken together, our results provide clear evidence of the existence of two molecular subtypes in NEPC model systems. These subtypes share NE phenotypic characteristics but differ in the expression of *ASCL1* and *NEUROD1*, which is associated with distinct chromatin landscapes and gene expression profiles.

Analysis of tumor heterogeneity in NEPC liver metastases

Next we aimed to determine if the results from the model systems can be extended to human clinical NEPC. First we interrogated expression levels of *ASCL1* and *NEUROD1* in tumor tissues from two cohorts of NEPC metastases[22,25]. In contrast to the mutually exclusive expression of the two TFs that we observed in NEPC PDXs (Figure 1g), clinical samples showed a range of coexpression. The *ASCL1* expression was higher in the majority of the metastases accompanied by a lower and more variable expression of *NEUROD1* for almost all the cases (Figure 3a, Suppl. Fig 3a).

To investigate whether these TFs are co-expressed in the same tumor cells or in distinct tumor sub-populations, we studied five distinct fragments of liver metastasis (FLM) obtained at autopsy from a patient diagnosed with NEPC available as both OCT frozen and formalin-fixed paraffin-embedded material (Suppl. Fig 3b) and performed RNA-seq to assess expression levels of *ASCL1* and *NEUROD1*. We observed a range of coexpression of the two TFs with FLM3 showing the highest relative expression of *NEUROD1* to *ASCL1* (Suppl. Fig 3c). We next performed immunohistochemical (IHC) analysis for ASCL1 and NEUROD1 protein expression on FLM3. The staining showed clear intratumoral heterogeneity in terms of ASCL1 and NEUROD1 staining that defined two separated tumor populations (Figure 3b). Correlation with histomorphological features showed that the two distinct cell populations also differ in their histological characteristics. ASCL1 positive cells had a sheet-like growth pattern and spindle cell morphology, whereas NEUROD1 positive cells appeared to grow in smaller cell clusters with pronounced nuclear molding and focal pleomorphic giant cells (Figure 3c). We next performed double staining of ASCL1 and NEUROD1 by immunofluorescence in FLM3 to investigate potential co-expression in tumor cells and observed that the vast majority of the cells showed an anticorrelated expression of the two TFs (Suppl. Fig 3d). We extended this analysis by investigation of two additional NEPCs from different patients by IHC and also observed the existence of this type of intratumor heterogeneity (Suppl. Fig 3e).

We next investigated the two observed intra-tumoral populations by single cell chromatin (scATAC-seq) and expression (snRNA-seq) analysis. We selected FLM3 that showed the highest NEUROD1 expression and FLM5 that had the lowest, almost 200-fold lower than *ASCL1* (Suppl. Fig 3c). We isolated nuclei from frozen sections of FLM3 and performed scATAC-seq and snRNA-seq to assign the *ASCL1* and *NEUROD1* expression with the corresponding chromatin state. The

unsupervised tSNE clustering of the scATAC-seq resulted in multiple clusters that we analyzed for differential accessibility at *SOX2* promoter to distinguish tumor and normal cells. Based on accessibility to *SOX2* the fraction of the tumor cells represented around 80% (Supp. Fig. 3f). Notably, we could distinguish the clusters that correspond to the two tumor subtypes based on the differential accessibility to the *ASCL1* and *NEUROD1* promoters (Suppl. Fig. 3f). The ASCL1 and NEUROD1 clusters also show differential accessibility at the top ATAC differential regions identified by bulk analysis (Suppl. Fig. 3g). A small number of cells displayed chromatin accessibility at both differential regions and the *ASCL1* and *NEUROD1* promoters, that we labelled as "dual" (Suppl. Fig. 3f,g). We next analyzed the snRNA-seq to identify the tumor cells that express *ASCL1* and *NEUROD1* and then integrated this dataset with the scATAC-seq using SEURAT[41] (Figure 3d). This integration enabled the assignment of normal cells based on the expression of specific markers. Crucially, we observed that cells with either the ASCL1 or NEUROD1 accessibility signature developed from bulk data preferentially express the corresponding TF (Figure 3d). Thus, the integrated single cell analysis clearly shows that the ASCL1 and NEUROD1 sub-types exist as separate subpopulations possessing similar epigenetic features as in their respective model systems.

We next investigated the FLM5 sample, which has the highest expression of *ASCL1* by scATAC-seq analysis. In accordance with the RNA-seq, the t-SNE analysis showed a single cluster of the FLM5 tumor cells with accessibility at *ASCL1* promoter but not at *NEUROD1* (Suppl. Fig. 3h). The integrated scATAC analysis of FLM3 and FLM5 revealed that 99% of the FLM5 tumor cells overlap the FLM3 ASCL1 cluster (Figure 3e) indicating that those cells have identical chromatin accessibility.

The identification of the two coexisting subtypes in NEPC metastases motivated the further investigation in the PDX model derived from this patient (Suppl. Fig 3b). This PDX model was characterized by RNA-seq analysis of multiple samples from passage 1 (p1) up to p17. The system showed dramatic changes in the relative expression of *ASCL1* and *NEUROD1* across passages, evolving from high *NEUROD1* expression and barely detectable expression of *ASCL1* at p1, to a relative co-expression at p3, to high *ASCL1* expression and extremely low *NEUROD1* expression from p5 onwards (Suppl. Fig 3i). We selected samples from p1, p3 and p5 for analysis by scATAC-seq analysis and performed the trajectory inference of the tumor cell evolution by analysis of the scATAC-seq results using Monocle[42] (Figure 3f). The tSNE analysis of the scATAC-seq data revealed homogenous populations at p1 and p5 consistent with the NEUROD1

and ASCL1 sub-types respectively, which was also validated by IHC (Suppl. Fig 3j), while p3 showed coexistence of the sub-type chromatin accessibility patterns (Figure 3f and 3g). Therefore, the inferred trajectory ordered the tumor cells along a continuum based on DNA accessibility with the pure ASCL1 and NEUROD1 tumor populations as the two edges of the trajectory (Figure 3f). Plotting the aggregated scATAC-seq by TF cluster from FLMs and PDXs in the PCA space defined by the model systems in Figure 2a further illustrates the PDX changes from NEUROD1 to ASCL1 while also validating the NEPC model systems we have used (Figure 3h). All together, these results demonstrate subtype heterogeneity in human NEPC metastases and that these subtypes show distinct epigenetic characteristics and a divergent dynamic behavior.

The NEPC sub-types are distinct but related clones in the patient metastasis

We next sought to investigate the genetic characteristics of the NEC samples using whole exome sequencing (WES) and copy number variation (CNV) inferred from the ATAC-seq data[43]. Inference of *RB1* genetic status from the bulk ATAC-seq data showed biallelic loss in all the NEPC PDX models but not in the ADPCs (Suppl. Fig. 4a) as previously reported for these NEPC models[20]. The same approach was applied genome-wide to the scATAC-seq clusters identified by the tSNE analysis on FLM3 and FLM5. The results show an overall similarity in the CNVs across these clusters (Figure 4a). For instance, we observed heterozygous losses in all of chr16 and parts of chr2 and chr13 in both ASCL1 and NEUROD1 clusters. In addition, we found a focal heterozygous loss at *PTEN* on chr10 in both clusters. However, clear CNV differences existed, including a 20MB amplification on chr14p and a chr7p amplification that are only present in the NEUROD1 cluster. Notably, CNVs of the ASCL1 component in FLM3 showed almost identical characteristics with the ASCL1 cluster in FLM5 (Pearson correlation = 0.97) while showing lower correlation to the NEUROD1 cluster within the same fragment (Pearson correlation = 0.81) (Figure 4b). These same CNV alterations along with others were observed in PDX passages. Specifically, p1 and a p3 cluster had CNVs that were most similar to the NEUROD1 profile, while the other p3 cluster and the p5 were most similar to the ASLC1 profile (Figure 4c, Suppl. Fig. 4b). This result supports the hypothesis that the dynamic changes in the PDX model are due to the preferential outgrowth of the ASCL1 clone that started at a very low fraction in the initial passage. WES analysis of the FLM samples, though derived from bulk tissue, validated the scATAC-seq inferred CNV alterations including amplifications on chr14p and chr7p (Suppl. Fig 4c).

Finally, we extended the CNV analysis of FLM3 to the single cell level following the method of Satpathy et al.[43,44]. The K-means clustering of the cells based on the CNVs distinguished normal cells with no alterations from two additional clusters, one with the chr7p and chr14p amplifications corresponding to the NEUROD1 tumor, and one without these alterations corresponding to the ASCL1 tumor (Figure 4d). We next marked the identity of the cells from the three clusters defined in the genetic analysis within the scATAC-seq tSNE plot from this sample and showed a strong correspondence to the groupings defined by the epigenetic analysis (Figure 4e). Importantly, this single cell resolution enabled a more refined interpretation of the dual accessibility initially identified in the analysis of the scATAC-seq results. The CNV inference of that cluster showed that it is composed of a mix of cells belonging to either ASCL1 or NEUROD1 clones (Suppl. Fig. 4d) and is not an independent clone. All together our results show the existence of distinct genetic clones associated with each of the two NEPC epigenetic subtypes in this patient, likely derived from a common ancestor given their substantial CNV profile overlap.

Discussion

Poorly differentiated NECs are a class of high-grade tumors that arise at different anatomical sites and typically express markers of neuroendocrine differentiation (CHGA, NCAM1, and SYP). Our results build considerably on previous work with RNA-seq and cell lines[19,45] and provide a molecular rationale for the shared histopathological behavior of these tumors based on a common epigenetic state regardless of anatomic origins or the distinct tumor-initiation mechanisms. This epigenetic convergence is associated with the expression of bHLH TFs known for specifying neural lineages. Notably, distinct members of the bHLH family are expressed by the different NECs, suggesting that a variety of TFs can maintain the NE fate.

The similarity in the chromatin state across NECs is particularly pronounced between NEPC and SCLC, which is surprising given the distinct cells of origin in these neoplasms. Although SCLC can arise as a treatment emergent phenotype, it mostly arises *de novo* and is thought to originate from resident neuroendocrine cells of the lung known to express either ASCL1 or NEUROD1[13,14,15]. In contrast, NEPC is primarily observed as a treatment emergent tumor that arises through a process of transdifferentiation from a preexisting adenocarcinoma[22]. Thus, despite the initial adenocarcinomas having a radically different chromatin state, the pattern of DNA accessibility observed in the resulting NEPC becomes almost indistinguishable from that in SCLC. More unexpected was our observation that treatment emergent NEPC also shows

subtypes based on the expression of *ASCL1* and *NEUROD1* as in SCLC, since those TFs have been previously associated with lung development. Despite both subtypes having the NE phenotype, with expression of *SYP*, *CHGA* and *INSM1* and exhibiting largely similar chromatin states, they still differ in the activation of specific sets of enhancer sites and transcriptional programs.

Importantly, we show a fundamental difference between the representation of the subtypes in the PDX models as compared to human clinical samples. In contrast to the mutually exclusive expression of *ASCL1* and *NEUROD1* in model systems, tissues from NEPC clinical cohorts show coexpression of *ASCL1* and *NEUROD1* at varying levels. Single cell analyses of a set of metastatic samples from the same patient revealed the presence of two distinct tumor populations that coexist within the metastasis. This observation emphasizes that PDXs, despite being good models of the human disease, still offer limitations to illustrate the complexity observed in primary tissues. In fact, those limitations could have precluded a better characterization of subtype coexistence in SCLC[46] that has mainly been described as homogeneous subtypes[12]. Our results show the existence of subtypes in clinical samples of NEPC and demonstrate heterogeneity in terms of the chromatin state.

CNV inference from scATAC-seq clusters demonstrated the existence of distinct clones associated with the two NEPC epigenetic subtypes. Based on our results we propose that the ASCL1 clone originated by transdifferentiation from an adenocarcinoma cell and there was subsequently subclonal evolution to the NEUROD1 state (Suppl. Fig 4e). This is supported by recent experiments that show tumors in Myc-driven SCLC GEMMs that start as pure *ASCL1* can develop *NEUROD1* expression both *in vitro* and *in vivo*[46]. Both NEPC clones then seed the liver metastases with the earlier ASCL1 clone being more common. It has been observed in CRPC studies that metastasis-to-metastasis seeding can occur and that may be contributing here as well[47]. However, we cannot rule out an alternative hypothesis where separate transdifferentiation events occur for the ASCL1 and NEUROD1 clones from the preexisting adenocarcinoma.

The intra-tumor heterogeneity in NEPC that we have observed here has direct clinical implications. In SCLC these subtypes have been linked to specific drug targets or predictors of drug response. For example, DLL3 is a target for an antibody-drug conjugate in the ASCL1 subtype[48,49] and AURKA is a target for small molecule inhibitors such as alisertib in the NEUROD1 subtype[50] . With this new understanding of sub-type heterogeneity based on NEUROD1 and

ASCL1 in NEPC, therapeutic strategies that target one but not the other will rapidly succumb to outgrowth of the resistant subpopulation. Altogether, our results illustrate the epigenetic complexity that exists in clinical tumors and provide a rational basis for targeting the inter and intra tumoral heterogeneity as a therapeutic strategy in NEPC.

ACKNOWLEDGEMENTS

REFERENCES

1. Oronsky, B., Ma, P. C., Morgensztern, D. & Carter, C. A. Nothing But NET: A Review of Neuroendocrine Tumors and Carcinomas. *Neoplasia* vol. 19 991–1002 (2017).

2. Klimstra, D. S., Modlin, I. R., Coppola, D., Lloyd, R. V. & Suster, S. The pathologic classification of neuroendocrine tumors: a review of nomenclature, grading, and staging systems. *Pancreas* **39**, 707–712 (2010).

3. Oronsky, B., Ma, P. C., Morgensztern, D. & Carter, C. A. Nothing But NET: A Review of Neuroendocrine Tumors and Carcinomas. *Neoplasia* **19**, 991–1002 (2017).

4. Rickman, D. S., Beltran, H., Demichelis, F. & Rubin, M. A. Biology and evolution of poorly differentiated neuroendocrine tumors. *Nat. Med.* **23**, 1–10 (2017).

5. Cheng, J. *et al.* Merkel cell polyomavirus recruits MYCL to the EP400 complex to promote oncogenesis. *PLoS Pathog.* **13**, e1006668 (2017).

6. Harms, P. W. *et al.* The Distinctive Mutational Spectra of Polyomavirus-Negative Merkel

Cell Carcinoma. *Cancer Res.* **75**, 3720–3727 (2015).

7. Aggarwal, R. *et al.* Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. *J. Clin. Oncol.* **36**, 2492–2503 (2018).

8. Rath, B., Plangger, A. & Hamilton, G. Non-small cell lung cancer-small cell lung cancer transformation as mechanism of resistance to tyrosine kinase inhibitors in lung cancer. *Cancer Drug Resistance* (2020) doi:10.20517/cdr.2019.85.

9. Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3**, 75ra26 (2011).

10. Oser, M. G., Niederst, M. J., Sequist, L. V. & Engelman, J. A. Transformation from non-small-cell lung cancer to small-cell lung cancer: molecular drivers and cells of origin. *Lancet Oncol.* **16**, e165–72 (2015).

11. Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3**, 75ra26 (2011).

12. Rudin, C. M. *et al.* Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat. Rev. Cancer* **19**, 289–297 (2019).

13. Neptune, E. R. *et al.* Targeted disruption of NeuroD, a proneural basic helix-loop-helix factor, impairs distal lung formation and neuroendocrine morphology in the neonatal lung. *J. Biol. Chem.* **283**, 21160–21169 (2008).

14. Borges, M. *et al.* An achaete-scute homologue essential for neuroendocrine differentiation in the lung. *Nature* **386**, 852–855 (1997).

15. Ito, T. *et al.* Basic helix-loop-helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. *Development* **127**, 3913–3921 (2000).

16. Borromeo, M. D. *et al.* ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. *Cell Rep.* **16**, 1259–1272 (2016).

17. Aggarwal, R. *et al.* Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. *J. Clin. Oncol.* **36**, 2492–2503 (2018).

18. Davies, A. H., Beltran, H. & Zoubeidi, A. Cellular plasticity and the neuroendocrine phenotype in prostate cancer. *Nat. Rev. Urol.* **15**, 271–286 (2018).

19. Balanis, N. G. *et al.* Pan-cancer Convergence to a Small-Cell Neuroendocrine Phenotype that Shares Susceptibilities with Hematological Malignancies. *Cancer Cell* **36**, 17–34.e7 (2019).

20. Nguyen, H. M. *et al.* LuCaP Prostate Cancer Patient-Derived Xenografts Reflect the Molecular Heterogeneity of Advanced Disease an--d Serve as Models for Evaluating Cancer Therapeutics. *Prostate* **77**, 654–671 (2017).

21. Drapkin, B. J. *et al.* Genomic and Functional Fidelity of Small Cell Lung Cancer Patient-Derived Xenografts. *Cancer Discov.* **8**, 600–615 (2018).

22. Beltran, H. *et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).

23. Oser, M. G., Niederst, M. J., Sequist, L. V. & Engelman, J. A. Transformation from non-small-cell lung cancer to small-cell lung cancer: molecular drivers and cells of origin. *Lancet Oncol.* **16**, e165–72 (2015).

24. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).

25. Labrecque, M. P. *et al.* Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J. Clin. Invest.* **129**, 4492–4505 (2019).

26. Pomerantz, M. M. *et al.* Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat. Genet.* **52**, 790–799 (2020).

27. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions.

Nat. Biotechnol. **28**, 495–501 (2010).

28. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

29. Fan, K. *et al.* MCPyV Large T Antigen-Induced Atonal Homolog 1 Is a Lineage-Dependency Oncogene in Merkel Cell Carcinoma. *Journal of Investigative Dermatology* vol. 140 56–65.e3 (2020).

30. Park, D. E. *et al.* Author Correction: Merkel cell polyomavirus activates LSD1-mediated blockade of non-canonical BAF to regulate transformation and tumorigenesis. *Nat. Cell Biol.* **22**, 752 (2020).

31. Augustyn, A. *et al.* ASCL1 is a lineage oncogene providing therapeutic targets for high-grade neuroendocrine lung cancers. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14788–14793 (2014).

32. Denny, S. K. *et al.* Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility. *Cell* **166**, 328–342 (2016).

33. Rudin, C. M. *et al.* Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat. Genet.* **44**, 1111–1116 (2012).

34. Horie, M. *et al.* An integrative transcriptome analysis reveals a functional role for thyroid transcription factor-1 in small cell lung cancer. *J. Pathol.* **246**, 154–165 (2018).

35. Zheng, R. *et al.* Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* **47**, D729–D735 (2019).

36. Layer, R. M. *et al.* GIGGLE: a search engine for large-scale integrated genome analysis. *Nat. Methods* **15**, 123–126 (2018).

37. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

38. Saint-André, V. *et al.* Models of human core transcriptional regulatory circuitries. *Genome

Res. **26**, 385–396 (2016).

39. Wapinski, O. L. *et al.* Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. *Cell Rep.* **20**, 3236–3247 (2017).

40. Matsushita, M. *et al.* Neural differentiation of human embryonic stem cells induced by the transgene-mediated overexpression of single transcription factors. *Biochem. Biophys. Res. Commun.* **490**, 296–301 (2017).

41. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

42. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

43. Kuilman, T. *et al.* CopywriteR: DNA copy number detection from off-target sequence data. *Genome Biol.* **16**, 49 (2015).

44. Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).

45. Park, J. W. *et al.* Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. *Science* **362**, 91–95 (2018).

46. Ireland, A. S. *et al.* MYC Drives Temporal Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate. *Cancer Cell* **38**, 60–78.e12 (2020).

47. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).

48. Saunders, L. R. *et al.* A DLL3-targeted antibody-drug conjugate eradicates high-grade pulmonary neuroendocrine tumor-initiating cells in vivo. *Sci. Transl. Med.* **7**, 302ra136 (2015).

49. Dylla, S. J. Toppling high-grade pulmonary neuroendocrine tumors with a DLL3-targeted trojan horse. *Molecular & Cellular Oncology* vol. 3 e1101515 (2016).

50. Mollaoglu, G. *et al.* MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. *Cancer Cell* **31**, 270–285 (2017).

51. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* vol. 33 2938–2940 (2017).

52. Corces, M. R. *et al.* Omni-ATAC-seq: Improved ATAC-seq protocol. *Protocol Exchange* (2017) doi:10.1038/protex.2017.096.

53. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

54. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

55. Qin, Q. *et al.* ChiLin: a comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. *BMC Bioinformatics* vol. 17 (2016).

56. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

57. Ji, X., Li, W., Song, J., Wei, L. & Liu, X. S. CEAS: cis-regulatory element annotation system. *Nucleic Acids Res.* **34**, W551–4 (2006).

58. Cornwell, M. *et al.* VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. *BMC Bioinformatics* **19**, 135 (2018).

59. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

60. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* vol. 36 411–420 (2018).

61. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).

62. Cibulskis, K. *et al.* ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* vol. 27 2601–2602 (2011).

63. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* vol. 31 213–219 (2013).

64. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).

65. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *Bioinformatics* (2019).

66. Taylor-Weiner, A. *et al.* DeTiN: overcoming tumor-in-normal contamination. *Nat. Methods* **15**, 531–534 (2018).

67. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

68. Jo, S.-Y., Kim, E. & Kim, S. Impact of mouse contamination in genomic profiling of patient-derived models and best practice for robust analysis. *Genome Biol.* **20**, 231 (2019).

Methods

Clinical samples.

Tissue samples were collected within 8 hours of death from patients who died of metastatic CRPC. All patients signed informed consent for a rapid autopsy, under the aegis of the Prostate Cancer Donor Program at the University of Washington. Hematoxylin and eosin-stained slides from each case were reviewed by a pathologist to confirm the presence of tumor cells. The Institutional Review Board of the University of Washington (IRB #2341) approved this study.

Nuclei preparation

Fragments of frozen tissues (PDX models) or 50 um sections (liver metastases) were cut and resuspended in 300 ul of cold 3-detergent-ATAC-Resuspension Buffer (RSB) containing 0.1% NP40, 0.1% Tween-20, and 0.01% Digitonin. Tissues were dounced 10 times each with a loose and a tight pestle each until homogenization was complete. The homogenate was then transferred

to a 1.5 ml pre-chilled microfuge tube and incubated on ice for 10 minutes. After lysis, 300 ul of ATAC-RSB containing 0.1% Tween-20 was added and the tubes were inverted to mix. Lysates were filtered through a 40 um cell strainer and nuclei were centrifuged for 10 minutes at 1500 RCF in a pre-chilled (4 ºC) fixed-angle centrifuge. Nuclei were resuspended with 300 ul of ATAC-RSB containing 0.1% Tween-20 and counted with a hemocytometer using Trypan blue stain.

ATAC-seq

100,000 nuclei were resuspended in 50 ul of transposition mix (25 ul 2x TD buffer, 2.5 ul transposase (100 nM final), 16.5 ul PBS, 0.5 ul 1% Digitonin, 0.5 ul 10% Tween-20, 5 ul H2O][52]. Transposition reactions were incubated at 37 ºC and shaken at 1000 RPM for 30 minutes on a thermomixer. Transposed DNA was purified using Qiagen columns. Libraries were amplified as described previously[53]. 35-bp paired-end reads were sequenced on a NextSeq instrument (Illumina).

ChIP-sequencing

Nuclei isolated as previously described were crosslinked with 1% Formaldehyde for 10 minutes for H3K27Ac ChIP-seq. For ASCL1 and NEUROD1 ChIP-seq nuclei were crosslinked in 2 steps with 2 mM of DSG (Pierce) for 45 minutes at room temperature followed by 1 ml of 1% Formaldehyde for 10 minutes. Crosslinked nuclei were then quenched with 0.125 M glycine for 5 minutes at room temperature and washed with PBS. After fixation, pellets were resuspended in 500 ul of 1% SDS (50 mM Tris-HCl pH 8, 10 mM EDTA) and sonicated for 5 (H3K27ac) or 10 (ASCL1 and NEUROD1) minutes using a Covaris E220 instrument (setting: 140 peak incident power, 5% duty factor and 200 cycles per burst) in 1 ml AFA fiber millitubes. Chromatin was immunoprecipitated with 10 ug of H3K27Ac antibody (Diagenode cat# C15410196), 10 ug of ASCL1 antibody (abcam ab74065) or 10 ug of NEUROD1 antibody (Cell Signaling mAb #4373). 5 ug of chromatin was used for H3K27Ac ChIPs, and 40 ug of chromatin was used for ASCL1 or NEUROD1 ChIPs. ChIP-seq libraries were made using Rubicon kit and purified. 75-bp single-end reads were sequenced on a Nextseq instrument (Illumina).

Single nuclei ATAC-seq and RNA-seq

Nuclei were prepared as described previously. For scATAC-seq nuclei were transposed according to the OMNI-ATAC protocol[52]. ~7,000 cells were targeted for each sample and processed according to the 10x Genomics scATAC-seq sample preparation protocol (Chromium Single Cell ATAC Library & Gel Bead Kit, 10x Genomics). For snRNA-seq, nuclei prepared the

same way were used directly in the 10x Genomics snRNA-seq protocol (Chromium Single Cell 3' v2 Reagent Kit, 10xGenomics).

RNA-seq

A fragment of frozen tissues (PDX models) or 50 um sections (liver metastases) were cut and homogenized in 1 ml of AllPrep DNA/RNA Mini Kit (Qiagen) using a plastic pestle (Cole-Palmer #44468-23). DNA and RNA were simultaneously isolated. 500 ng RNA was used to prepare libraries using the NEBNext. Ultra™ RNA Library Prep Kit for Illumina. RNA quantity and quality were assessed on an Agilent 2100 Bioanalyzer. For all RNA-seq, reads were sequenced on a NextSeq 500 instrument (Illumina).

Whole exome sequencing

DNA extraction on frozen PDXs, human FLMs and adjacent normal tissue was performed using the AllPrep DNA/RNA Mini Kit (Qiagen). Whole exome sequencing was performed by Novogene using their standard protocols. Briefly, 1000 ng of genomic DNA were used as input to generate sequencing libraries using the Agilent SureSelect Human All Exon Kit. Captured libraries were enriched by PCR, purified, quantified using the Agilent Bioanalyzer 2100 system, and subsequently sequenced using the NextSeq 500 instrument (Illumina).

IHC

Immunohistochemical and immunofluorescence studies using ASCL1 (clone 24B72D11.1, cat. no. 556604, BD biosciences, San Jose, CA) and NEUROD1 (clone EPR17084, cat. no. ab205300, Abcam, Cambridge, MA) specific antibodies were carried out on archival formalin fixed paraffin embedded tissues. In brief, 5-micron paraffin sections were de-waxed and rehydrated following standard protocols. Antigen retrieval consisted of steaming for 40 min in Target Retrieval Solution (S1700, Agilent, Santa Clara, CA). Slides were then washed and equilibrated in TBS-Tween buffer (Sigma, St. Louis, MO) for 10 min. Primary antibodies were applied at a dilution of 1:25 at 37C for 60 min. For chromogenic studies, immunocomplexes were visualized by applying secondary detection reagents of the UltraVision™ Quanto Detection System (cat. no. TL-060-QHD, Thermo Fisher, Waltham, MA) following manufacturer instructions. Sequential dual-immunofluorescence labeling studies were carried out using Tyramide SuperBoost kits (Thermo Fisher, Waltham, MA). All bright field slides were imaged using a Ventana DP200 system (Roche Diagnostics, Indianapolis, IN). Fluorescence images were acquired on a Cytation 5 Cell Imager (Biotek, Winooski, VT).

Computational and statistical analysis

Analysis of ATAC-seq and ChIP-seq data

A modified version of the ChiLin pipeline was used for quality control and pre-processing of the data[54,55]. We used Burrows-Wheeler Aligner (BWA Version: 0.7.17-r1188) as a read mapping tool to align to hg19 using default parameters. Unique reads for a position for peak calling were used to reduce false positive peaks, and statistically significant peaks were finally selected by calculating a false discovery rate (FDR) of reported peaks. ATAC peaks were called using MACS2 (v2.1.2) with a cut-off of FDR<0.01. H3K27ac, ASCL1 and NEUROD1 peaks were called using MACS2 using the same cut-off. DESeq2 was used to identify differential peaks in ATAC-seq and ChIP-seq, where gained or lost peaks were defined with the threshold of log2 fold change of 1 or 2 and an adjusted p-value<0.05. PCA was performed using princomp in R.

CEAS analysis is used to annotate resulting peaks with genome features. Cistrome Toolkit (dbtoolkit.cistrome.org) was used to probe which factors might regulate the user-defined genes. Genomic Regions Enrichment of Annotations Tool (GREAT) was used to annotate peaks with their biological functions. Conservation plots were obtained with the Conservation Plot (version 1.0.0) tool available in Cistrome[54,55]

Analysis of Super-enhancers

Bed files for H3K27ac peaks created by MACS2 were used as input to by ROSE[54] to call Super-enhancers (SEs) in H3K27ac ChIP-seq data.

Visualization of ChIP-seq and ATAC-seq data

Read depth normalized profiles corresponding to read coverage per 1 million reads were used for heatmaps and for visualization using the integrative genomics viewer (IGV)[56]. Heat maps were prepared using deepTools (version 2.5.4) and aggregation plots for ChIP-seq signals were generated using Sitepro in CEAS[57]. In the volcano plots, ATAC-seq peak summits were associated with the nearest TSS within a distance of +/- 50 kb, and incorporating DESeq2 output from RNA-seq, with the final plot generated using ggplot2 in R.

Analysis and Visualization of RNA-seq data

For RNA-seq data, read alignment, quality control and data analysis were performed using VIPER[58]. RNA-seq reads were mapped by STAR[59] to hg19 and read counts for each gene were

generated by Cufflinks. Differential gene expression analyses were performed on absolute gene counts for RNA-seq data using DESeq2.

Single cell ATAC-seq and RNA-seq

Single cell RNA-seq data generated by 10x Genomics were preprocessed using the Cell Ranger (https://www.10xgenomics.com/) to obtain the UMI (unique molecular identifier) counts for each gene. To get a reliable single cell transcriptome dataset, we excluded the cells with less than 200 genes expressed (UMI > 0) or the cells with more than 80% UMIs from mitochondrial genes. The filtered data was then normalized and scaled by using Seurat[60] to remove unwanted sources of variations[61]. t-SNE was performed on the normalized data to visualize the single cells in two-dimensional space by using the top 10 dimensions of principal component analysis (PCA). Unsupervised clustering was performed by using the "FindClusters" function in the Seurat package with parameter of resolution = 0.8. Cell cycle phases of all single cells were assigned by using the cyclone function in scran package [61]. Genes with differential expression between clusters were obtained by using Wilcoxon rank-sum test. FDR was then calculated to correct for multiple testing.

Single-cell ATAC-seq data were processed using the Cell Ranger ATAC pipeline v1.1.0, which provides QC and clustering. Any cell that had Fraction of reads in peaks (FRiP) < 0.2 or total fragments < 1,000 was removed from the analysis. The t-SNE analysis was performed using the implementation from the Loupe Cell Browser 3.1.0.

scATAC-seq and scRNA-seq data integration was performed by Seurat. The scATAC-seq peak matrix provided by 10x was loaded and collapsed to a "gene activity matrix". The processed data was then scaled and normalized. To help understand the internal structure of the ATAC-seq data, the "RunLSI" function was run. "FindTransferAnchors" function identified 'anchors' between the ATAC-seq and RNA-seq datasets, and finally ATAC-seq and RNA-seq data are able to be co-embedded in the same tSNE plot.

scCNV

By modifying an existing method used for bulk ATAC-seq data, we created a way to use off target scATAC-seq reads to infer DNA copy number amplifications. This approach first breaks the genome into many large intervals and finds the coverage of each window. The coverage of one

hundred GC-matched intervals are then averaged together as background. The coverage of each interval will be compared to each GC matched background to estimate CNV fold change. The size of each interval was set to 1-2 Mb to account for the sparsity of the scATAC-seq data with 'ChunkGRanges' function in GenomicRange. For each window, the 'GCcontent' function of biovizBase was used to calculate the percentage GC content. The coverage was compensated for removed peaks by using the effective window size in coverage calculation.

Whole exome sequencing

Reads were aligned using BWA v0.5.9 and somatic mutations called using a customized version of the Getz Lab CGA WES Characterization pipeline (https://portal.firecloud.org/#methods/getzlab/CGA_WES_Characterization_Pipeline_v0.1_Dec2 018/). We used ContEst[62] to estimate cross sample contamination, MuTect[63] v1.1.6 to call single nucleotide variants, and Strelka[64] v1.0.11 to call indels. MuTect2.1[65] was used to confirm Strelka indel calls. We applied DeTiN[66] to rescue true somatic variants that were removed due to tumor-in-normal contamination. Variant calls were filtered through a panel of normal samples to remove artifacts from miscalled germline alterations and other rare error modes. Variants were annotated using VEP, Oncotator, and vcf2maf v1.6.17 (https://github.com/mskcc/vcf2maf). Allelic copy number, tumor purity and ploidy were analyzed using ABSOLUTE[67].

Prior to characterizing somatic mutations and copy number profiles from patient derived xenograft samples, we removed potentially confounding mouse DNA sequences using ConcatRef[68]. Briefly, WES results were aligned to a concatenated hg19 and mm9 reference genome, and only reads for which both pairs uniquely aligned to just the hg19 reference sequences using BWA. The resultant high-confidence human paired end sequences were then used for downstream analysis as above.
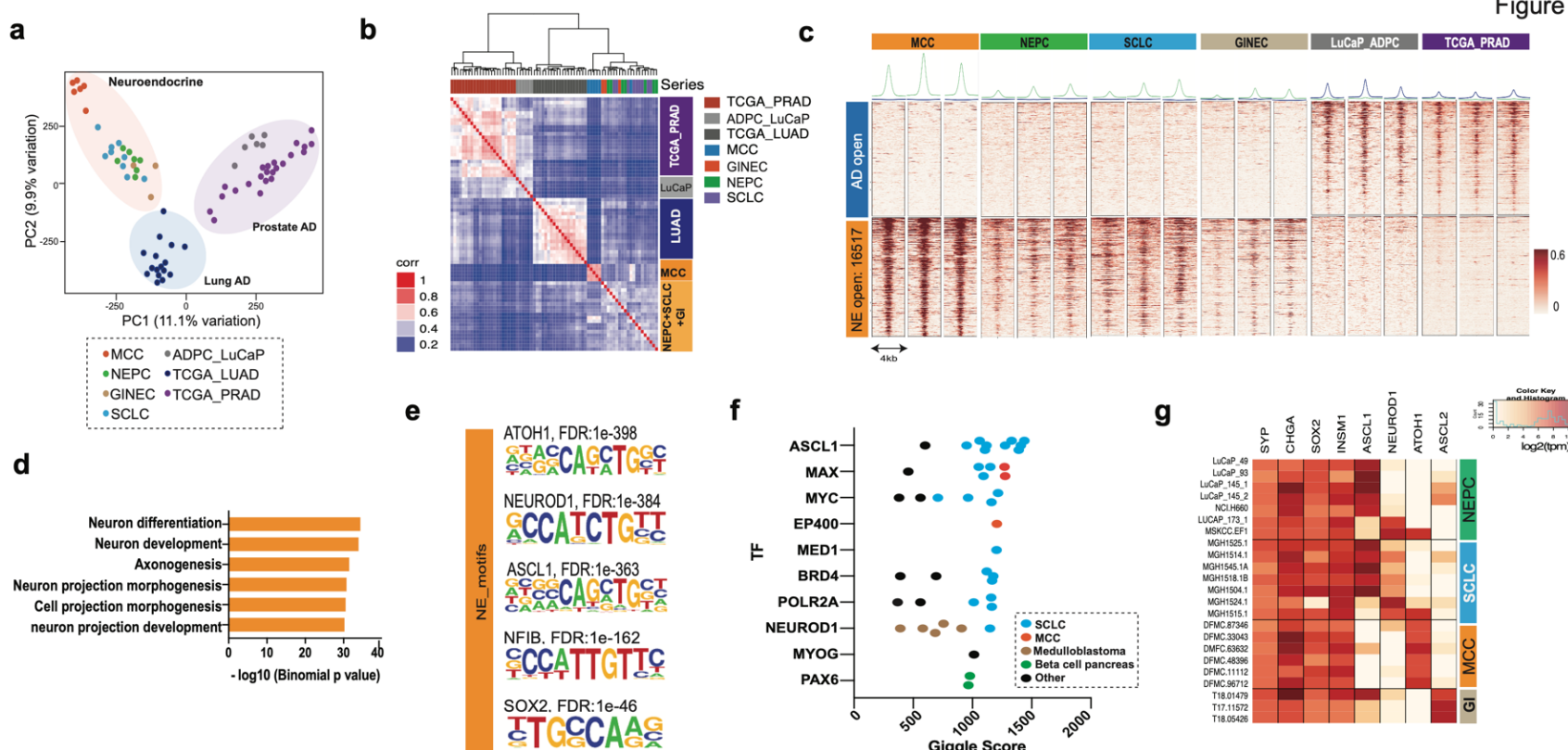
Figure 1. NE carcinomas share a common chromatin state independent of their anatomical origin. (a) PCA analysis of ATAC-seq data of NECs including Merkel Cell Carcinoma (MCC), Neuroendocrine prostate cancer (NEPC), Neuroendocrine gastrointestinal (GINE) and Small cell lung cancer (SCLC). The plot also includes Prostate adenocarcinoma (PDX models and TCGA primary tissues) and Lung adenocarcinoma (TCGA primary tissues). **(b)** Hierarchical clustering of the pairwise Pearson's correlation of the ATAC-seq signal across the distinct tumor types. **(c)** Heatmap representation of the differential regions between ADs and NECs. Each row is a peak location and each column is a sample. Shown above each column are the composite tag density plots for the AD sites (blue) and NE sites (green). **(d)** Gene Ontology pathways enriched in genes with nearby NE-specific accessible regions shown in (c). **(e)** Top results from motif analysis of the NE-specific accessible regions. **(f)** Public ChIP-seq datasets showing the highest overlap with the NE-specific accessible regions annotated by tissue type as determined by CistromeDB toolkit. The TFs are ordered by the top scoring dataset of each type. **(g)** Expression of NE markers and bHLH TFs across all the NEC samples in our study displayed as a heatmap.
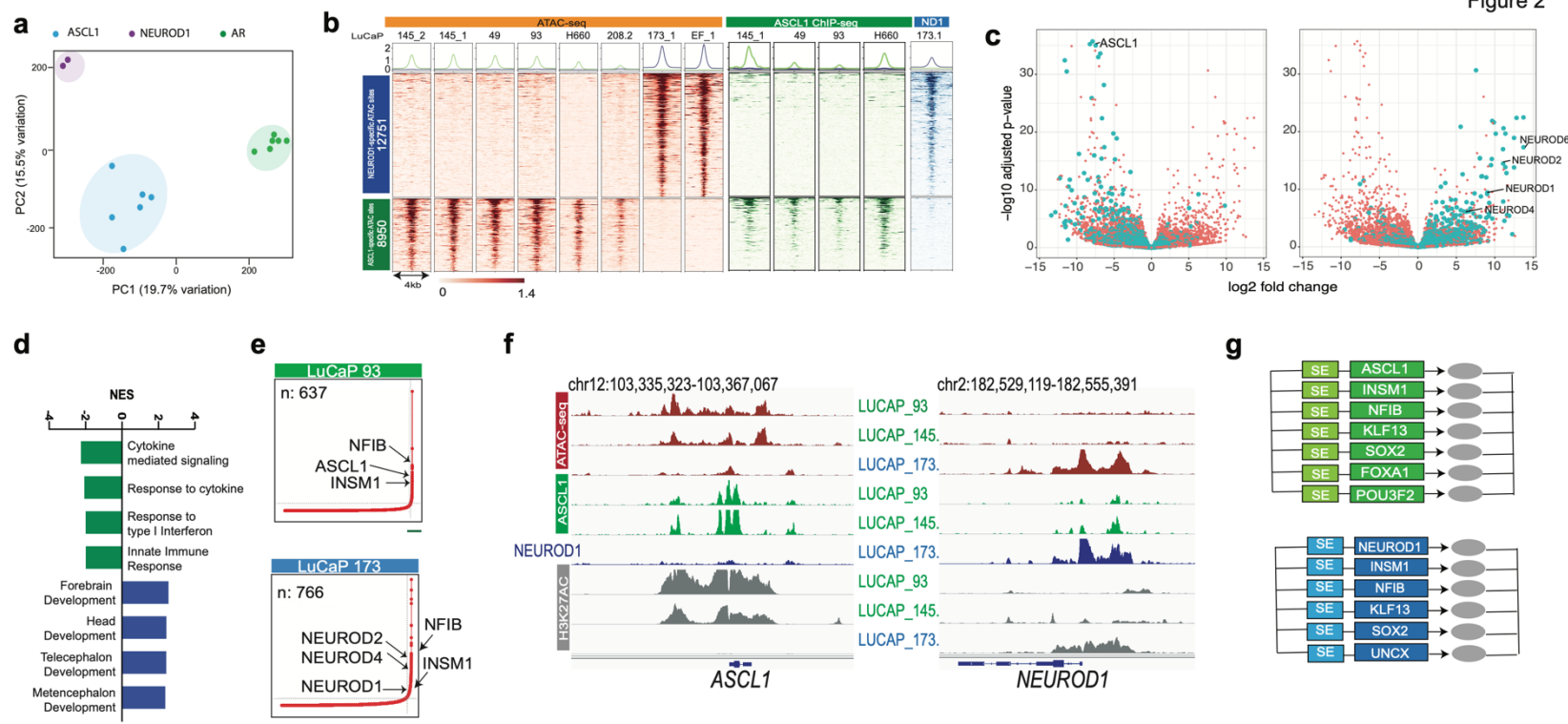
Figure 2. **NEPC shows tumor subtypes based on the differential expression of the transcription factors ASCL1 and NEUROD1. (a)** PCA analysis of NEPC and ADPC PDXs ATAC-seq data. Samples are color coded by the dominant TF expressed in that sample. **(b)** Heatmap representing the differential ATAC-seq regions between NEPC subtypes. ASCL1 ChIP-seq (green) and NEUROD1 ChIP-seq (blue) are shown for the indicated samples at the same sites. **(c)** Association between differential ATAC accessible sites and differential gene expression. Each volcano plot depicts RNA-seq log2 fold change and adjusted p-value calculated by DESeq2. Each dot represents one gene: green indicates a differential ATAC peak is with 50kB and orange indicates no such peak. ASCL1-specific ATAC accessible regions and genes upregulated in ASCL1 subtype (left) or NEUROD1-specific accessible regions and genes upregulated in NEUROD1 subtype (right). **(d)** GSEA pathway analysis enriched in ASCL1 subtype (green) and NEUROD1 subtype (blue). **(e)** Superenhancer analysis of a representative case of the ASCL1 subtype (top) and NEUROD1 subtype (bottom). **(f)** Representative IGV tracks at the *ASCL1* and *NEUROD1* gene. ATAC-seq tracks are in green, ASCL1 ChIP-seq in green, NEUROD1 ChIP-seq in blue and H3K27ac in gray. The loci are marked by subtype specific SEs with preferential binding of their respective TF. **(g)** Circuits of lineage transcription factors specific for the ASCL1 subtype (green) and NEUROD1 subtype (blue).
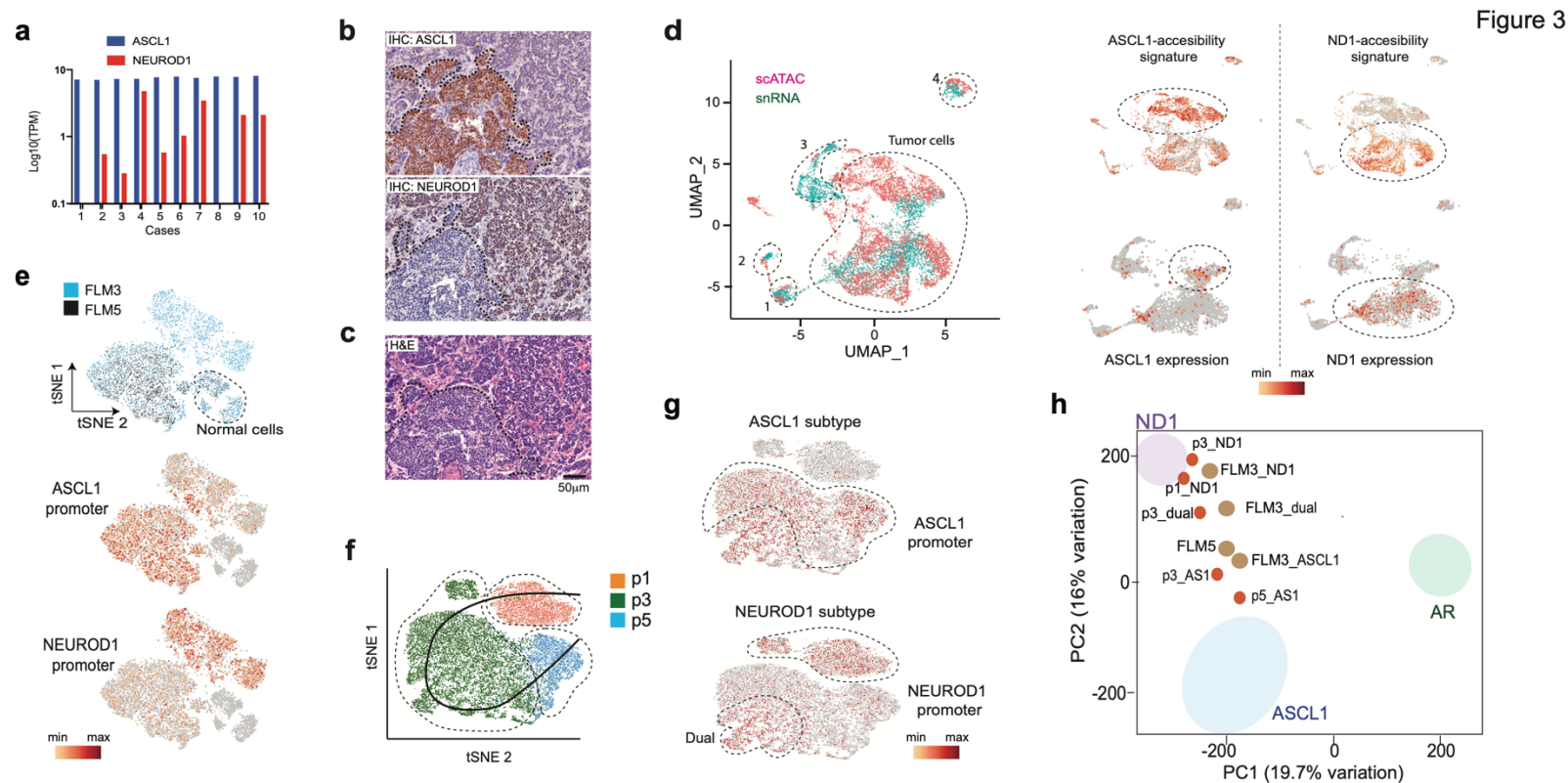
Figure 3. Single cell analysis reveals that NEPC sub-types co-exist in human metastasis and contribute to inter- and intra-tumoral heterogeneity. (**a**) Plot of *ASCL1* and *NEUROD1* expression in NEPC tissues from a clinical cohort (Labreque *et al*. 2019)[25]. TPM: Transcripts per million. (**b**) Immunohistochemical analysis of FLM3 (ASCL1 staining in top panel and NEUROD1 staining in the middle panel) showing intratumor heterogeneity. (**c**) Hematoxylin and eosin staining illustrating the distinct histologies for the two subpopulations. (**d**) Combined analysis of the scATAC-seq and snRNA-seq in FLM3 using SEURAT (left). Markers specific for normal cell populations enabled assignment of clusters: 1, vascular cells; 2, stromal cells; 3, hepatic cells; 4, macrophages. Accessibility at the top 30 differential ATAC-seq regions between ASCL1 and NEUROD1 subtypes identified by bulk analysis (top right). Analysis of *ASCL1* and *NEUROD1* expression in the snRNA-seq analysis (bottom right). This analysis matches cells with TF expression and the corresponding differential DNA accessibility for each subtype. (**e**) t-SNE analysis of the combined FLM3 (blue) and FLM5 (black) scATAC-seq data (top). The lower two plots show the differential accessibility at *ASCL1* promoter (middle) and *NEUROD1* promoter (bottom). (**f**) t-SNE plot of the combined PDX data showing the projected trajectory across all of the cells in the PDX analysis. (**g**) t-SNE plot of the combined PDX showing the differential accessibility of the combined PDX dataset at the *ASCL1* promoter indicating the ASCL1 subtype (top) and *NEUROD1* promoter showing the NEUROD1 subtype (bottom) the results also show a cluster with dual accessibility at *ASCL1* and *NEUROD1* promoters. (**h**) Projection of the scATAC-seq clusters for FLM3 and 5 (light brown dots) and PDX passes (orange dots) within the PCA space defined in Figure 2a.
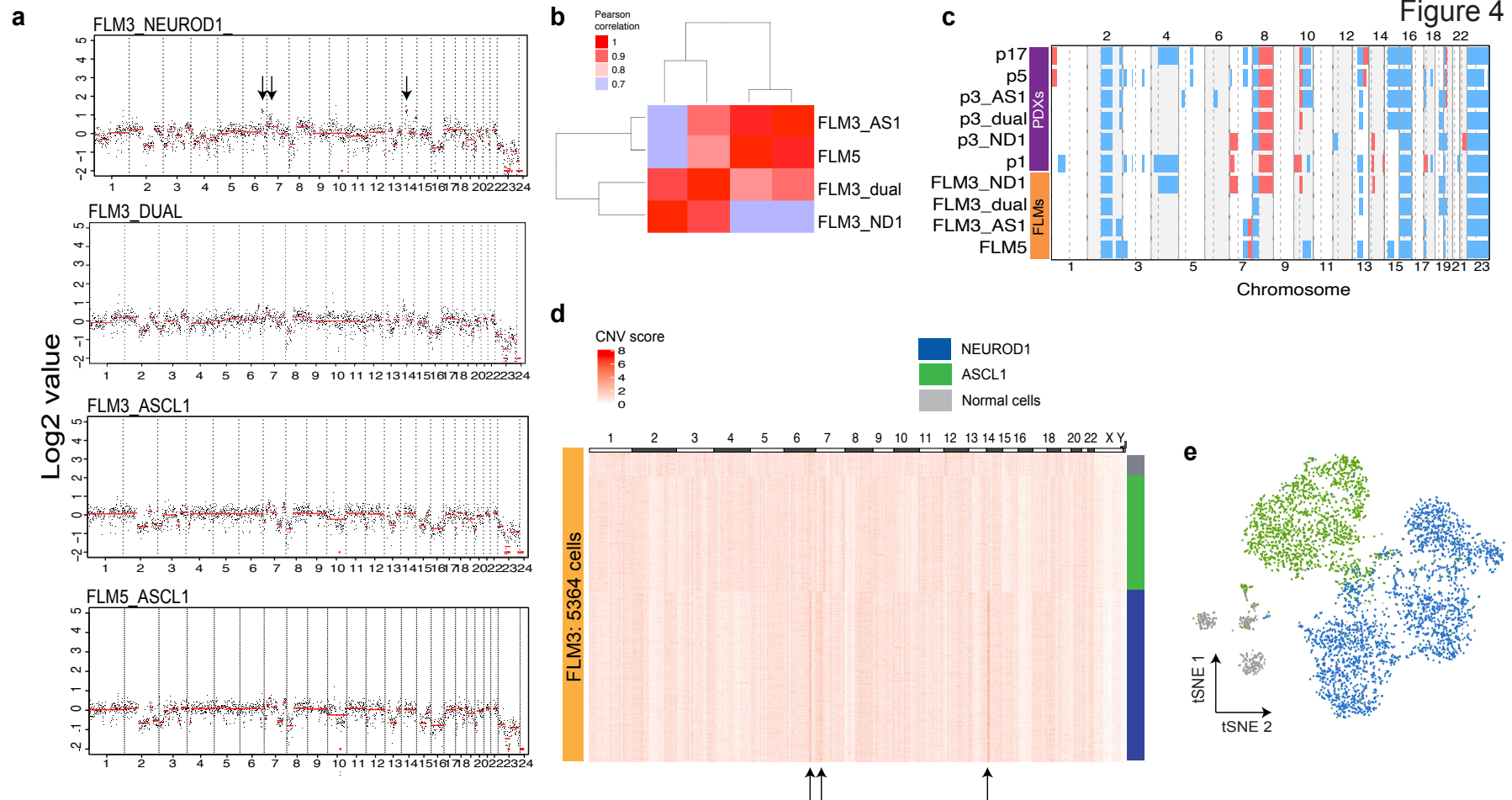
Figure 4. The NEPC sub-types are distinct clones. (a) Genome wide CNV profiles inferred from the scATAC-seq clusters in FLM3 and FLM5. Black dots are values in 1MB regions and the red line is the result of running a segmentation algorithm on the data (see Methods). Arrows point to differences seen in CNVs across the clusters. **(b)** Sample pairwise Pearson's correlation of the CNV profiles. **(c)** Summary heatmap of the scATAC-seq inferred CNV alterations across all of the patient samples and PDX models profiled (blue represents losses and red represents gains). **(d)** Heatmap of the single cell CNV analysis of FLM3 where each column is a 2MB bin tiled across the genome and the rows are individual cells that have been clustered with K-means. Arrows point to CNV differences observed here and in the cluster level analysis. **(e)** tSNE plot of FLM3 scATAC-seq data colored by the cluster each cell was partioned into by the inferred CNV alterations. Those three clusters clearly correspond to NEUROD1 (blue), ASCL1 (green) and normal cells (gray).
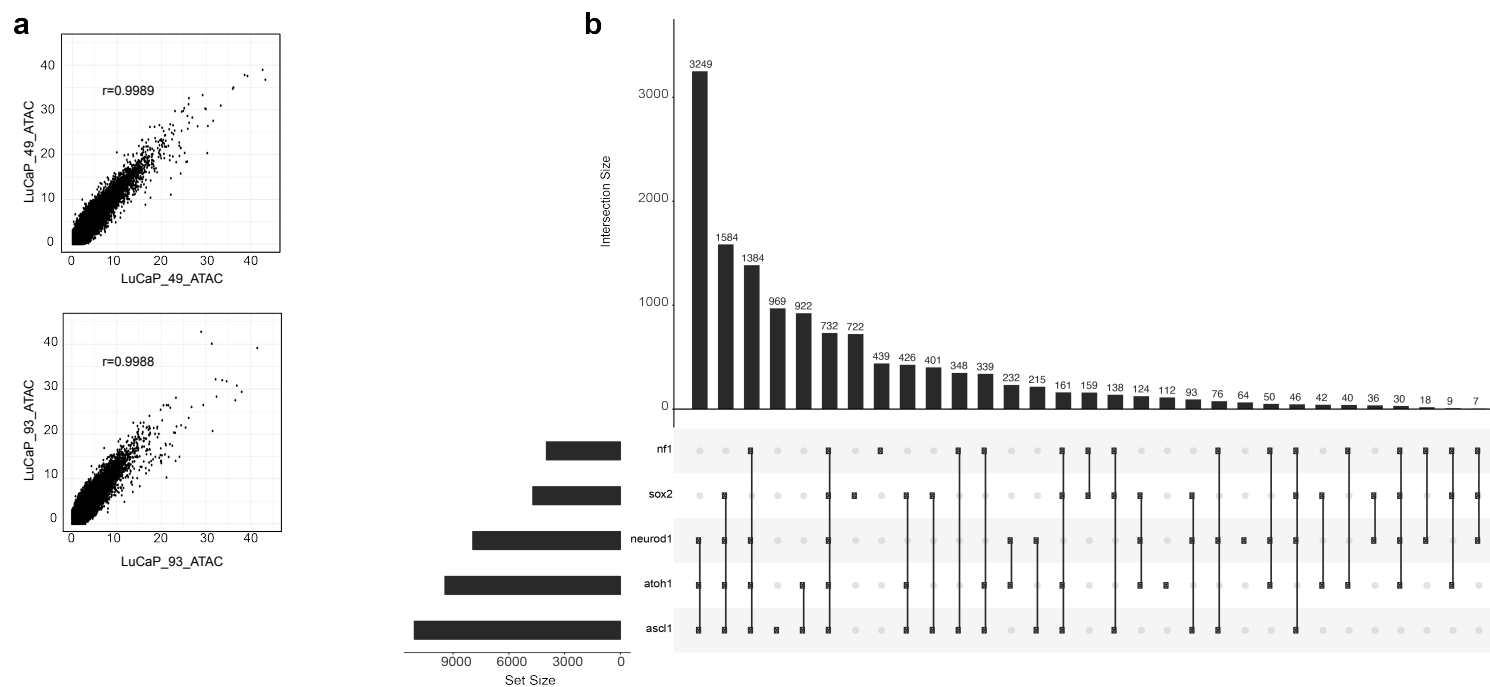
Supplemental Figure 1. (a) Correlation plots of the ATAC-seq signal for two LuCaP_49 replicates (Pearson's correlation r = 0.9989, top) and LuCaP_93 replicates (Pearson's correlation r = 0.9988, bottom). **(b)** UpSetR plot[51] showing the intersections of ATAC-seq peaks containing the specified motifs and their Intersection sizes.
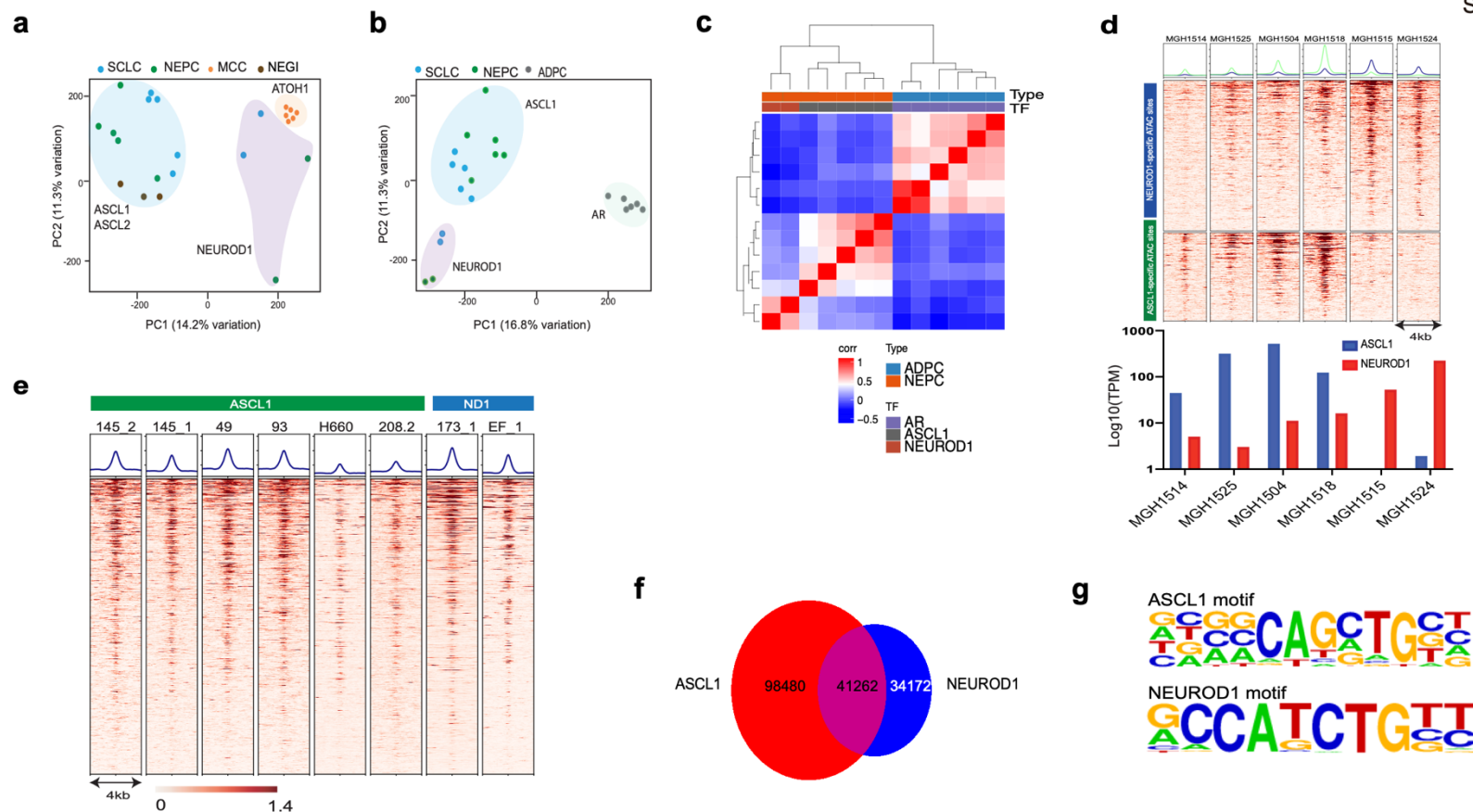
Supplemental Figure 2. (a) PCA analysis of ATAC-seq data of NECs, samples are color coded by tumor type and clusters are highlighted based on the expression of the dominant bHLH TF in each sample. (**b**) PCA analysis of ATAC-seq data of SCLC, NEPC and ADPC, samples are color coded by tumor type and clusters are highlighted based on the expression of the dominant bHLH TF in each sample. (**c**) Hierarchical clustering of the pairwise Pearson's correlation of the RNA-seq signal across the NEPCs and ADPC. (**d**) Heatmap of the ATAC-seq signal in SCLCs at the ASCL1- and NEUROD1-specific DNA accessible regions identified in NEPC (on top) and corresponding expression of *ASCL1* and *NEUROD1* of the same SCLC tissues (bottom). (**e**) Heatmap of the ATAC-seq signal at the shared accessible regions in the ASCL1 and NEUROD1 NEPC subtypes. (**f**) Venn diagram representing the union of the ASCL1 binding sites obtained by ChIP-seq in PDXs (145.1, 93, 43 and EF1) in red and the NEUROD1 bindings by ChIP-seq analysis in PDX 173.1 (in blue) (**g**) Top enriched consensus motifs identified in ASCL1 ChIP-seq (top) and NEUROD1 ChIP-seq data (bottom).
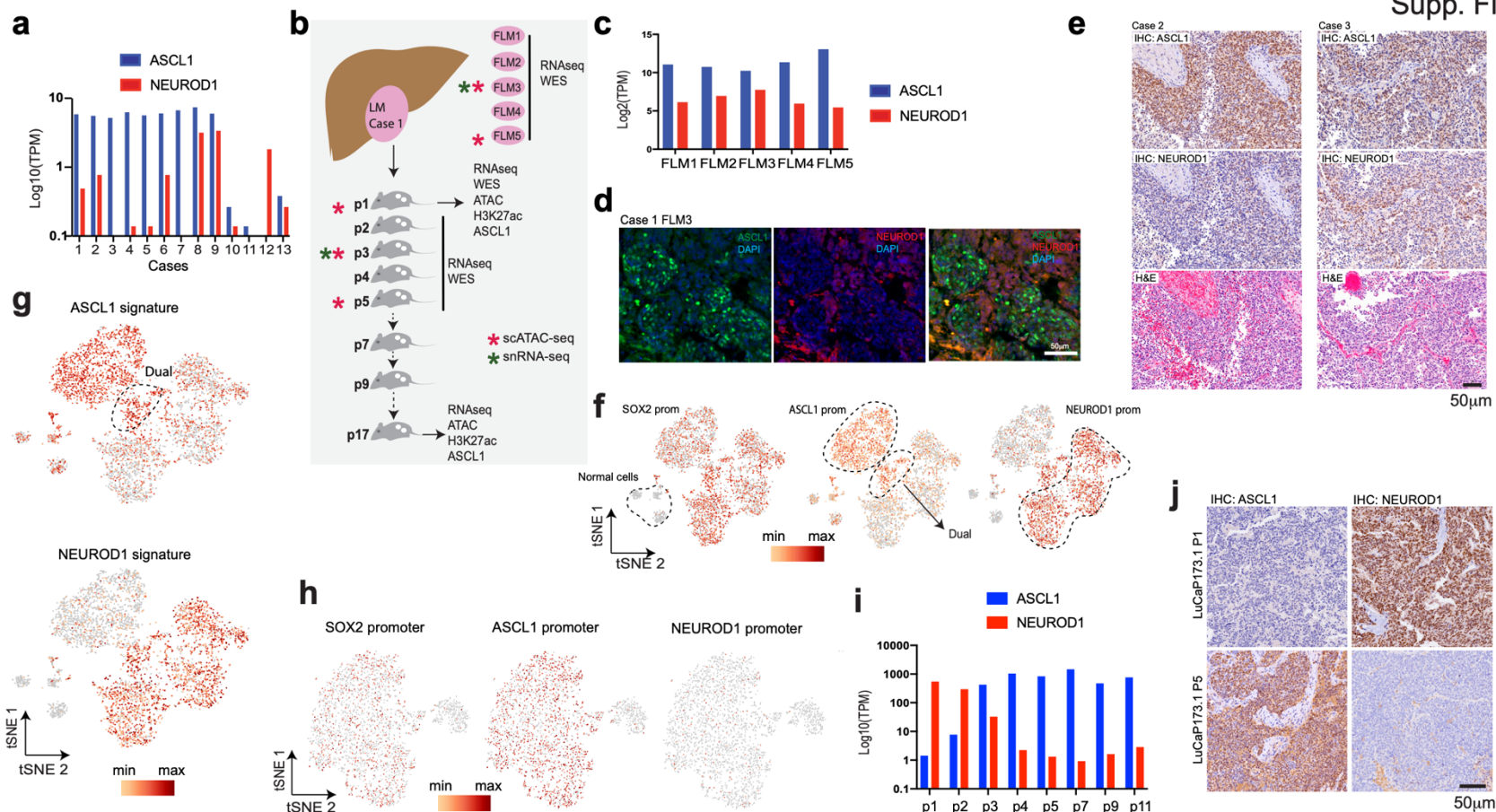
Supplemental Figure 3. (a) Plot of *ASCL1* and *NEUROD1* expression in NEPC tissues from a clinical cohort (Beltran *et al*. 2016)[22]. **(b)** Schematic of the tissues (liver metastasis and tumor passages) analyzed in this study. **(c)** Plot of *ASCL1* and *NEUROD1* expression in the five fragments of liver metastases identified in the case studied (case 1). **(d)** Double immunofluorescence for FLM3 (case 1) with ASCL1 (green) and NEUROD1 (red). **(e)** Immunohistochemical analysis of two additional NEPC liver metastasis cases, ASCL1 on top, NEUROD1 middle panel and H&E bottom. **(f)** t-SNE analysis of the scATAC-seq data of FLM3, showing accessibility at the *SOX2* promoter (marking tumor cells) and differential analysis at *ASCL1* (left) and *NEUROD1* (right) promoters marking ASCL1, "dual" and NEUROD1 subpopulations respectively. **(g)** t-SNE analysis of the scATAC-seq data of FLM3 showing accessibility at the top 30 differential ATAC-seq regions in ASCL1 (top) and NEUROD1 (bottom) subtypes identified by bulk analysis. **(h)** t-SNE analysis of the scATAC-seq data of FLM5, showing accessibility at the *SOX2* promoter (marking tumor cells) and differential analysis at *NEUROD1* (left) and *ASCL1* (right) promoters marking NEUROD1 and ASCL1 subpopulations respectively. **(i)** Analysis of *ASCL1* and *NEUROD1* expression across PDXs passages. **(j)** Immunohistochemical analysis for ASCL1 (left) and NEUROD1 (right) of PDX p1 and p5 showing anticorrelated expression of NEUROD1 and ASCL1 respectively.
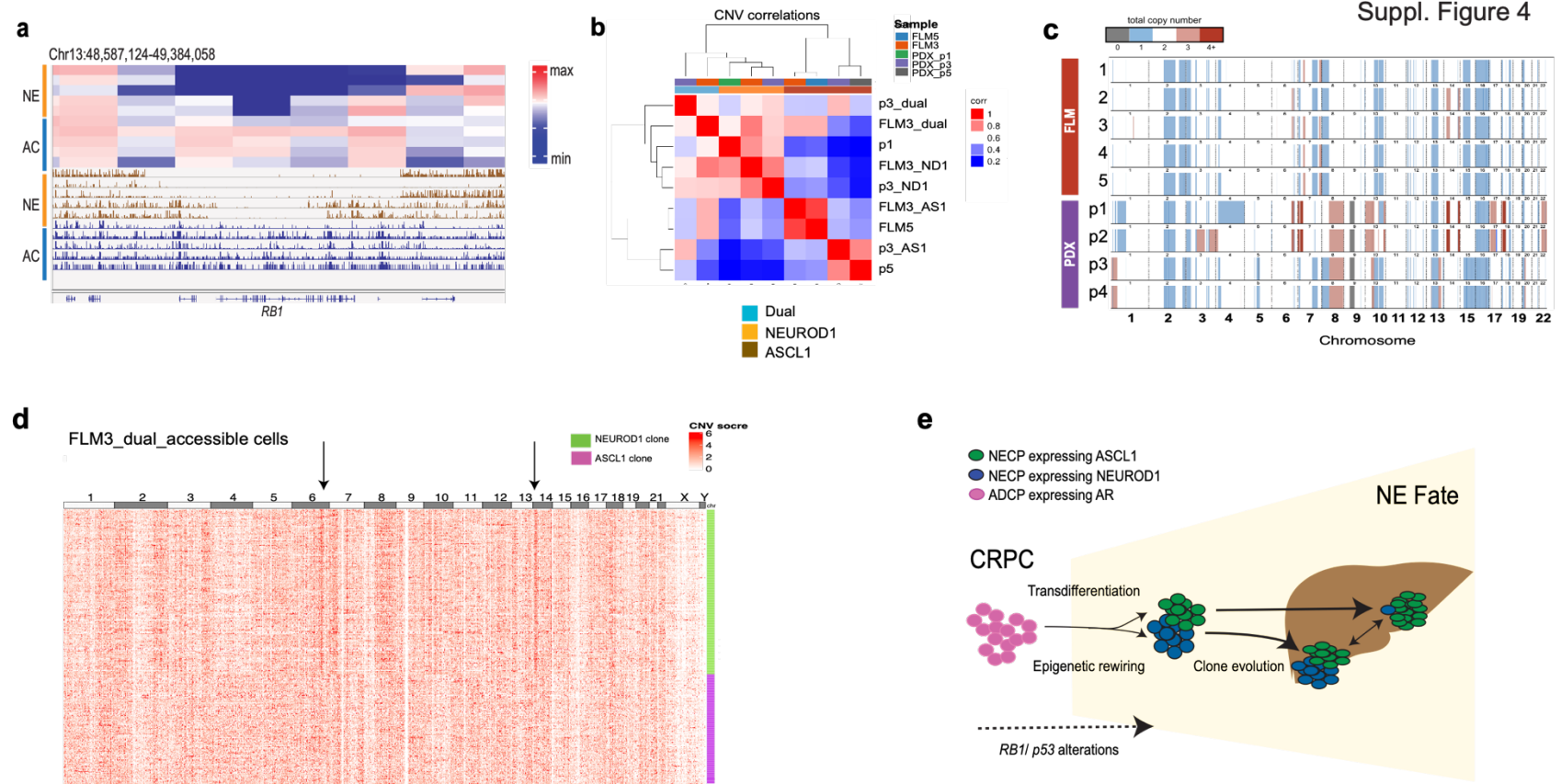
Supplemental Figure 4. (a) Inference of CNV at *RB1* locus from ATAC-seq bulk signal in NEPC and ADPC. (top) Heatmap representation of the inferred data and (bottom) IGV track of the actual ATAC-seq data at that locus. **(b)** Hierarchical clustering of the pairwise Pearson's correlation of the CNV inference from scATAC-seq t-SNE clusters in FLMs and PDX passages from the case represented in Figure 3b. **(c)** Heatmap of the CNV alterations determined by WES across FLM3 (by clusters), FLM5 and PDX passages (by clusters). **(d)** Heatmap of the single cell CNV analysis of cells showing "dual" accessibility to *ASCL1* and *NEUROD1* promoters. Two dominant clusters are evident that are the same as the larger clusters of ASCL1 and NEUROD1 subclones. **(e)** Model of the origin of the two NEPC subtypes.

Supplemental Table1. Tumor samples used for the ATAC-seq analysis

Sample	Tumor	Tissue	ATAC-seq				RNA-seq
			UniquelyMapped	Total_Peaks	DHS	FRiP	
DFMC_11112	MCC	PDX	45497753	91671	4897	36.4	Yes
DFMC_33043	MCC	PDX	50628567	90443	4828	41.1	Yes
DFMC_48396	MCC	PDX	56332609	94578	4871	34.8	Yes
DFMC_63632	MCC	PDX	44959450	66762	4671	26.7	Yes
DFMC_87346	MCC	PDX	47845776	61516	4887	38.6	Yes
DFMC_96712	MCC	PDX	44559386	72993	4832	29.4	Yes
EF1_cfce	NEPC	Cell line	95546177	76884	4890	39.1	
LuCaP_145_1_rep1	NEPC	PDX	27578828	25268	4806	12.6	
LuCaP_145_1_rep2	NEPC	PDX	40326914	41537	4819	4.6	
LuCaP_145_2	NEPC	PDX	21372235	32372	4892	16.6	Yes
LuCaP_173_1_p17	NEPC	PDX	37204748	31379	4828	33.3	Yes
LuCaP_173_1_p1	NEPC	PDX	173962297	79631	4782	8	Yes
LuCaP_208	NEPC	PDX	53527713	36861	4859	9	
LuCaP_49_rep1	NEPC	PDX	40159074	35772	4881	15	Yes
LuCaP_49_rep2	NEPC	PDX	44956639	51883	4893	21	
LuCaP_23	ADPC	PDX	30997298	49662	4868	24	Yes
LuCaP_77	ADPC	PDX	19527592	54493	4731	26.3	Yes
LuCaP_77CR	ADPC	PDX	52410587	56397	4838	20.4	
LuCaP_78	ADPC	PDX	29173802	48785	4641	17.6	Yes
LuCaP_81	ADPC	PDX	47086721	40311	4784	13.5	Yes
LuCaP_93_rep1	ADPC	PDX	39109506	35158	4664	10.5	Yes
LuCaP_93_rep2	ADPC	PDX	33565564	42470	4638	14.2	
LuCaP_96	ADPC	PDX	66307812	51946	4926	18.1	Yes
MGH_1504	SCLC	PDX	37791305	35413	4628	13.6	Yes
MGH_1514	SCLC	PDX	62263801	37249	4649	8.9	Yes
MGH_1515	SCLC	PDX	35865754	32243	4465	11.1	Yes
MGH_1518	SCLC	PDX	33289869	38847	4472	15.6	Yes
MGH_1524-1	SCLC	PDX	45397760	25676	4704	5.7	Yes
MGH_1525	SCLC	PDX	42628509	27972	4758	6.8	Yes
MGH_1545	SCLC	PDX	35339906	51209	4697	22	Yes
MGH_1567-1	SCLC	PDX	58821999	21845	4910	5.8	
NCI-H660	NEPC	Cell line	310339362	70104	4423	15.1	
T17-11572_NEC	GINEC	Primary tissue	62189628	54148	4959	22.6	Yes
T18-01479_NEC	GINEC	Primary tissue	52641558	80490	4853	30	Yes
T18-05426_I_NEC	GINEC	Primary tissue	58703600	55989	4896	36.2	Yes

DHS: DNase I hypersensitive site (per 5000 peaks)
FRIP: fraction of reads that fall into a peak

Supplemental table 2. Motif analysis at NE-specific peaks

Motif Name	Consensus	P-value	Log P-value	q-value (Benjamini)	# of Target Sequences with Motif(of 4540)	% of Target Sequences with Motif	# of Background Sequences with Motif(of 44756)	% of Background Sequences with Motif
Atoh1(bHLH)/Cerebellum-Atoh1-ChIP-Seq(GSE22111)/Homer	VNRVCAGCTGGY	1e-398	-9.17E+02	0	2781	61.26%	13641.7	30.48%
Ascl1(bHLH)/NeuralTubes-Ascl1-ChIP-Seq(GSE55840)/Homer	NNVVCAGCTGBN	1e-384	-8.85E+02	0	3364	74.10%	19420.5	43.39%
NeuroD1(bHLH)/Islet-NeuroD1-ChIP-Seq(GSE30298)/Homer	GCCATCTGTT	1e-363	-8.38E+02	0	2308	50.84%	10266.3	22.94%
NeuroG2(bHLH)/Fibroblast-NeuroG2-ChIP-Seq(GSE75910)/Homer	ACCATCTGTT	1e-328	-7.57E+02	0	3056	67.31%	17367	38.80%
HEB(bHLH)/mES-Heb-ChIP-Seq(GSE53233)/Homer	VCAGCTGBNN	1e-326	-7.52E+02	0	3542	78.02%	22448.3	50.16%
E2A(bHLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer	DNRCAGCTGY	1e-324	-7.48E+02	0	3173	69.89%	18591.1	41.54%
Olig2(bHLH)/Neuron-Olig2-ChIP-Seq(GSE30882)/Homer	RCCATMTGTT	1.00E-291	-6.71E+02	0	3228	71.10%	19821.1	44.29%
Ptf1a(bHLH)/Panc1-Ptf1a-ChIP-Seq(GSE47459)/Homer	ACAGCTGTTN	1.00E-290	-6.69E+02	0	3899	85.88%	27464	61.37%
MyoG(bHLH)/C2C12-MyoG-ChIP-Seq(GSE36024)/Homer	AACAGCTG	1.00E-266	-6.15E+02	0	2494	54.93%	13402.9	29.95%
Tcf12(bHLH)/GM12878-Tcf12-ChIP-Seq(GSE32465)/Homer	VCAGCTGYTG	1.00E-262	-6.04E+02	0	2448	53.92%	13095.1	29.26%
Tcf21(bHLH)/ArterySmoothMuscle-Tcf21-ChIP-Seq(GSE61369)/Ho	NAACAGCTGG	1.00E-258	-5.95E+02	0	2335	51.43%	12205.7	27.27%
Myf5(bHLH)/GM-Myf5-ChIP-Seq(GSE24852)/Homer	BAACAGCTGT	1.00E-240	-5.55E+02	0	1900	41.85%	9022.1	20.16%
MyoD(bHLH)/Myotube-MyoD-ChIP-Seq(GSE21614)/Homer	RRCAGCTGYTSY	1.00E-212	-4.89E+02	0	1986	43.74%	10214.4	22.82%
Ap4(bHLH)/AML-Tfap4-ChIP-Seq(GSE45738)/Homer	NAHCAGCTGD	1.00E-206	-4.75E+02	0	2524	55.59%	14927.7	33.35%
Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer	TAATTAGN	1.00E-178	-4.11E+02	0	1694	37.31%	8557.6	19.12%
Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/H	NNYTAATTAR	1.00E-169	-3.90E+02	0	1715	37.78%	8904	19.90%
Slug(Zf)/Mesoderm-Snai2-ChIP-Seq(GSE61475)/Homer	SNGCACCTGCHS	1.00E-166	-3.83E+02	0	1590	35.02%	7995.8	17.87%
Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer	ADBTAATTAR	1.00E-166	-3.83E+02	0	2166	47.71%	12695.4	28.37%
NF1(CTF)/LNCAP-NF1-ChIP-Seq(Unpublished)/Homer	CYTGGCABNSTGCCAR	1.00E-162	-3.75E+02	0	1214	26.74%	5314.3	11.87%
E2A(bHLH),near_PU.1/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	NVCACCTGBN	1.00E-148	-3.43E+02	0	2676	58.94%	17791.2	39.75%
Nkx6.1(Homeobox)/Islet-Nkx6.1-ChIP-Seq(GSE40975)/Homer	GKTAATGR	1.00E-135	-3.12E+02	0	2758	60.75%	18977	42.40%
NF1-halfsite(CTF)/LNCaP-NF1-ChIP-Seq(Unpublished)/Homer	YTGCCAAG	1.00E-131	-3.02E+02	0	2838	62.51%	19872.4	44.40%
SCL(bHLH)/HPC7-Scl-ChIP-Seq(GSE13511)/Homer	AVCAGCTG	1.00E-121	-2.79E+02	0	4232	93.22%	36268	81.04%
Tgif2(Homeobox)/mES-Tgif2-ChIP-Seq(GSE55404)/Homer	TGTCANYT	1.00E-101	-2.34E+02	0	3460	76.21%	27394.9	61.21%
Tlx?(NR)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	CTGGCAGSCTGCCA	1.00E-99	-2.28E+02	0	1114	24.54%	5775.3	12.90%
Isl1(Homeobox)/Neuron-Isl1-ChIP-Seq(GSE31456)/Homer	CTAATKGV	1.00E-93	-2.14E+02	0	2285	50.33%	15853	35.42%
ZEB1(Zf)/PDAC-ZEB1-ChIP-Seq(GSE64557)/Homer	VCAGGTRDRY	1.00E-92	-2.12E+02	0	2701	59.49%	19857.5	44.37%
Pit1+1bp(Homeobox)/GCrat-Pit1-ChIP-Seq(GSE58009)/Homer	ATGCATAATTCA	1.00E-75	-1.73E+02	0	587	12.93%	2521.5	5.63%
Tbx5(T-box)/HL1-Tbx5.biotin-ChIP-Seq(GSE21529)/Homer	AGGTGTCA	1.00E-74	-1.72E+02	0	3473	76.50%	28575.3	63.85%
Meis1(Homeobox)/MastCells-Meis1-ChIP-Seq(GSE48085)/Homer	VGCTGWCAVB	1.00E-66	-1.54E+02	0	2238	49.30%	16415.6	36.68%
ZBTB18(Zf)/HEK293-ZBTB18.GFP-ChIP-Seq(GSE58341)/Homer	AACATCTGGA	1.00E-64	-1.49E+02	0	1054	23.22%	6173.1	13.79%
Tgif1(Homeobox)/mES-Tgif1-ChIP-Seq(GSE55404)/Homer	YTGWCADY	1.00E-60	-1.39E+02	0	3119	68.70%	25409.9	56.78%
Sox6(HMG)/Myotubes-Sox6-ChIP-Seq(GSE32627)/Homer	CCATTGTTNY	1.00E-57	-1.32E+02	0	1794	39.52%	12720.4	28.42%
Pdx1(Homeobox)/Islet-Pdx1-ChIP-Seq(SRA008281)/Homer	YCATYAATCA	1.00E-56	-1.30E+02	0	1237	27.25%	7919.4	17.70%
Nanog(Homeobox)/mES-Nanog-ChIP-Seq(GSE11724)/Homer	RGCCATTAAC	1.00E-55	-1.29E+02	0	3659	80.59%	31453.8	70.28%
Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer	CCWTTGTY	1.00E-51	-1.18E+02	0	1921	42.31%	14152.9	31.62%
Bapx1(Homeobox)/VertebralCol-Bapx1-ChIP-Seq(GSE36672)/Hom	TTRAGTGSYK	1.00E-50	-1.16E+02	0	2356	51.89%	18272.6	40.83%
Foxo1(Homeobox)/RAW-Foxo1-ChIP-Seq(Fan_et_al.)/Homer	CTGTTTAC	1.00E-48	-1.11E+02	0	2159	47.56%	16511.4	36.89%
Smad4(MAD)/ESC-SMAD4-ChIP-Seq(GSE29422)/Homer	VBSYGTCTGG	1.00E-46	-1.08E+02	0	2053	45.22%	15572.2	34.79%
Sox2(HMG)/mES-Sox2-ChIP-Seq(GSE11431)/Homer	BCCATTGTTC	1.00E-46	-1.07E+02	0	1116	24.58%	7264.3	16.23%
Nkx2.1(Homeobox)/LungAC-Nkx2.1-ChIP-Seq(GSE43252)/Homer	RSCACTYRAG	1.00E-46	-1.07E+02	0	2961	65.22%	24486.1	54.71%
OCT:OCT-short(POU,Homeobox)/NPC-OCT6-ChIP-Seq(GSE43916)/	ATGCATWATGCATRW	1.00E-45	-1.05E+02	0	889	19.58%	5436.9	12.15%
AP-2gamma(AP2)/MCF7-TFAP2C-ChIP-Seq(GSE21234)/Homer	SCCTSAGGSCAW	1.00E-43	-1.01E+02	0	1859	40.95%	13932.1	31.13%
Nkx2.2(Homeobox)/NPC-Nkx2.2-ChIP-Seq(GSE61673)/Homer	BTBRAGTGSN	1.00E-43	-1.01E+02	0	2343	51.61%	18499.1	41.33%
EBF1(EBF)/Near-E2A-ChIP-Seq(GSE21512)/Homer	GTCCCCWGGGGA	1.00E-43	-9.99E+01	0	1934	42.60%	14642.8	32.72%
Sox10(HMG)/SciaticNerve-Sox3-ChIP-Seq(GSE35132)/Homer	CCWTTGTYYB	1.00E-42	-9.84E+01	0	1834	40.40%	13751.8	30.73%
Smad2(MAD)/ES-SMAD2-ChIP-Seq(GSE29422)/Homer	CTGTCTGG	1.00E-42	-9.72E+01	0	1977	43.55%	15103	33.75%
Smad3(MAD)/NPC-Smad3-ChIP-Seq(GSE36673)/Homer	TWGTCTGV	1.00E-41	-9.62E+01	0	3041	66.98%	25557.4	57.11%
Nkx2.5(Homeobox)/HL1-Nkx2.5.biotin-ChIP-Seq(GSE21529)/Home	RRSCACTYAA	1.00E-41	-9.51E+01	0	2515	55.40%	20300.3	45.36%
AR-halfsite(NR)/LNCaP-AR-ChIP-Seq(GSE27824)/Homer	CCAGGAACAG	1.00E-40	-9.42E+01	0	3619	79.71%	31749.7	70.94%
Sox4(HMG)/proB-Sox4-ChIP-Seq(GSE50066)/Homer	YCTTTGTTCC	1.00E-40	-9.41E+01	0	1086	23.92%	7227.7	16.15%
OCT:OCT(POU,Homeobox,IR1)/NPC-Brn2-ChIP-Seq(GSE35496)/Ho	ATGAATWATTCATGA	1.00E-40	-9.28E+01	0	79	1.74%	106.6	0.24%
FoxL2(Forkhead)/Ovary-FoxL2-ChIP-Seq(GSE60858)/Homer	WWTRTAAACAVG	1.00E-39	-8.99E+01	0	949	20.90%	6154.1	13.75%
EBF(EBF)/proBcell-EBF-ChIP-Seq(GSE21978)/Homer	DGTCCCYRGGGA	1.00E-38	-8.96E+01	0	553	12.18%	3031.5	6.77%
AP-2alpha(AP2)/Hela-AP2alpha-ChIP-Seq(GSE31477)/Homer	ATGCCCTGAGGC	1.00E-37	-8.58E+01	0	1526	33.61%	11217.3	25.06%
Prop1(Homeobox)/GHFT1-PROP1.biotin-ChIP-Seq(GSE77302)/Hor	NTAATBNAATTA	1.00E-37	-8.58E+01	0	733	16.15%	4467.3	9.98%
Sox15(HMG)/CPA-Sox15-ChIP-Seq(GSE62909)/Homer	RAACAATGGN	1.00E-37	-8.56E+01	0	1263	27.82%	8900	19.89%
Twist(bHLH)/HMLE-TWIST1-ChIP-Seq(Chang_et_al)/Homer	VCAKCTGGNNNCCAGN	1.00E-36	-8.51E+01	0	382	8.41%	1847.6	4.13%
Six2(Homeobox)/NephronProgenitor-Six2-ChIP-Seq(GSE39837)/H	GWAAYHTGAKMC	1.00E-34	-7.92E+01	0	1130	24.89%	7871.9	17.59%
FOXK1(Forkhead)/HEK293-FOXK1-ChIP-Seq(GSE51673)/Homer	NVWTGTTTAC	1.00E-34	-7.92E+01	0	1153	25.40%	8072.8	18.04%
Foxf1(Forkhead)/Lung-Foxf1-ChIP-Seq(GSE77951)/Homer	WWATRTAAACAN	1.00E-33	-7.73E+01	0	979	21.56%	6622.5	14.80%
Rfx6(HTH)/Min6b1-Rfx6.HA-ChIP-Seq(GSE62844)/Homer	TGTTKCCTAGCAACM	1.00E-33	-7.71E+01	0	1551	34.16%	11644.4	26.02%
FOXA1(Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/Homer	WAAGTAAACA	1.00E-33	-7.67E+01	0	1244	27.40%	8915.2	19.92%
Hoxc9(Homeobox)/Ainv15-Hoxc9-ChIP-Seq(GSE21812)/Homer	GGCCATAAATCA	1.00E-32	-7.43E+01	0	644	14.19%	3928.1	8.78%
Sox17(HMG)/Endoderm-Sox17-ChIP-Seq(GSE61475)/Homer	CCATTGTTYB	1.00E-32	-7.41E+01	0	896	19.74%	5983.1	13.37%
Fox:Ebox(Forkhead,bHLH)/Panc1-Foxa2-ChIP-Seq(GSE47459)/Hom	NNNVCTGWGYAAACA	1.00E-31	-7.25E+01	0	1245	27.42%	9013	20.14%
RBPJ:Ebox(?,bHLH)/Panc1-Rbpj1-ChIP-Seq(GSE47459)/Homer	GGGRAARRGRMCAGN	1.00E-31	-7.21E+01	0	696	15.33%	4375.2	9.78%
Nkx3.1(Homeobox)/LNCaP-Nkx3.1-ChIP-Seq(GSE28264)/Homer	AAGCACTTAA	1.00E-30	-6.99E+01	0	2371	52.22%	19550.1	43.68%
Foxo3(Forkhead)/U2OS-Foxo3-ChIP-Seq(E-MTAB-2701)/Homer	DGTAAACA	1.00E-30	-6.93E+01	0	859	18.92%	5759.9	12.87%
Six1(Homeobox)/Myoblast-Six1-ChIP-Chip(GSE20150)/Homer	GKVTCADRTTWC	1.00E-29	-6.82E+01	0	371	8.17%	1935.6	4.32%
Phox2a(Homeobox)/Neuron-Phox2a-ChIP-Seq(GSE31456)/Homer	YTAATYNRATTA	1.00E-29	-6.81E+01	0	486	10.70%	2787.4	6.23%
Unknown(Homeobox)/Limb-p300-ChIP-Seq/Homer	SSCMATWAAA	1.00E-29	-6.78E+01	0	767	16.89%	5024.5	11.23%
Barx1(Homeobox)/Stomach-Barx1.3xFlag-ChIP-Seq(GSE69483)/H	AAACMATTAN	1.00E-29	-6.75E+01	0	598	13.17%	3664.3	8.19%
FOXA1(Forkhead)/MCF7-FOXA1-ChIP-Seq(GSE26831)/Homer	WAAGTAAACA	1.00E-29	-6.68E+01	0	1030	22.69%	7259.6	16.22%
Eomes(T-box)/H9-Eomes-ChIP-Seq(GSE26097)/Homer	ATTAACACCT	1.00E-28	-6.61E+01	0	2277	50.15%	18745.6	41.89%
HOXA9(Homeobox)/HSC-Hoxa9-ChIP-Seq(GSE33509)/Homer	GGCCATAAATCA	1.00E-28	-6.59E+01	0	807	17.78%	5388.1	12.04%

Motif	Consensus	P-value	logP		Target	Target %	Bkg	Bkg %
PRDM9(Zf)/Testis-DMC1-ChIP-Seq(GSE35498)/Homer	ADGGYAGYAGCATCT	1.00E-28	-6.56E+01	0	769	16.94%	5078.7	11.35%
Pit1(Homeobox)/GCrat-Pit1-ChIP-Seq(GSE58009)/Homer	ATGMATATDC	1.00E-28	-6.49E+01	0	1073	23.63%	7672.5	17.14%
Znf263(Zf)/K562-Znf263-ChIP-Seq(GSE31477)/Homer	CVGTSCTCCC	1.00E-27	-6.42E+01	0	2470	54.41%	20682.3	46.21%
Zic3(Zf)/mES-Zic3-ChIP-Seq(GSE37889)/Homer	GGCCYCCTGCTGDGH	1.00E-27	-6.28E+01	0	1197	26.37%	8805	19.67%
Foxa2(Forkhead)/Liver-Foxa2-ChIP-Seq(GSE25694)/Homer	CYTGTTTACWYW	1.00E-27	-6.23E+01	0	920	20.26%	6410	14.32%
HOXD13(Homeobox)/Chicken-Hoxd13-ChIP-Seq(GSE38910)/Home	NCYAATAAAA	1.00E-26	-6.13E+01	0	1091	24.03%	7907.2	17.67%
Oct6(POU,Homeobox)/NPC-Pou3f1-ChIP-Seq(GSE35496)/Homer	WATGCAAATGAG	1.00E-26	-6.12E+01	0	533	11.74%	3245.5	7.25%
Unknown-ESC-element(?)/mES-Nanog-ChIP-Seq(GSE11724)/Hom	CACAGCAGGGGG	1.00E-24	-5.71E+01	0	1141	25.13%	8445.8	18.87%
Tbr1(T-box)/Cortex-Tbr1-ChIP-Seq(GSE71384)/Homer	AAGGTGTKAA	1.00E-24	-5.65E+01	0	1566	34.49%	12302.5	27.49%
NPAS2(bHLH)/Liver-NPAS2-ChIP-Seq(GSE39860)/Homer	KCCACGTGAC	1.00E-23	-5.51E+01	0	1416	31.19%	10965.8	24.50%
Hoxb4(Homeobox)/ES-Hoxb4-ChIP-Seq(GSE34014)/Homer	TGATTRATGGCY	1.00E-22	-5.28E+01	0	296	6.52%	1560.8	3.49%
FOXK2(Forkhead)/U2OS-FOXK2-ChIP-Seq(E-MTAB-2204)/Homer	SCHTGTTTACAT	1.00E-22	-5.25E+01	0	768	16.92%	5317.8	11.88%
Pitx1(Homeobox)/Chicken-Pitx1-ChIP-Seq(GSE38910)/Homer	TAATCCCN	1.00E-22	-5.25E+01	0	3372	74.27%	30210.4	67.50%
AMYB(HTH)/Testes-AMYB-ChIP-Seq(GSE44588)/Homer	TGGCAGTTGG	1.00E-22	-5.13E+01	0	1630	35.90%	13044.7	29.15%
Zic(Zf)/Cerebellum-ZIC1.2-ChIP-Seq(GSE60731)/Homer	CCTGCTGAGH	1.00E-21	-5.01E+01	0	1294	28.50%	9995	22.33%
FOXP1(Forkhead)/H9-FOXP1-ChIP-Seq(GSE31006)/Homer	NYYTGTTTACHN	1.00E-21	-4.98E+01	0	522	11.50%	3335.5	7.45%
FOXM1(Forkhead)/MCF7-FOXM1-ChIP-Seq(GSE72977)/Homer	TRTTTACTTW	1.00E-21	-4.90E+01	0	1040	22.91%	7751.5	17.32%
Pax8(Paired,Homeobox)/Thyroid-Pax8-ChIP-Seq(GSE26938)/Home	GTCATGCHTGRCTGS	1.00E-20	-4.78E+01	0	618	13.61%	4154.6	9.28%
Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer	GGGGGGGG	1.00E-20	-4.76E+01	0	1681	37.03%	13632	30.46%
Brn2(POU,Homeobox)/NPC-Brn2-ChIP-Seq(GSE35496)/Homer	ATGAATATTC	1.00E-19	-4.57E+01	0	165	3.63%	725.5	1.62%
BMAL1(bHLH)/Liver-Bmal1-ChIP-Seq(GSE39860)/Homer	GNCACGTG	1.00E-19	-4.45E+01	0	1959	43.15%	16367.2	36.57%
Sox9(HMG)/Limb-SOX9-ChIP-Seq(GSE73225)/Homer	AGGVNCCTTTGT	1.00E-19	-4.41E+01	0	1088	23.96%	8304.9	18.56%
Pitx1:Ebox(Homeobox,bHLH)/Hindlimb-Pitx1-ChIP-Seq(GSE41591)	YTAATTRAWWCCAGA	1.00E-18	-4.35E+01	0	287	6.32%	1601.3	3.58%
Rfx5(HTH)/GM12878-Rfx5-ChIP-Seq(GSE31477)/Homer	SCCTAGCAACAG	1.00E-18	-4.33E+01	0	445	9.80%	2826.3	6.31%
Foxa3(Forkhead)/Liver-Foxa3-ChIP-Seq(GSE77670)/Homer	BSNTGTTTACWYWGN	1.00E-18	-4.32E+01	0	395	8.70%	2431.2	5.43%
Brn1(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Homer	TATGCWAATBAV	1.00E-18	-4.27E+01	0	380	8.37%	2321.6	5.19%
p63(p53)/Keratinocyte-p63-ChIP-Seq(GSE17611)/Homer	NNDRCATGCYCNRRCA	1.00E-18	-4.18E+01	0	536	11.81%	3591.9	8.03%
ETS:E-box(ETS,bHLH)/HPC7-Scl-ChIP-Seq(GSE22178)/Homer	AGGAARCAGCTG	1.00E-17	-4.02E+01	0	207	4.56%	1059.8	2.37%
PAX5(Paired,Homeobox)/GM12878-PAX5-ChIP-Seq(GSE32465)/Ho	GCAGCCAAGCRTGACH	1.00E-17	-3.92E+01	0	582	12.82%	4025.6	8.99%
CRX(Homeobox)/Retina-Crx-ChIP-Seq(GSE20012)/Homer	GCTAATCC	1.00E-16	-3.85E+01	0	2126	46.83%	18188.7	40.64%
Tbet(T-box)/CD8-Tbet-ChIP-Seq(GSE33802)/Homer	AGGTGTGAAM	1.00E-16	-3.84E+01	0	1225	26.98%	9708.3	21.69%
BMYB(HTH)/Hela-BMYB-ChIP-Seq(GSE27030)/Homer	NHAACBGYYV	1.00E-16	-3.83E+01	0	1502	33.08%	12267.6	27.41%
NPAS(bHLH)/Liver-NPAS-ChIP-Seq(GSE39860)/Homer	NVCACGTG	1.00E-16	-3.70E+01	0	1733	38.17%	14481.5	32.36%
PAX6(Homeobox)/Forebrain-Pax6-ChIP-Seq(GSE66961)/Ho	NGTGTTCAVTSAAGCG	1.00E-16	-3.69E+01	0	166	3.66%	806.4	1.80%
CDX4(Homeobox)/ZebrafishEmbryos-Cdx4.Myc-ChIP-Seq(GSE4825	NGYCATAAAWCH	1.00E-16	-3.63E+01	0	880	19.38%	6669.2	14.90%
p73(p53)/Trachea-p73-ChIP-Seq(PRJNA310161)/Homer	NRRRCAWGTCCDGRC	1.00E-15	-3.61E+01	0	112	2.47%	459.9	1.03%
MYB(HTH)/ERMYB-Myb-ChIPSeq(GSE22095)/Homer	GGCVGTTR	1.00E-15	-3.60E+01	0	1718	37.84%	14377.9	32.13%
Cux2(Homeobox)/Liver-Cux2-ChIP-Seq(GSE35985)/Homer	HNRAATCAAT	1.00E-15	-3.59E+01	0	488	10.75%	3311.8	7.40%
Tbx20(T-box)/Heart-Tbx20-ChIP-Seq(GSE29636)/Homer	GGTGYTGACAGS	1.00E-15	-3.56E+01	0	358	7.89%	2261.4	5.05%
RFX(HTH)/K562-RFX3-ChIP-Seq(SRA012198)/Homer	CGGTTGCCATGGCAAC	1.00E-15	-3.54E+01	0	134	2.95%	605.3	1.35%
ZNF416(Zf)/HEK293-ZNF416.GFP-ChIP-Seq(GSE58341)/Homer	WDNCTGGGCA	1.00E-15	-3.51E+01	0	1917	42.22%	16307.5	36.44%
HOXA2(Homeobox)/mES-Hoxa2-ChIP-Seq(Donaldson_et_al.)/Hom	GYCATCMATCAT	1.00E-14	-3.40E+01	0	161	3.55%	799.2	1.79%
Rfx1(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	KGTTGCCATGGCAA	1.00E-14	-3.32E+01	0	302	6.65%	1858.1	4.15%
KLF14(Zf)/HEK293-KLF14.GFP-ChIP-Seq(GSE58341)/Homer	RGKGGGCGKGGC	1.00E-13	-3.21E+01	0	1950	42.95%	16742.8	37.41%
p53(p53)/Saos-p53-ChIP-Seq(GSE15780)/Homer	RRCATGYCYRGRCATG	1.00E-13	-3.17E+01	0	162	3.57%	829.2	1.85%
p53(p53)/Saos-p53-ChIP-Seq/Homer	RRCATGYCYRGRCATG	1.00E-13	-3.17E+01	0	162	3.57%	829.2	1.85%
X-box(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	GGTTGCCATGGCAA	1.00E-13	-3.16E+01	0	174	3.83%	917	2.05%
Sp5(Zf)/mES-Sp5.Flag-ChIP-Seq(GSE72989)/Homer	RGKGGGCGGAGC	1.00E-13	-3.15E+01	0	1188	26.17%	9598.3	21.45%
Oct4(POU,Homeobox)/mES-Oct4-ChIP-Seq(GSE11431)/Homer	ATTTGCATAW	1.00E-13	-3.15E+01	0	512	11.28%	3604.7	8.05%
HNF6(Homeobox)/Liver-Hnf6-ChIP-Seq(ERP000394)/Homer	NTATYGATCH	1.00E-13	-3.14E+01	0	565	12.44%	4056.5	9.06%
Rfx2(HTH)/LoVo-RFX2-ChIP-Seq(GSE49402)/Homer	GTTGCCATGGCAACM	1.00E-13	-3.13E+01	0	149	3.28%	743.1	1.66%
Mef2b(MADS)/HEK293-Mef2b.V5-ChIP-Seq(GSE67450)/Homer	GCTATTTTTGGM	1.00E-13	-3.00E+01	0	841	18.52%	6505.5	14.54%
Mef2c(MADS)/GM12878-Mef2c-ChIP-Seq(GSE32465)/Homer	DCYAAAAATAGM	1.00E-12	-2.88E+01	0	455	10.02%	3184.8	7.12%
Tbox:Smad(T-box,MAD)/ESCd5-Smad2_3-ChIP-Seq(GSE29422)/Ho	AGGTGHCAGACA	1.00E-12	-2.87E+01	0	304	6.70%	1946.2	4.35%
Otx2(Homeobox)/EpiLC-Otx2-ChIP-Seq(GSE56098)/Homer	NYTAATCCYB	1.00E-11	-2.69E+01	0	791	17.42%	6152	13.75%
ZNF467(Zf)/HEK293-ZNF467.GFP-ChIP-Seq(GSE58341)/Homer	TGGGGAAGGGCM	1.00E-11	-2.58E+01	0	1272	28.02%	10586.3	23.65%
GSC(Homeobox)/FrogEmbryos-GSC-ChIP-Seq(DRA000576)/Home	RGGATTAR	1.00E-11	-2.57E+01	0	1073	23.63%	8748.4	19.55%
Cdx2(Homeobox)/mES-Cdx2-ChIP-Seq(GSE14586)/Homer	GYMATAAAAH	1.00E-11	-2.55E+01	0	672	14.80%	5141.2	11.49%
Hand2(bHLH)/Mesoderm-Hand2-ChIP-Seq(GSE61475)/Homer	TGACANARRCCAGRC	1.00E-10	-2.43E+01	0	697	15.35%	5397.2	12.06%
MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer	TGCTGACTCA	1.00E-10	-2.42E+01	0	1036	22.82%	8467.5	18.92%
Rbpj1(?)/Panc1-Rbpj1-ChIP-Seq(GSE47459)/Homer	HTTTCCCASG	1.00E-10	-2.39E+01	0	1748	38.50%	15163.1	33.88%
PAX3:FKHR-fusion(Paired,Homeobox)/Rh4-PAX3:FKHR-ChIP-Seq(G	ACCRTGACTAATTNN	1.00E-10	-2.32E+01	0	229	5.04%	1445.3	3.23%
Tcf4(HMG)/Hct116-Tcf4-ChIP-Seq(SRA012054)/Homer	ASATCAAAGGVA	1.00E-09	-2.31E+01	0	506	11.15%	3781.7	8.45%
HOXB13(Homeobox)/ProstateTumor-HOXB13-ChIP-Seq(GSE56288	TTTTATKRGG	1.00E-09	-2.10E+01	0	988	21.76%	8149.4	18.21%
HNF1b(Homeobox)/PDAC-HNF1B-ChIP-Seq(GSE64557)/Homer	GTTAATNATTAA	1.00E-08	-2.06E+01	0	166	3.66%	995	2.22%
OCT4-SOX2-TCF-NANOG(POU,Homeobox,HMG)/mES-Oct4-ChIP-S	ATTTGCATAACAATG	1.00E-08	-2.04E+01	0	213	4.69%	1365.9	3.05%
Foxh1(Forkhead)/hESC-FOXH1-ChIP-Seq(GSE29422)/Homer	NNTGTGGATTSS	1.00E-08	-1.97E+01	0	642	14.14%	5058.8	11.30%
Mef2d(MADS)/Retina-Mef2d-ChIP-Seq(GSE61391)/Homer	GCTATTTTTAGC	1.00E-08	-1.96E+01	0	215	4.74%	1395.2	3.12%
Pbx3(Homeobox)/GM12878-PBX3-ChIP-Seq(GSE32465)/Homer	SCTGTCAMTCAN	1.00E-08	-1.92E+01	0	334	7.36%	2386.4	5.33%
NFAT(RHD)/Jurkat-NFATC1-ChIP-Seq(Jolma_et_al.)/Homer	ATTTTCCATT	1.00E-08	-1.90E+01	0	969	21.34%	8056.8	18.00%
PBX1(Homeobox)/MCF7-PBX1-ChIP-Seq(GSE28007)/Homer	GSCTGTCACTCA	1.00E-08	-1.86E+01	0	136	3.00%	794.8	1.78%
Brachyury(T-box)/Mesoendoderm-Brachyury-ChIP-exo(GSE54963)	ANTTMRCASBNNNGT	1.00E-07	-1.81E+01	0	323	7.11%	2318.3	5.18%
Pax7(Paired,Homeobox)/Myoblast-Pax7-ChIP-Seq(GSE25064)/Hor	TAATCAATTA	1.00E-07	-1.80E+01	0	120	2.64%	683.7	1.53%
ZNF415(Zf)/HEK293-ZNF415.GFP-ChIP-Seq(GSE58341)/Homer	GRTGMTRGAGCC	1.00E-07	-1.75E+01	0	795	17.51%	6523.5	14.58%
ZNF189(Zf)/HEK293-ZNF189.GFP-ChIP-Seq(GSE58341)/Homer	TGGAACAGMA	1.00E-07	-1.75E+01	0	1010	22.25%	8504	19.00%
HIF-1b(HLH)/T47D-HIF1b-ChIP-Seq(GSE59937)/Homer	RTACGTGC	1.00E-07	-1.74E+01	0	1053	23.19%	8907.3	19.90%
TATA-Box(TBP)/Promoter/Homer	CCTTTTAWAGSC	1.00E-07	-1.70E+01	0	1228	27.05%	10566.3	23.61%
GATA3(Zf)/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	AGATAASR	1.00E-07	-1.67E+01	0	1328	29.25%	11527.6	25.76%
Arnt:Ahr(bHLH)/MCF7-Arnt-ChIP-Seq(Lo_et_al.)/Homer	TBGCACGCAA	1.00E-07	-1.65E+01	0	662	14.58%	5352.9	11.96%
PR(NR)/T47D-PR-ChIP-Seq(GSE31130)/Homer	VAGRACAKNCTGTBC	1.00E-07	-1.65E+01	0	2150	47.36%	19449.9	43.46%
Stat3+il21(Stat)/CD4-Stat3-ChIP-Seq(GSE19198)/Homer	SVYTTCCNGGAARB	1.00E-06	-1.52E+01	0	836	18.41%	6999.5	15.64%

Name	Motif	P	Log P	q-value	# Target	% Target	# Bg	% Bg
Oct4:Sox17(POU,Homeobox,HMG)/F9-Sox17-ChIP-Seq(GSE44553	CCATTGTATGCAAAT	1.00E-06	-1.50E+01	0	154	3.39%	990.2	2.21%
Bcl6(Zf)/Liver-Bcl6-ChIP-Seq(GSE31578)/Homer	NNNCTTTCCAGGAAA	1.00E-06	-1.50E+01	0	1558	34.32%	13809.6	30.86%
Zfp281(Zf)/ES-Zfp281-ChIP-Seq(GSE81042)/Homer	CCCCTCCCCCAC	1.00E-06	-1.44E+01	0	271	5.97%	1969.8	4.40%
Oct2(POU,Homeobox)/Bcell-Oct2-ChIP-Seq(GSE21512)/Homer	ATATGCAAAT	1.00E-06	-1.42E+01	0	299	6.59%	2214.3	4.95%
Hnf1(Homeobox)/Liver-Foxa2-Chip-Seq(GSE25694)/Homer	GGTTAAWCATTAA	1.00E-06	-1.41E+01	0	134	2.95%	847.7	1.89%
Pax7(Paired,Homeobox),long/Myoblast-Pax7-ChIP-Seq(GSE25064	TAATCHGATTAC	1.00E-06	-1.40E+01	0	42	0.93%	177.9	0.40%
HLF(bZIP)/HSC-HLF.Flag-ChIP-Seq(GSE69817)/Homer	RTTATGYAAB	1.00E-06	-1.39E+01	0	721	15.88%	5996.3	13.40%
Tcf3(HMG)/mES-Tcf3-ChIP-Seq(GSE11724)/Homer	ASWTCAAAGG	1.00E-05	-1.34E+01	0	286	6.30%	2124.5	4.75%
ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer	AGGCCTRG	1.00E-05	-1.32E+01	0	1851	40.77%	16748	37.42%
STAT4(Stat)/CD4-Stat4-ChIP-Seq(GSE22104)/Homer	NYTTCCWGGAAR	1.00E-05	-1.31E+01	0	1040	22.91%	9004	20.12%
Atf7(bZIP)/3T3L1-Atf7-ChIP-Seq(GSE56872)/Homer	NGRTGACGTCAY	1.00E-05	-1.30E+01	0	479	10.55%	3838.2	8.58%
ZNF165(Zf)/WHIM12-ZNF165-ChIP-Seq(GSE65937)/Homer	AAGGKGRCGCAGGCA	1.00E-05	-1.28E+01	0	258	5.68%	1902.6	4.25%
ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	ACAGGAAGTG	1.00E-05	-1.27E+01	0	1780	39.21%	16094.1	35.96%
Mef2a(MADS)/HL1-Mef2a.biotin-ChIP-Seq(GSE21529)/Homer	CYAAAAATAG	1.00E-05	-1.25E+01	0	420	9.25%	3328.7	7.44%
ZNF322(Zf)/HEK293-ZNF322.GFP-ChIP-Seq(GSE58341)/Homer	GAGCCTGGTACTGWG	1.00E-05	-1.22E+01	0	512	11.28%	4168.7	9.31%
Pknox1(Homeobox)/ES-Prep1-ChIP-Seq(GSE63282)/Homer	SCTGTCAVTCAV	1.00E-05	-1.21E+01	0	313	6.89%	2398.4	5.36%
Max(bHLH)/K562-Max-ChIP-Seq(GSE31477)/Homer	RCCACGTGGYYN	1.00E-05	-1.19E+01	0	785	17.29%	6684.6	14.94%
PRDM1(Zf)/Hela-PRDM1-ChIP-Seq(GSE31477)/Homer	ACTTTCACTTTC	1.00E-05	-1.19E+01	0	573	12.62%	4734.5	10.58%
CLOCK(bHLH)/Liver-Clock-ChIP-Seq(GSE39860)/Homer	GHCACGTG	1.00E-05	-1.18E+01	0	602	13.26%	5002.5	11.18%
c-Jun-CRE(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	ATGACGTCATCY	1.00E-05	-1.17E+01	0	323	7.11%	2498.8	5.58%
Tcfcp2l1(CP2)/mES-Tcfcp2l1-ChIP-Seq(GSE11431)/Homer	NRAACCRGTTYRAACCI	1.00E-04	-1.15E+01	0	191	4.21%	1367	3.05%
Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer	NRRTGACGTCAT	1.00E-04	-1.14E+01	0	344	7.58%	2694.6	6.02%
MafB(bZIP)/BMM-Mafb-ChIP-Seq(GSE75722)/Homer	WNTGCTGASTCAGCAI	1.00E-04	-1.08E+01	0	481	10.59%	3945.2	8.82%
GATA:SCL(Zf,bHLH)/Ter119-SCL-ChIP-Seq(GSE18720)/Homer	CRGCTGBNGNSNNSA	1.00E-04	-1.06E+01	0	163	3.59%	1151.6	2.57%
E2F3(E2F)/MEF-E2F3-ChIP-Seq(GSE71376)/Homer	BTKGGCGGGAAA	1.00E-04	-1.05E+01	0.0001	578	12.73%	4842.6	10.82%
E2F6(E2F)/Hela-E2F6-ChIP-Seq(GSE31477)/Homer	GGCGGGAARN	1.00E-04	-1.04E+01	0.0001	496	10.93%	4097.2	9.15%
CEBP:AP1(bZIP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer	DRTGTTGCAA	1.00E-04	-9.64E+00	0.0001	726	15.99%	6255.5	13.98%
Usf2(bHLH)/C2C12-Usf2-ChIP-Seq(GSE36030)/Homer	GTCACGTGGT	1.00E-04	-9.46E+00	0.0002	343	7.56%	2754.9	6.16%
Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	BBCTTATCTS	1.00E-04	-9.37E+00	0.0002	615	13.55%	5239.3	11.71%
PAX5(Paired,Homeobox),condensed/GM12878-PAX5-ChIP-Seq(GS	GTCACGCTCSCTGM	1.00E-04	-9.24E+00	0.0002	149	3.28%	1067.1	2.38%
Atf1(bZIP)/K562-ATF1-ChIP-Seq(GSE31477)/Homer	GATGACGTCA	1.00E-03	-9.18E+00	0.0002	624	13.74%	5332.6	11.92%
E2F4(E2F)/K562-E2F4-ChIP-Seq(GSE31477)/Homer	GGCGGGAAAH	1.00E-03	-9.03E+00	0.0002	343	7.56%	2771.3	6.19%
KLF5(Zf)/LoVo-KLF5-ChIP-Seq(GSE49402)/Homer	DGGGYGKGGC	1.00E-03	-8.56E+00	0.0004	1442	31.76%	13128.7	29.33%
NFY(CCAAT)/Promoter/Homer	RGCCAATSRG	1.00E-03	-8.53E+00	0.0004	764	16.83%	6675.1	14.91%
MITF(bHLH)/MastCells-MITF-ChIP-Seq(GSE48085)/Homer	RTCATGTGAC	1.00E-03	-8.40E+00	0.0004	969	21.34%	8617.8	19.26%
Gata6(Zf)/HUG1N-GATA6-ChIP-Seq(GSE51936)/Homer	YCTTATCTBN	1.00E-03	-8.39E+00	0.0004	817	18.00%	7182.4	16.05%
Atf4(bZIP)/MEF-Atf4-ChIP-Seq(GSE35681)/Homer	MTGATGCAAT	1.00E-03	-8.37E+00	0.0004	276	6.08%	2197.2	4.91%
KLF6(Zf)/PDAC-KLF6-ChIP-Seq(GSE64557)/Homer	MKGGGYGTGGCC	1.00E-03	-8.02E+00	0.0006	1255	27.64%	11373.4	25.41%
Gata4(Zf)/Heart-Gata4-ChIP-Seq(GSE35151)/Homer	NBWGATAAGR	1.00E-03	-7.64E+00	0.0009	907	19.98%	8082.7	18.06%
Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	SAGATAAGRV	1.00E-03	-7.51E+00	0.001	541	11.92%	4654.9	10.40%
bZIP:IRF(bZIP,IRF)/Th17-BatF-ChIP-Seq(GSE39756)/Homer	NAGTTTCABTHTGACTI	1.00E-03	-7.42E+00	0.0011	384	8.46%	3212.6	7.18%
ZNF143|STAF(Zf)/CUTLL-ZNF143-ChIP-Seq(GSE29600)/Homer	ATTTCCCAGVAKSCY	1.00E-03	-7.24E+00	0.0013	480	10.57%	4105	9.17%
Egr1(Zf)/K562-Egr1-ChIP-Seq(GSE32465)/Homer	TGCGTGGGYG	1.00E-03	-7.19E+00	0.0014	765	16.85%	6771.3	15.13%
Stat3(Stat)/mES-Stat3-ChIP-Seq(GSE11431)/Homer	CTTCCGGGAA	1.00E-03	-7.14E+00	0.0014	563	12.40%	4880.5	10.90%
ZSCAN22(Zf)/HEK293-ZSCAN22.GFP-ChIP-Seq(GSE58341)/Homer	SMCAGTCWGAKGGAG	1.00E-02	-6.64E+00	0.0024	132	2.91%	990.3	2.21%
DMRT1(DM)/Testis-DMRT1-ChIP-Seq(GSE64892)/Homer	TWGHWACAWTGTWD	1.00E-02	-6.41E+00	0.0029	188	4.14%	1487.9	3.32%
JunD(bZIP)/K562-JunD-ChIP-Seq/Homer	ATGACGTCATCN	1.00E-02	-6.36E+00	0.0031	87	1.92%	615.5	1.38%
NF1:FOXA1(CTF,Forkhead)/LNCAP-FOXA1-ChIP-Seq(GSE27824)/H	WNTGTTTRYTTTGGCA	1.00E-02	-6.34E+00	0.0031	62	1.37%	410.8	0.92%
Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	NNATGASTCATH	1.00E-02	-6.14E+00	0.0038	506	11.15%	4409.4	9.85%
PU.1-IRF(ETS:IRF)/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	MGGAAGTGAAAC	1.00E-02	-6.10E+00	0.0039	1347	29.67%	12423.1	27.76%
USF1(bHLH)/GM12878-Usf1-ChIP-Seq(GSE32465)/Homer	SGTCACGTGR	1.00E-02	-5.82E+00	0.0052	503	11.08%	4401.6	9.83%
CHR(?)/Hela-CellCycle-Expression/Homer	SRGTTTCAAA	1.00E-02	-5.73E+00	0.0057	558	12.29%	4922.3	11.00%
GATA(Zf),IR3/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	NNNNNBAGATAWYAT	1.00E-02	-5.59E+00	0.0065	143	3.15%	1118.5	2.50%
p53(p53)/mES-cMyc-ChIP-Seq(GSE11431)/Homer	ACATGCCCGGCAT	1.00E-02	-5.58E+00	0.0065	34	0.75%	202.6	0.45%
ZNF136(Zf)/HEK293-ZNF136.GFP-ChIP-Seq(GSE58341)/Homer	YTKGATAHAGTATTCTV	1.00E-02	-5.57E+00	0.0066	85	1.87%	617.3	1.38%
DMRT6(DM)/Testis-DMRT6-ChIP-Seq(GSE60440)/Homer	YDGHTACAWTGTADC	1.00E-02	-5.45E+00	0.0073	154	3.39%	1220.4	2.73%
n-Myc(bHLH)/mES-nMyc-ChIP-Seq(GSE11431)/Homer	VRCCACGTGG	1.00E-02	-5.33E+00	0.0082	800	17.62%	7242.7	16.18%
CEBP(bZIP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer	ATTGCGCAAC	1.00E-02	-5.06E+00	0.0107	556	12.25%	4950.5	11.06%
EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	AVCAGGAAGT	1.00E-02	-4.87E+00	0.0129	1340	29.52%	12479.9	27.89%
MafF(bZIP)/HepG2-MafF-ChIP-Seq(GSE31477)/Homer	HWWGTCAGCAWWTT	1.00E-02	-4.82E+00	0.0136	273	6.01%	2325.5	5.20%
KLF10(Zf)/HEK293-KLF10.GFP-ChIP-Seq(GSE58341)/Homer	GGGGGTGTGTCC	1.00E-02	-4.79E+00	0.0139	628	13.83%	5652.3	12.63%
c-Myc(bHLH)/LNCAP-cMyc-ChIP-Seq(Unpublished)/Homer	VCCACGTG	1.00E-02	-4.66E+00	0.0158	441	9.71%	3897.4	8.71%
Etv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer(0.967)	NNAYTTCCTGHN	1.00E-02	-4.64E+00	0.016	1019	22.44%	9403.1	21.01%
ZNF711(Zf)/SHSY5Y-ZNF711-ChIP-Seq(GSE20673)/Homer	AGGCCTAG	1.00E-01	-4.59E+00	0.0168	2152	47.40%	20441.9	45.68%
Fli1(ETS)/CD8-FLI-ChIP-Seq(GSE20898)/Homer	NRYTTCCGGH	1.00E-01	-4.55E+00	0.0173	1130	24.89%	10482.3	23.42%
CArG(MADS)/PUER-Srf-ChIP-Seq(Sullivan_et_al.)/Homer	CCATATATGGNM	1.00E-01	-4.54E+00	0.0174	307	6.76%	2654.4	5.93%
HIF2a(bHLH)/785_O-HIF2a-ChIP-Seq(GSE34871)/Homer	GCACGTACCC	1.00E-01	-4.35E+00	0.0209	333	7.33%	2907.2	6.50%
ZNF317(Zf)/HEK293-ZNF317.GFP-ChIP-Seq(GSE58341)/Homer	GTCWGCTGTYYCTCT	1.00E-01	-4.32E+00	0.0215	110	2.42%	870.4	1.94%
Zfp809(Zf)/ES-Zfp809-ChIP-Seq(GSE70799)/Homer	GGGGCTYGKCTGGGA	1.00E-01	-4.22E+00	0.0236	364	8.02%	3205.5	7.16%
NFAT:AP1(RHD,bZIP)/Jurkat-NFATC1-ChIP-Seq(Jolma_et_al.)/Hon	SARTGGAAAAWRTGA	1.00E-01	-4.17E+00	0.0248	174	3.83%	1451.4	3.24%
IRF:BATF(IRF:bZIP)/pDC-Irf8-ChIP-Seq(GSE66899)/Homer	CTTTCANTATGACTV	1.00E-01	-4.13E+00	0.0258	109	2.40%	868.3	1.94%
GATA3(Zf),DR8/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	AGATSTNDNNDSAGAT	1.00E-01	-4.11E+00	0.0261	72	1.59%	544.9	1.22%
BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	DATGASTCAT	1.00E-01	-4.07E+00	0.0271	569	12.53%	5151.5	11.51%
HIF-1a(bHLH)/MCF7-HIF1a-ChIP-Seq(GSE28352)/Homer	TACGTGCV	1.00E-01	-4.05E+00	0.0275	248	5.46%	2136.4	4.77%
c-Myc(bHLH)/mES-cMyc-ChIP-Seq(GSE11431)/Homer	VVCCACGTGG	1.00E-01	-4.05E+00	0.0275	607	13.37%	5514.7	12.32%
AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	VTGACTCATC	1.00E-01	-3.83E+00	0.0341	666	14.67%	6097.4	13.62%
OCT:OCT(POU,Homeobox)/NPC-Brn1-ChIP-Seq(GSE35496)/Home	ATGAATATTCATGAG	1.00E-01	-3.82E+00	0.0341	11	0.24%	54	0.12%
ZNF519(Zf)/HEK293-ZNF519.GFP-ChIP-Seq(GSE58341)/Homer	GAGSCCGAGC	1.00E-01	-3.81E+00	0.0345	272	5.99%	2373.9	5.30%
CRE(bZIP)/Promoter/Homer	CSGTGACGTCAC	1.00E-01	-3.79E+00	0.0349	193	4.25%	1643.3	3.67%
Erra(NR)/HepG2-Erra-ChIP-Seq(GSE31477)/Homer	CAAAGGTCAG	1.00E-01	-3.74E+00	0.0365	2237	49.27%	21389.5	47.79%
bHLHE40(bHLH)/HepG2-BHLHE40-ChIP-Seq(GSE31477)/Homer	KCACGTGMCN	1.00E-01	-3.68E+00	0.0385	300	6.61%	2642.2	5.90%

Motif	Consensus	p-value	log p	q-value	# Target	% Target	# Bg	% Bg
JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	RATGASTCAT	1.00E-01	-3.63E+00	0.0403	495	10.90%	4485.6	10.02%
EWS:FLI1-fusion(ETS)/SK_N_MC-EWS:FLI1-ChIP-Seq(SRA014231)	VACAGGAAAT	1.00E-01	-3.63E+00	0.0403	604	13.30%	5523.6	12.34%
TCFL2(HMG)/K562-TCF7L2-ChIP-Seq(GSE29196)/Homer	ACWTCAAAGG	1.00E-01	-3.61E+00	0.0409	80	1.76%	629.9	1.41%
BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer	CNNBRGCGCCCCCTGS	1.00E-01	-3.57E+00	0.0422	252	5.55%	2201.9	4.92%
IRF3(IRF)/BMDM-Irf3-ChIP-Seq(GSE67343)/Homer	AGTTTCAKTTTC	1.00E-01	-3.47E+00	0.0464	278	6.12%	2450.1	5.47%
ZNF692(Zf)/HEK293-ZNF692.GFP-ChIP-Seq(GSE58341)/Homer	GTGGGCCCCA	1.00E-01	-3.40E+00	0.0496	250	5.51%	2193.2	4.90%
Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer	DATGASTCATHN	1.00E-01	-3.36E+00	0.0515	574	12.64%	5262.3	11.76%
THRb(NR)/Liver-NR1A2-ChIP-Seq(GSE52613)/Homer	TRAGGTCA	1.00E-01	-3.36E+00	0.0515	3292	72.51%	31903.4	71.28%
EWS:ERG-fusion(ETS)/CADO_ES1-EWS:ERG-ChIP-Seq(SRA014231	ATTTCCTGTN	1.00E-01	-3.34E+00	0.0522	761	16.76%	7057	15.77%
ETV1(ETS)/GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	AACCGGAAGT	1.00E-01	-3.29E+00	0.0545	1397	30.77%	13224.6	29.55%
E2F1(E2F)/Hela-E2F1-ChIP-Seq(GSE22478)/Homer	CWGGCGGGAA	1.00E-01	-3.23E+00	0.0577	168	3.70%	1442.8	3.22%
CTCF-SatelliteElement(Zf?)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/H	TGCAGTTCCMVNWRT	1.00E-01	-3.11E+00	0.0652	21	0.46%	137.4	0.31%
EBNA1(EBV-virus)/Raji-EBNA1-ChIP-Seq(GSE30709)/Homer	GGYAGCAYDTGCTDCC	1.00E-01	-3.10E+00	0.0653	27	0.59%	186.1	0.42%
ZNF41(Zf)/HEK293-ZNF41.GFP-ChIP-Seq(GSE58341)/Homer	CCTCATGGTGYCYTWY	1.00E-01	-3.09E+00	0.0656	37	0.81%	270.7	0.60%
Chop(bZIP)/MEF-Chop-ChIP-Seq(GSE35681)/Homer	ATTGCATCAT	1.00E-01	-2.95E+00	0.0752	194	4.27%	1699.7	3.80%
NFkB-p65-Rel(RHD)/ThioMac-LPS-Expression(GSE23622)/Homer	GGAAATTCCC	1.00E-01	-2.95E+00	0.0754	67	1.48%	536.1	1.20%
ELF3(ETS)/PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	ANCAGGAAGT	1.00E-01	-2.93E+00	0.0764	766	16.87%	7152.9	15.98%
NFkB-p65(RHD)/GM12787-p65-ChIP-Seq(GSE19485)/Homer	WGGGGATTTCCC	1.00E-01	-2.85E+00	0.0822	613	13.50%	5688.1	12.71%
RXR(NR),DR1/3T3L1-RXR-ChIP-Seq(GSE13511)/Homer	TAGGGCAAAGGTCA	1.00E-01	-2.85E+00	0.0822	1410	31.06%	13415.5	29.98%
IRF2(IRF)/Erythroblas-IRF2-ChIP-Seq(GSE36985)/Homer	GAAASYGAAASY	1.00E-01	-2.74E+00	0.0914	90	1.98%	751.5	1.68%
Pax7(Paired,Homeobox),longest/Myoblast-Pax7-ChIP-Seq(GSE250	NTAATTDGCYAATTAN	1.00E-01	-2.71E+00	0.0933	25	0.55%	177.4	0.40%
Klf9(Zf)/GBM-Klf9-ChIP-Seq(GSE62211)/Homer	GCCACRCCCACY	1.00E-01	-2.66E+00	0.0979	397	8.74%	3639.2	8.13%
IRF4(IRF)/GM12878-IRF4-ChIP-Seq(GSE32465)/Homer	ACTGAAACCA	1.00E-01	-2.55E+00	0.1092	392	8.63%	3602.1	8.05%
ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	ACAGGAAGTG	1.00E-01	-2.46E+00	0.1191	1072	23.61%	10181.7	22.75%
ARE(NR)/LNCAP-AR-ChIP-Seq(GSE27824)/Homer	RGRACASNSTGTYCYB	1.00E+00	-2.30E+00	0.1388	276	6.08%	2519.8	5.63%
IRF1(IRF)/PBMC-IRF1-ChIP-Seq(GSE43036)/Homer	GAAAGTGAAAGT	1.00E+00	-2.28E+00	0.1415	124	2.73%	1087.3	2.43%
PRDM14(Zf)/H1-PRDM14-ChIP-Seq(GSE22767)/Homer	RGGTCTCTAACY	1.00E+00	-2.22E+00	0.1487	303	6.67%	2783.5	6.22%
RORgt(NR)/EL4-RORgt.Flag-ChIP-Seq(GSE56019)/Homer	AAYTAGGTCA	1.00E+00	-2.22E+00	0.1487	123	2.71%	1081.3	2.42%
E2F(E2F)/Hela-CellCycle-Expression/Homer	TTSGCGCGAAAA	1.00E+00	-2.22E+00	0.1487	37	0.81%	293.8	0.66%
NFkB-p50,p52(RHD)/Monocyte-p50-ChIP-Chip(Schreiber_et_al.)/H	GGGGGAATCCCC	1.00E+00	-2.16E+00	0.1568	156	3.44%	1394.8	3.12%
DUX4(Homeobox)/Myoblasts-DUX4.V5-ChIP-Seq(GSE75791)/Hom	NWTAAYCYAATCAWN	1.00E+00	-2.13E+00	0.1611	35	0.77%	278.5	0.62%
GRHL2(CP2)/HBE-GRHL2-ChIP-Seq(GSE46194)/Homer	AAACYKGTTWDACMR(	1.00E+00	-2.07E+00	0.1708	338	7.44%	3133.5	7.00%
SpiB(ETS)/OCILY3-SPIB-ChIP-Seq(GSE56857)/Homer	AAAGRGGAAGTG	1.00E+00	-2.04E+00	0.1744	227	5.00%	2075.6	4.64%
ZNF264(Zf)/HEK293-ZNF264.GFP-ChIP-Seq(GSE58341)/Homer	RGGGCACTAACY	1.00E+00	-2.03E+00	0.176	674	14.85%	6379.2	14.25%
Sp1(Zf)/Promoter/Homer	GGCCCCGCCCCC	1.00E+00	-2.03E+00	0.176	178	3.92%	1611.7	3.60%
REST-NRSF(Zf)/Jurkat-NRSF-ChIP-Seq/Homer	GGMGCTGTCCATGGT(	1.00E+00	-1.71E+00	0.2394	12	0.26%	87.4	0.20%
ZNF675(Zf)/HEK293-ZNF675.GFP-ChIP-Seq(GSE58341)/Homer	ARGAGGMCAAAATGV	1.00E+00	-1.71E+00	0.2394	147	3.24%	1341.9	3.00%
Nr5a2(NR)/Pancreas-LRH1-ChIP-Seq(GSE34295)/Homer	BTCAAGGTCA	1.00E+00	-1.70E+00	0.2413	774	17.05%	7402.1	16.54%
Nr5a2(NR)/mES-Nr5a2-ChIP-Seq(GSE19019)/Homer	BTCAAGGTCA	1.00E+00	-1.65E+00	0.2514	577	12.71%	5493.9	12.28%
Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	HACTTCCGGY	1.00E+00	-1.64E+00	0.2554	485	10.68%	4603.4	10.29%
SPDEF(ETS)/VCaP-SPDEF-ChIP-Seq(SRA014231)/Homer	ASWTCCTGBT	1.00E+00	-1.61E+00	0.2599	1065	23.46%	10258.9	22.92%
CEBP:CEBP(bZIP)/MEF-Chop-ChIP-Seq(GSE35681)/Homer	NTNATGCAAYMNNHT	1.00E+00	-1.61E+00	0.261	119	2.62%	1083.4	2.42%
ZNF382(Zf)/HEK293-ZNF382.GFP-ChIP-Seq(GSE58341)/Homer	GNCTGTASTRNTGBCT	1.00E+00	-1.54E+00	0.2767	30	0.66%	252.8	0.56%
GFY(?)/Promoter/Homer	ACTACAATTCCC	1.00E+00	-1.49E+00	0.2906	55	1.21%	486.5	1.09%
PGR(NR)/EndoStromal-PGR-ChIP-Seq(GSE69539)/Homer	AAGAACATWHTGTTC	1.00E+00	-1.46E+00	0.2975	194	4.27%	1812.8	4.05%
TFE3(bHLH)/MEF-TFE3-ChIP-Seq(GSE75757)/Homer	GTCACGTGACYV	1.00E+00	-1.38E+00	0.3228	60	1.32%	539.1	1.20%
T1ISRE(IRF)/ThioMac-Ifnb-Expression/Homer	ACTTTCGTTTCT	1.00E+00	-1.25E+00	0.3643	11	0.24%	88.5	0.20%
MafK(bZIP)/C2C12-MafK-ChIP-Seq(GSE36030)/Homer	GCTGASTCAGCA	1.00E+00	-1.20E+00	0.384	230	5.07%	2189.2	4.89%
GABPA(ETS)/Jurkat-GABPa-ChIP-Seq(GSE17954)/Homer	RACCGGAAGT	1.00E+00	-1.18E+00	0.388	881	19.41%	8548.6	19.10%
IRF8(IRF)/BMDM-IRF8-ChIP-Seq(GSE77884)/Homer	GRAASTGAAAST	1.00E+00	-1.17E+00	0.392	281	6.19%	2688	6.01%
Srebp1a(bHLH)/HepG2-Srebp1a-ChIP-Seq(GSE31477)/Homer	RTCACSCCAY	1.00E+00	-1.14E+00	0.403	242	5.33%	2313.6	5.17%
PU.1:IRF8(ETS:IRF)/pDC-Irf8-ChIP-Seq(GSE66899)/Homer	GGAAGTGAAAST	1.00E+00	-1.08E+00	0.4249	182	4.01%	1737.4	3.88%
GRE(NR),IR3/RAW264.7-GRE-ChIP-Seq(Unpublished)/Homer	VAGRACAKWCTGTYC	1.00E+00	-1.08E+00	0.4249	211	4.65%	2019.6	4.51%
ERE(NR),IR3/MCF7-ERa-ChIP-Seq(Unpublished)/Homer	VAGGTCACNSTGACC	1.00E+00	-1.01E+00	0.4521	301	6.63%	2906.2	6.49%
ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	ACVAGGAAGT	1.00E+00	-9.78E-01	0.4671	711	15.66%	6929.3	15.48%
ZBTB12(Zf)/HEK293-ZBTB12.GFP-ChIP-Seq(GSE58341)/Homer	NGNTCTAGAACCNGV	1.00E+00	-9.67E-01	0.4709	407	8.96%	3951.3	8.83%
SF1(NR)/H295R-Nr5a1-ChIP-Seq(GSE44220)/Homer	CAAGGHCANV	1.00E+00	-8.67E-01	0.5187	492	10.84%	4805.8	10.74%
GATA(Zf),IR4/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	NAGATWNBNATCTNN	1.00E+00	-8.55E-01	0.5232	64	1.41%	613.9	1.37%
GRE(NR),IR3/A549-GR-ChIP-Seq(GSE32465)/Homer	NRGVACABNVTGTYCY	1.00E+00	-8.26E-01	0.5367	120	2.64%	1163.2	2.60%
ELF1(ETS)/Jurkat-ELF1-ChIP-Seq(SRA014231)/Homer	AVCCGGAAGT	1.00E+00	-8.21E-01	0.5374	441	9.71%	4314.2	9.64%
Mouse_Recombination_Hotspot(Zf)/Testis-DMC1-ChIP-Seq(GSE24	ACTYKNATTCGTGNTA(	1.00E+00	-8.02E-01	0.5461	53	1.17%	510.1	1.14%
KLF3(Zf)/MEF-Klf3-ChIP-Seq(GSE44748)/Homer	NRGCCCCRCCCHBNN	1.00E+00	-7.95E-01	0.5481	476	10.48%	4664.8	10.42%
GATA3(Zf),DR4/iTreg-Gata3-ChIP-Seq(GSE20898)/Homer	AGATGKDGAGATAAG	1.00E+00	-7.58E-01	0.5666	61	1.34%	592.1	1.32%
RUNX-AML(Runt)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	GCTGTGGTTW	1.00E+00	-7.51E-01	0.5689	718	15.81%	7057.4	15.77%
OCT:OCT(POU,Homeobox)/NPC-OCT6-ChIP-Seq(GSE43916)/Home	YATGCATATRCATRT	1.00E+00	-7.03E-01	0.5947	68	1.50%	666.4	1.49%
STAT6(Stat)/CD4-Stat6-ChIP-Seq(GSE22104)/Homer	ABTTCYYRRGAA	1.00E+00	-7.01E-01	0.5947	488	10.75%	4805.7	10.74%
Egr2(Zf)/Thymocytes-Egr2-ChIP-Seq(GSE34254)/Homer	NGCGTGGGCGGR	1.00E+00	-6.90E-01	0.5985	148	3.26%	1456.8	3.26%
RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq(GSE29180)/Homer	AAACCACARM	1.00E+00	-6.82E-01	0.6015	997	21.96%	9828.7	21.96%
E-box(bHLH)/Promoter/Homer	SSGGTCACGTGA	1.00E+00	-6.38E-01	0.6263	55	1.21%	544.4	1.22%
ISRE(IRF)/ThioMac-LPS-Expression(GSE23622)/Homer	AGTTTCASTTTC	1.00E+00	-6.38E-01	0.6263	45	0.99%	445.7	1.00%
FOXA1:AR(Forkhead,NR)/LNCAP-AR-ChIP-Seq(GSE27824)/Homer	AGTAAACAAAAAAGAA	1.00E+00	-5.95E-01	0.6496	41	0.90%	409.4	0.91%
LXRE(NR),DR4/RAW-LXRb.biotin-ChIP-Seq(GSE21512)/Homer	RGGTTACTANAGGTCA	1.00E+00	-5.41E-01	0.6837	42	0.93%	424.8	0.95%
PPARE(NR),DR1/3T3L1-Pparg-ChIP-Seq(GSE13511)/Homer	TGACCTTTGCCCCA	1.00E+00	-5.35E-01	0.6856	1125	24.78%	11148.8	24.91%
Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	GGATGACTCATC	1.00E+00	-5.31E-01	0.6858	394	8.68%	3922.6	8.76%
PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	AGAGGAAGTG	1.00E+00	-5.22E-01	0.6902	485	10.68%	4826.5	10.78%
ZNF528(Zf)/HEK293-ZNF528.GFP-ChIP-Seq(GSE58341)/Homer	AGAAATGACTTCCCT	1.00E+00	-5.20E-01	0.6902	7	0.15%	72.5	0.16%
Nur77(NR)/K562-NR4A1-ChIP-Seq(GSE31363)/Homer	TGACCTTTNCNT	1.00E+00	-5.15E-01	0.6908	146	3.22%	1465.9	3.28%
STAT5(Stat)/mCD4+-Stat5-ChIP-Seq(GSE12346)/Homer	RTTTCTNAGAAA	1.00E+00	-5.11E-01	0.6909	296	6.52%	2956.5	6.61%
Esrrb(NR)/mES-Esrrb-ChIP-Seq(GSE11431)/Homer	KTGACCTTGA	1.00E+00	-4.11E-01	0.761	623	13.72%	6234.5	13.93%
Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	NRYTTCCGGY	1.00E+00	-4.06E-01	0.7625	450	9.91%	4518.3	10.10%
VDR(NR),DR3/GM10855-VDR+vitD-ChIP-Seq(GSE22484)/Homer	ARAGGTCANWGAGTT	1.00E+00	-3.77E-01	0.7827	232	5.11%	2355.6	5.26%

ETS(ETS)/Promoter/Homer	AACCGGAAGT	1.00E+00	-3.07E-01	0.8372	260	5.73%	2658.5	5.94%
Ets1-distal(ETS)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	MACAGGAAGT	1.00E+00	-2.98E-01	0.8419	287	6.32%	2932.8	6.55%
Srebp2(bHLH)/HepG2-Srebp2-ChIP-Seq(GSE31477)/Homer	CGGTCACSCCAC	1.00E+00	-2.45E-01	0.885	151	3.33%	1580.9	3.53%
GFX(?)/Promoter/Homer	ATTCTCGCGAGA	1.00E+00	-2.33E-01	0.8926	2	0.04%	29.1	0.07%
E2F7(E2F)/Hela-E2F7-ChIP-Seq(GSE32673)/Homer	VDTTTCCCGCCA	1.00E+00	-2.31E-01	0.8926	68	1.50%	735.7	1.64%
GFY-Staf(?,Zf)/Promoter/Homer	RACTACAATTCCCAGAA	1.00E+00	-2.20E-01	0.8988	25	0.55%	287.5	0.64%
Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	NATGASTCABNN	1.00E+00	-1.97E-01	0.9166	263	5.79%	2734.9	6.11%
RAR:RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	RGGTCADNNAGAGGT	1.00E+00	-1.97E-01	0.9166	31	0.68%	355.5	0.79%
ZBTB33(Zf)/GM12878-ZBTB33-ChIP-Seq(GSE32465)/Homer	GGVTCTCGCGAGAAC	1.00E+00	-1.90E-01	0.9175	10	0.22%	127.4	0.28%
CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer	AYAGTGCCMYCTRGTG	1.00E+00	-1.73E-01	0.9311	123	2.71%	1320.7	2.95%
STAT6(Stat)/Macrophage-Stat6-ChIP-Seq(GSE38377)/Homer	TTCCKNAGAA	1.00E+00	-1.54E-01	0.9457	477	10.51%	4919.1	10.99%
THRa(NR)/C17.2-THRa-ChIP-Seq(GSE38347)/Homer	GGTCANYTGAGGWCA	1.00E+00	-1.47E-01	0.9493	485	10.68%	5005.4	11.18%
STAT1(Stat)/HelaS3-STAT1-ChIP-Seq(GSE12782)/Homer	NATTTCCNGGAAAT	1.00E+00	-1.17E-01	0.9758	240	5.29%	2549.9	5.70%
RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer	SAAACCACAG	1.00E+00	-1.10E-01	0.9794	669	14.74%	6891.3	15.40%
Reverb(NR),DR2/RAW-Reverba.biotin-ChIP-Seq(GSE45914)/Home	GTRGGTCASTGGGTCA	1.00E+00	-1.07E-01	0.9794	153	3.37%	1662.1	3.71%
RARg(NR)/ES-RARg-ChIP-Seq(GSE30538)/Homer	AGGTCAAGGTCA	1.00E+00	-9.03E-02	0.9928	26	0.57%	327.5	0.73%
HNF4a(NR),DR1/HepG2-HNF4a-ChIP-Seq(GSE25021)/Homer	CARRGKBCAAAGTYCA	1.00E+00	-7.46E-02	1	480	10.57%	5032.2	11.24%
GLI3(Zf)/Limb-GLI3-ChIP-Chip(GSE11077)/Homer	CGTGGGTGGTCC	1.00E+00	-7.43E-02	1	149	3.28%	1645.5	3.68%
ZNF669(Zf)/HEK293-ZNF669.GFP-ChIP-Seq(GSE58341)/Homer	GARTGGTCATCGCCC	1.00E+00	-6.94E-02	1	63	1.39%	741.5	1.66%
ETS:RUNX(ETS,Runt)/Jurkat-RUNX1-ChIP-Seq(GSE17954)/Homer	RCAGGATGTGGT	1.00E+00	-5.20E-02	1	86	1.89%	1001.9	2.24%
NRF(NRF)/Promoter/Homer	STGCGCATGCGC	1.00E+00	-4.84E-02	1	80	1.76%	940.8	2.10%
RUNX2(Runt)/PCa-RUNX2-ChIP-Seq(GSE33889)/Homer	NWAACCACADNN	1.00E+00	-4.82E-02	1	758	16.70%	7890	17.63%
Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	GATGASTCATCN	1.00E+00	-4.33E-02	1	173	3.81%	1930.5	4.31%
HRE(HSF)/HepG2-HSF1-ChIP-Seq(GSE31477)/Homer	BSTTCTRGAABVTTCYA	1.00E+00	-4.16E-02	1	142	3.13%	1607.9	3.59%
Klf4(Zf)/mES-Klf4-ChIP-Seq(GSE11431)/Homer	GCCACACCCA	1.00E+00	-3.82E-02	1	378	8.33%	4058.7	9.07%
RAR:RXR(NR),DR5/ES-RAR-ChIP-Seq(GSE56893)/Homer	AGGTCAAGGTCA	1.00E+00	-3.77E-02	1	98	2.16%	1145.1	2.56%
ZNF16(Zf)/HEK293-ZNF16.GFP-ChIP-Seq(GSE58341)/Homer	MACCTTCYATGGCTCC	1.00E+00	-3.29E-02	1	11	0.24%	176.2	0.39%
EKLF(Zf)/Erythrocyte-Klf1-ChIP-Seq(GSE20478)/Homer	NWGGGTGTGGCY	1.00E+00	-3.05E-02	1	223	4.91%	2475.6	5.53%
Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	AWWNTGCTGAGTCAT	1.00E+00	-2.34E-02	1	33	0.73%	447.3	1.00%
Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	TGCTGAGTCA	1.00E+00	-2.10E-02	1	157	3.46%	1804	4.03%
LRF(Zf)/Erythroblasts-ZBTB7A-ChIP-Seq(GSE74977)/Homer	AAGACCCYYN	1.00E+00	-1.93E-02	1	1626	35.81%	16689	37.29%
Ronin(THAP)/ES-Thap11-ChIP-Seq(GSE51522)/Homer	RACTACAACTCCCAGVA	1.00E+00	-1.83E-02	1	7	0.15%	134.6	0.30%
NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	GATGACTCAGCA	1.00E+00	-1.44E-02	1	35	0.77%	484.7	1.08%
FXR(NR),IR1/Liver-FXR-ChIP-Seq(Chong_et_al.)/Homer	AGGTCANTGACCTB	1.00E+00	-1.25E-02	1	438	9.65%	4767.5	10.65%
TR4(NR),DR1/Hela-TR4-ChIP-Seq(GSE24685)/Homer	GAGGTCAAAGGTCA	1.00E+00	-8.28E-03	1	96	2.11%	1191	2.66%
NRF1(NRF)/MCF7-NRF1-ChIP-Seq(Unpublished)/Homer	CTGCGCATGCGC	1.00E+00	-6.46E-03	1	35	0.77%	506.4	1.13%
Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	HTGCTGAGTCAT	1.00E+00	-1.00E-03	1	26	0.57%	439.5	0.98%
Gfi1b(Zf)/HPC7-Gfi1b-ChIP-Seq(GSE22178)/Homer	MAATCACTGC	1.00E+00	-7.28E-04	1	548	12.07%	6113	13.66%
HRE(HSF)/Striatum-HSF1-ChIP-Seq(GSE38000)/Homer	TTCTAGAABNTTCTA	1.00E+00	-6.54E-04	1	163	3.59%	2032.7	4.54%
YY1(Zf)/Promoter/Homer	CAAGATGGCGGC	1.00E+00	-8.20E-05	1	24	0.53%	462.7	1.03%
TEAD4(TEA)/Tropoblast-Tead4-ChIP-Seq(GSE37350)/Homer	CCWGGAATGY	1.00E+00	-2.00E-06	1	717	15.79%	8228.2	18.39%
TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	YCWGGAATGY	1.00E+00	-2.00E-06	1	524	11.54%	6206.3	13.87%
COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer	AGRGGTCA	1.00E+00	0.00E+00	1	1516	33.39%	16643.4	37.19%
TEAD2(TEA)/Py2T-Tead2-ChIP-Seq(GSE55709)/Homer	CCWGGAATGY	1.00E+00	0.00E+00	1	384	8.46%	5067.8	11.32%
ZFP3(Zf)/HEK293-ZFP3.GFP-ChIP-Seq(GSE58341)/Homer	GGGTTTTGAAGGATGA	1.00E+00	0.00E+00	1	0	0.00%	24.9	0.06%

Supplemental table 3. Characteristics of the ASCL1 and NEUROD1 ChIP-seq datasets

Sample	ChIP-seq	Uniquely Mapped Reads	Total_Peaks	DHS (per 5000 peaks)	FRiP
LuCaP49	ASCL1	26093095	46859	4732	32.5
LuCaP93	ASCL1	25852594	49994	4664	21.1
LuCaP145_1	ASCL1	38818608	46303	4259	38.9
H660	ASCL1	24979578	61338	4869	34.4
LuCaP173_p1	NEUROD1	31869709	78124	4857	35.1

DHS: DNase I hypersensitive site (per 5000 peaks)
FRIP: fraction of reads that fall into a peak