



HHS Public Access

Author manuscript

Nat Cell Biol. Author manuscript; available in PMC 2022 April 21.

Published in final edited form as:

Nat Cell Biol. 2021 November ; 23(11): 1199–1211. doi:10.1038/s41556-021-00766-y.

Dynamic transcriptional reprogramming leads to immunotherapeutic vulnerabilities in myeloma

Julia Frede^{1,2,3}, **Praveen Anand**^{1,2,3}, **Noori Sotudeh**^{1,2,3}, **Ricardo A. Pinto**^{3,4,5}, **Monica S. Nair**¹, **Hannah Stuart**¹, **Andrew J. Yee**^{2,6}, **Tushara Vijaykumar**¹, **Johannes M. Waldschmidt**^{1,2,3}, **Sayalee Potdar**⁷, **Jake A. Kloeber**¹, **Antonios Kokkalis**^{1,2,3}, **Valeriya Dimitrova**^{2,3,7}, **Mason Mann**⁶, **Jacob P. Laubach**^{1,2}, **Paul G. Richardson**^{1,2}, **Kenneth C. Anderson**^{1,2}, **Noopur S. Raje**^{2,6}, **Birgit Knoechel**^{2,3,7,8,*}, **Jens G. Lohr**^{1,2,3,8,*}

¹Department of Medical Oncology, Jerome Lipper Multiple Myeloma Center, Dana Farber Cancer Institute, Boston, MA, USA

²Harvard Medical School, Boston, MA, USA

³Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁴Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, USA

⁵Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

⁶Massachusetts General Hospital, Boston, MA, USA

⁷Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

⁸These authors jointly supervised this work.

Summary

While there is extensive evidence for genetic variation as a basis for treatment resistance, other sources of variation result from cellular plasticity. Using multiple myeloma as an example of an incurable lymphoid malignancy, we show how cancer cells modulate lineage restriction, adapt their enhancer usage and employ cell-intrinsic diversity for survival and treatment escape. By using single cell transcriptome and chromatin accessibility profiling, we show that distinct transcriptional states co-exist in individual cancer cells and that differential transcriptional regulon usage and enhancer rewiring underlie these alternate transcriptional states. We demonstrate that exposure to standard treatment further promotes transcriptional reprogramming and differential

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

*Correspondence: jensg_lohr@dfci.harvard.edu, birgit_knoechel@dfci.harvard.edu.

Author Contributions

J.F. designed and performed experiments, analyzed data, and wrote the manuscript. P.A., R. P.R., N.S. and H.S. performed data analysis. M.S.N. performed experiments and data analysis. T.V., J.M.W., S.P., J.A.K., A.K., and V.D. performed sample processing, sequencing or provided scientific input. N.S.R. and A.J.Y. provided project leadership, coordinated sample acquisition and clinical data integration. M.M., K.C.A., P.G.R. and J.P.L. oversaw clinical data analysis and integration. B.K. and J.G.L. designed experiments, provided project leadership, supervised the analysis and wrote the manuscript. All authors discussed the results and reviewed the manuscript.

Declaration of Interests

A.J.Y. has a consulting role for Adaptive Biotechnologies, Amgen, BMS, Celgene, GSK, Janssen, Karyopharm, Sanofi and Takeda. The remaining authors declare no competing interests.

enhancer recruitment, while simultaneously reducing developmental potential. Importantly, treatment generates a distinct complement of actionable immunotherapy targets, such as CXCR4, which can be exploited to overcome treatment resistance. Our studies therefore delineate how to transform the cellular plasticity that underlies drug resistance into immuno-oncologic therapeutic opportunities.

Introduction

Although most cancer patients respond to initial treatment, responses are often short-lived as drug resistance frequently develops. Non-genetic cellular plasticity and adaptive state changes have recently emerged as a basis for therapeutic resistance in cancer¹⁻³, and a better understanding of how cell state transitions are regulated is critical to develop therapeutic approaches that can overcome drug resistance. These transitions are mediated by dynamic transcriptional changes and can involve epigenetic remodeling of the open chromatin landscape, or changes in the activity of transcriptional regulators⁴⁻⁷.

To investigate lineage infidelity and drug resistance at the single cell level, we studied multiple myeloma (MM), a malignancy of antibody-producing plasma cells in the bone marrow. The notion that plasma cells are terminally differentiated cells has long prevented intensive research into phenotypic plasticity and differentiation in multiple myeloma; however, it has recently been demonstrated that normal and malignant plasma cells differ in their differentiation state^{8,9}. Furthermore, it has been suggested that more immature states may correlate with drug resistance⁸. Yet, it remains unclear how heterogeneous differentiation states are regulated at the transcriptional and epigenetic level, and whether they change with treatment.

Here we use single cell transcriptome and chromatin accessibility profiles to define the transcriptional and cis-regulatory alterations that underlie cell state transitions in malignant plasma cells. We infer gene regulatory network configurations during the establishment of distinct cell states, highlighting regulatory factors driving phenotypic changes. These phenotypic changes in turn provide the rationale for targeted therapeutic strategies to overcome drug resistance.

Results

Coexisting transcriptional programs in single myeloma cells.

To define transcriptional states in myeloma at single cell resolution, we performed full-length single cell RNA sequencing of myeloma cells and CD45+ immune cells from the bone marrow or blood of 8 patients with relapsed/refractory multiple myeloma (RRMM) and two normal donors by SmartSeq2¹⁰ (Fig. 1a, Extended Data Fig. 1). Patients were treated with elotuzumab, pomalidomide, bortezomib and dexamethasone on a clinical trial (Elo-PVD; [NCT02718833](#); Supplementary Table 1), and we obtained bone marrow samples at screening and following treatment. A total of 6,955 cells were analyzed and underwent QC filtering (Extended Data Fig. 2, Supplementary Table 2). As an initial examination of transcriptional heterogeneity, we performed PAGODA2 clustering¹¹ (Fig.

1b,c). Reference-based annotation of the single cell profiles using a dataset of 21 immune cell populations¹² showed that clusters 20,1,5,7,14,12,3, and 13 corresponded to CD4+ and CD8+ T cells, NK cells, B cells, monocytes, and neutrophils, respectively, while clusters 2,4,6,8,10,11,15,16,17, and 19 were annotated as plasma cells (Fig. 1d).

We next set out to distinguish malignant from non-malignant cells. We estimated copy number variations (CNVs), which occur in virtually all myeloma patients. CNVs were highly enriched in presumed malignant clusters, highly concordant with results from clinical cytogenetics (Extended Data Fig. 3a, Supplementary Table 3), and were distinct between individual clusters (Fig. 1e, Extended Data Fig. 3b). As the immunoglobulin sequences are typically monoclonal or rarely oligoclonal in myeloma cells, we reconstructed the CDR3 region for each cell. Indeed, presumed malignant cells from individual patients displayed a high degree of clonality, with the most common CDR3 sequence found in the majority of cells (Fig. 1e). As MM cells frequently harbor translocations involving the immunoglobulin heavy chain (IgH) locus, we next assessed expression of translocation target genes *CCND1* and *MAF* and detected overexpression in clusters 6,17 and 2, and cluster 8, respectively, concordant with FISH results (Fig. 1e, Extended Data Fig. 3a).

We then employed a random forest (RF) model to determine genes that best discriminate presumed malignant cells and normal plasma cells (Fig. 1e, Extended Data Fig. 3c–e). The RF model identified several genes which have previously been described as drivers in MM, such as *CCND1* and the modulator of Wnt signaling *FRZB*¹³. Myeloma marker genes, which thus far have not been described in MM, include the cytoskeletal linker protein dystonin (*DST*). Using an aggregate approach combining myeloma-specific gene expression with copy number and clonotype analysis, we were able to robustly distinguish myeloma cells from normal plasma cells, even within patients, including cases for which PAGODA2 clustering alone was ambiguous (Fig. 1f, Extended Data Fig. 3f,g).

Characterization of malignant clusters through marker gene analysis identified known myeloma driver genes, such as *MAF* consistent with a t(14;16) translocation in MM7 (Extended Data Fig. 3h; Supplementary Table 4). Interestingly, some driver genes were highly expressed in more than one MM cluster, e.g. *CCND1* in 3/8 patients (Fig. 1e). We therefore asked if the distinct clustering of myeloma cells from patients with the same established genomic classifier t(11;14), which results in overexpression of *CCND1* (Extended Data Fig. 3a,d,h) could be due to the presence of multiple gene expression signatures in the same cells. We indeed observed that individuals with *CCND1* translocation expressed the corresponding *CDI* signature, but further manifested expression of distinct other signatures, such as the *NFKB* signature (MM1) or the proliferation signature *PR* (MM8) (Fig. 1g). We detected co-expression of unrelated signatures in the same single cells in each individual, arguing that there are widespread transcriptional differences not captured by established gene expression classifiers¹⁴.

To investigate transcriptional heterogeneity in greater detail, we first assessed the effect of the cell cycle as a potential source of heterogeneity. We did not observe cell cycle driven effects on PAGODA2 clustering with the exception of cluster 10, which consisted of cycling cells derived from multiple patients (Fig. 2a,b). We next scored expression of 11 recurrent

heterogeneity programs identified across 198 cell lines reflecting 22 cancer types¹⁵ (Fig. 2c). These programs representing diverse biological processes including the cell cycle, stress response, and protein maturation revealed widespread inter- and intra-patient heterogeneity.

To identify other cellular states displaying variability in myeloma cells, we employed nonnegative matrix factorization to identify expression programs in our dataset (Fig. 2d). Using this approach, we detected 6 robust expression programs across patients. Apart from the cell cycle, these reflected cell signaling states, such as KRAS/MAPK signaling, IL2/STAT5 signaling, interferon response, and IL6/STAT3 signaling, a known survival pathway in multiple myeloma (Fig. 2d–f). The sixth program consisted of genes related to the endoplasmic reticulum, a cellular component involved in protein maturation and secretion, processes of paramount importance for myeloma cells. Scoring of these expression signatures revealed widespread transcriptional variability across the identified expression programs between and within patients (Fig. 2g).

Lineage infidelity and plasticity in myeloma cells.

We next sought to define the nature of the transcriptional profiles that are co-expressed in individual cancer cells. We observed expression of genes not normally detectable in plasma cells, such as CD20 (*MS4A1*), the B lymphoid tyrosine kinase *BLK* and the serine protease C (*CTSC*) (Extended Data Fig. 4a). To investigate whether myeloma cells acquire transcriptional states from less mature or entirely different hematopoietic lineages, we defined expression signatures derived from the BLUEPRINT dataset¹². We found that the transcriptional states of myeloma cells indeed diverged from normal plasma cells towards more immature progenitor populations of the lymphoid or entirely different hematopoietic precursors (Fig. 3a,b). Notably, these transcriptional signatures were co-expressed in the same single cells. We validated our observations using signatures from the Human Cell Atlas¹⁶ (Extended Data Fig. 4b,c).

In order to capture the differentiation trajectories between different developmental stages, we performed cluster analysis based on expression of progenitor signatures (Supplementary Table 5) and performed RNA velocity analysis¹⁷. We identified a differentiation trajectory from less to more differentiated cells (Fig. 3c,d). The plasma cells from normal donors clustered towards the endpoint of the trajectory, consistent with their state of terminal differentiation (Extended Data Fig. 4d,e). Cycling cells were enriched towards the starting point of the trajectory (Extended Data Fig. 4f). Since transcriptional diversity is a hallmark of developmental potential which is progressively restricted as cells differentiate¹⁸, we assessed the number of genes expressed per cell and transcriptional entropy^{19,20} (Extended Data Fig. 4g–j) and detected increases in both measures in MM ($p < 2.2e-16$ by two-sided Wilcoxon-test.). Using CytoTRACE¹⁸, we predicted a higher developmental potential for myeloma cells than for normal plasma cells (Fig. 3e, Extended Data Fig. 4k). The cells with the highest CytoTRACE score, i.e. the most immature cells, mapped to the starting point of the trajectory by RNA velocity.

To further define the two divergent differentiation endpoints, we determined the top ranked genes correlated with developmental potential (Supplementary Table 6), which included *LAMP5*, a recently identified marker of MM upregulated through epigenetic

activation¹³ (Fig. 3f). Gene sets associated with immaturity included metabolic processes, Myc activation, and a plasmablast-like signature (Fig. 3g). Interestingly, we observed that the starting and endpoints of the differentiation trajectory were associated with divergent signaling states, characterized by mutually exclusive activation of either the MAPK pathway or PI3K pathway (Fig. 3h,i). While MAPK activity was detected in a subset of cycling and immature cells, the more mature cells showed activation of PI3K, consistent with previous reports showing that PI3K activity induces *PRDM1* and terminal differentiation of B cells²¹. These findings suggest that MM cells arrange along a spectrum of two divergent states, which are characterized by heterogeneity in differentiation state and signaling activity.

To distinguish whether the observed results reflect defective differentiation or a loss of lineage restriction, we assessed expression of plasma cell lineage transcriptional regulators, including *XBPI*, *PRDM1* and *IRF4*. Expression of these regulators was preserved in MM (Extended Data Fig. 4l), arguing that MM cells retain plasma cell identity and manifest a loss of lineage restriction, rather than a block in differentiation.

Regulatory network usage and chromatin accessibility in MM.

As gene-regulatory networks (GRN) have been shown to play a key role in the regulation of cell fates, we set out to define the underlying GRN in MM. We therefore defined transcription factor (regulon) activity using SCENIC²². We identified transcriptional modules present in normal hematopoiesis in the BLUEPRINT dataset and determined the dominating module containing plasma cell transcriptional regulators (Fig. 4a, Extended Data Fig. 5a). In our normal donor plasma cells we detected activity predominantly in the plasma cell module (Fig. 4b). In contrast, gene regulatory network activity of myeloma cells was no longer confined to the plasma cell module, but was observed across other modules assigned to hematopoietic cells including HSC and macrophage modules (Fig. 4b, Extended Data Fig. 5b–d).

To identify state-specific gene-regulatory networks in MM, we next constructed a network based on MM cell regulon activity as well as a network based on normal hematopoietic cells (Extended Data Fig. 5e), and created an overlay (Fig. 4c). Interestingly, while the majority of regulons have activity in normal hematopoietic and myeloma cells, we identified many regulons that were largely inactive or entirely absent in normal hematopoiesis (Fig. 4c,d, Extended Data Fig. 5e, Supplementary Tables 7,8). ELF3 and TEAD4 are two examples of regulons, which gain activity in MM cells (Extended Data Fig. 5f). ELF3 is a member of the epithelium-specific ETS (ESE) transcription factors expressed predominantly in epithelial tissues²³, while TEAD4 acts as a downstream regulator of the Hippo pathway, and binds to the M-CAT motif found primarily in muscle-specific genes²⁴. Among their target genes are genes not normally expressed in plasma cells including *CD3E*, *CD5* and *CD300A* (Supplementary Table 9). These data illustrate an aberrant and promiscuous activation of transcriptional regulators in MM that are not active in normal hematopoietic cells, in line with a loss of lineage restriction.

Importantly, we also identified a shared set of regulons between malignant and normal plasma cells, including *XBPI*, *IRF4* and *PRDM1*, indicating that myeloma cells retain lineage-specific regulons (Extended Data Fig. 5g). This is consistent with our previous

finding that MM cells retain plasma cell identity. Interestingly, even plasma cell lineage defining regulators, such as XBP1, gained network connections in myeloma cells, suggesting that while myeloma cells maintain activity of canonical plasma cell regulators, these acquire additional target genes outside of the plasma cell module (Fig. 4d).

We therefore determined the transcription factors with the highest rewiring score for which connections between transcriptional regulators and target genes are altered²⁵ (Fig. 4e, Supplementary Table 10). We identified ELF3 and TEAD4 as the topmost rewired TFs, followed by XBP1 (Fig. 4e, Supplementary Table 10), indicating that rewired TFs include lineage-defining and non-lineage TFs. Other rewired TFs included E2F1, which has been described as a master regulator of proliferation²⁶. E2F1 has been shown to be essential in myeloma cells and has been identified as a potential therapeutic target²⁷, which suggests that other transcriptional nodes, in particular TFs with a high rewiring score, could represent attractive therapeutic targets.

We next identified the top rewired cell surface markers which may represent potential therapeutic targets. Deregulated surface molecules included the growth factor receptors PDGF receptor A and B (*PDGFRA* and *PDGFRB*), as well as B cell maturation antigen BCMA (*TNFRSF17*) (Fig. 4f). These data illustrate that substantial transcriptional rewiring occurs in myeloma cells although expression of plasma cell lineage-defining TFs is preserved, and misexpression of surface proteins appears to be an important consequence of this rewiring.

To assess whether the identified rewired TFs represent therapeutic vulnerabilities, we investigated CRISPR screening data from the Depmap portal (<https://depmap.org/portal/>). We found that some rewired TFs, such as XBP1 were essential in MM (Fig. 4g,h). Essential TFs were strongly enriched in B cell/plasma cell TFs (Supplementary Table 11), including XBP1, IRF4, and IKZF1 (Fig. 4i,j), reinforcing the notion that MM cells do not lose their plasma cell identity, and the essentiality of plasma cell defining TFs for cell survival. However, other rewired TFs that do not define the plasma or B-cell lineages, such as TAL1 do not generate lineage dependencies in myeloma, likely due to plasticity and redundancy in the transcriptional network (Fig. 4g,k).

To assess whether increased transcriptional diversity could be attributed to greater chromatin accessibility, we performed single cell ATAC sequencing on myeloma cells from a subset of patients as well as normal donor plasma cells (Fig. 5a). A total of 1,483 cells were analyzed and underwent QC filtering (Methods, Extended Data Fig. 6a–d, Supplementary Table 2). As an initial assessment of heterogeneity, we performed clustering and dimensionality reduction (Extended Data Fig. 6e–h). Annotation of the single cell ATAC profiles through integration with our single cell RNA data revealed that clusters 1–6 corresponded to plasma cells, while clusters 7,8, and 9 corresponded to monocytes, B cells and NK cells (Extended Data Fig. 6i–k). Visual inspection of tracks with aggregated ATAC profiles and gene scores for marker genes *SDCI*, *LYZ*, *MS4A1* and *FCGR3A* confirmed these cell type annotations (Extended Data Fig. 6l–s). For further characterization of the single cell accessibility profiles of myeloma cells, we retained clusters annotated as plasma cells only. Similar to our

findings in transcriptional profiling, we observed patient-specific clustering, arguing that individual patients have distinct chromatin profiles (Fig. 5b,c).

We next performed peak calling (Extended Data Fig. 7a–c) and found that the majority of accessible regions in normal plasma cells were shared with MM cells, with only 7,944 of 46,935 peaks unique to the normal plasma cells (Extended Data Fig. 7b). However, we identified a substantial number of peaks unique to MM (29,761; Extended Data Fig. 7b), suggesting that MM cells retain accessible regions of normal plasma cells, but gain additional peaks. Consistent with this finding, we observed a higher number of differentially accessible peaks by DESeq2 gained in MM (3,421 peaks gained vs 887 lost in MM compared to normal donor plasma cells, FDR = 0.1; Extended Data Fig. 7c,d).

To characterize genomic regions with peak enrichment, we annotated chromatin states using ChromHMM annotations in GM12878, a widely used model for normal B cells, as reference. We observed that a greater fraction of accessible regions fell into heterochromatin in MM (Extended Data Fig. 7e,f; $p=0.0029$ by t-test). Furthermore, a greater fraction of peaks was localized in intronic and intergenic regions (Extended Data Fig. 7g), in line with a gain of enhancers. Accessible regions in myeloma cells were also located at a greater distance from the transcription start site (TSS) (Extended Data Fig. 7h; $p < 2.2 \times 10^{-16}$ by Wilcoxon test), consistent with reports describing increased H3K27 acetylation levels in myeloma²⁸.

We next used GREAT to assign genes to enhancers²⁹ and noted that many genes in MM were associated with ten or more enhancers (Fig. 5d). When comparing accessibility of enhancers associated with the top multi-enhancer genes, we found that a subset was differentially accessible in MM (Extended Data Fig. 7i). Gene set enrichment analysis showed that the top multi-enhancer genes were involved in cellular processes such as differentiation, cell death and signaling, indicating that changes in chromatin accessibility modulate these critical processes. As these results depend on accurate enhancer-gene maps, we validated our findings using the Activity-by-Contact (ABC) model to link regulatory regions to target genes³⁰ (Extended Data Fig. 7j–l).

To assess variability in the scATAC dataset, we next performed pseudotime ordering of the single cell ATAC profiles (Fig. 5e). We found that genes known to play a role in MM, such as *CCND1*, as well as *ELF3* displayed higher accessibility towards the start of the trajectory (Fig. 5f). Analyzing motif accessibility across pseudotime (Fig. 5g), we further observed that NFkB family members *REL*, *RELA* and *NFKB1* scored higher towards the end of the trajectory in the normal donor plasma cells. We therefore performed differential motif enrichment analysis using *chromVAR*³¹ and detected differential enrichment of a total of 276 TFs (FDR = 0.01, Extended Data Fig. 7m). We found that 120 of a total of 429 rewired TFs displayed differential motif accessibility (Extended Data Fig. 7n), including TEAD4, ELF3, and other TFs, such as IRF8, FOXO3 and MEF2A (Fig. 5h). This finding indicates that a gain in accessibility of transcription factor binding sites contributes to rewiring and transcriptional reprogramming.

The single cell chromatin accessibility data further allowed us to predict cis-regulatory interactions by defining peak-to-gene linkage. The loci of transcription factors *ELF3* and *TEAD4* manifested further evidence for enhancer activation in MM (Fig. 5i,j) demonstrating that there are widespread alterations in cis-regulatory regions in MM.

Taken together, these data suggest that de-repression of heterochromatin in MM allows for binding of transcription factors and expression of genes not normally expressed in this lineage, ultimately leading to promiscuous acquisition of alternate cell states (Fig. 5k).

Treatment reduces developmental potential.

To assess how treatment modulates the differentiation state of single myeloma cells, we investigated myeloma cells from patients treated with Elo-PVD at Cycle 5 Day 1 (C5D1), i.e. persisting after 4 months of treatment. Alternative splicing resulting in multiple transcript variants is known to regulate key developmental decisions, including maintenance of pluripotency and differentiation³². Differential isoform expression analysis showed differential expression of 312 transcripts with treatment (Extended Data Fig. 8a). Gene set enrichment analysis revealed that transcripts involved in processes like RNA binding, cell activation and regulation of gene expression were differentially spliced between timepoints (Extended Data Fig. 8b).

To further investigate differential splicing probabilities in single cells, we performed Louvain clustering based on the splicing probability ψ and identified 6 clusters, two of which were enriched in treated cells suggesting treatment as cause of cluster formation (Extended Data Fig. 8c–e). Indeed, we observed that transcripts with a high splicing probability showed exon inclusion less frequently upon treatment, i.e. had higher probability of exon skipping (Fig. 6a). *PDCD4* and *TSCI*, a negative regulator of mTORC1 signaling were among the transcripts with reduced splicing probabilities at C5D1, while *CALU*, which is involved in protein folding³³ showed higher splicing probabilities (Extended Data Fig. 8f).

We also generated miso plots to visualize potential exon skipping events or alternate splice site usage in individual transcripts following treatment (Extended Data Fig. 8g–i). We hypothesized that *SLAMF7*, the target of elotuzumab treatment, might be alternatively spliced upon treatment. We identified various splice isoforms resulting in skipping of exon 3 which encodes the antibody binding site³⁴ (Extended Data Fig. 8j). However, we did not observe differential splicing upon treatment in this transcript (Extended Data Fig. 8f), arguing that alternative splicing of *SLAMF7* does not significantly contribute to treatment evasion in this cohort.

We next investigated whether developmental potential changed with treatment. Notably, we observed a decrease in the number of genes expressed (5,700 at screening vs 4,473 at C5D1, $p < 2.2 \times 10^{-16}$ by t-test) (Extended Data Fig. 9a). CytoTRACE values were decreased following treatment ($p < 2.2 \times 10^{-16}$ by t-test; Fig. 6b), suggesting that developmental potential and transcriptional diversity were reduced by treatment, consistent with persisting cells assuming a more quiescent state. We further observed a minor reduction of cycling cells upon treatment (Extended Data Fig. 9b).

We therefore hypothesized that treatment might modulate TF activity. Notably, we observed that overall regulon activity was increased upon treatment, with 160/365 decreased and 205/365 regulons upregulated (Fig. 6c). Regulons that increased in activity included the B cell TF FOXO3 and the regulators of plasma cell fate IRF4 and PRDM1, while downregulated regulons included the rewired non-lineage TFs ELF3 and TEAD4 (Fig. 6d). Among the lineage TFs, we observed an increase in regulon activity in 17/38 (including IRF family members IRF4, IRF2 and IRF8) (Extended Data Fig. 9c,d).

To investigate whether cells with increased regulon activity were selected for by treatment or whether treatment directly induced these transcriptional changes, we modelled short-term drug treatment *in vitro*. Treatment resulted in a reduction in developmental potential, indicating that this change is a direct effect of treatment rather than an effect of selection ($p < 2.2 \times 10^{-16}$ by t-test; Fig. 6e). Short-term treatment further resulted in changes in regulon activity consistent with those observed in MM patients (245/434 up- and 189/434 downregulated), including upregulation of IRF factors and B cell lineage transcription factors, such as FOXO3 (Fig. 6f,g).

We hypothesized that changes in regulon activity could be attributed to differential chromatin accessibility and performed ATAC sequencing of MMCL MOLP2 after 72h treatment with PVD (Fig. 6h,i). Differential accessibility analysis by DESeq2 revealed 5,268 peaks gained and 3,154 peaks lost with treatment (Fig. 6j), indicating that treatment induced significant changes in chromatin accessibility. GO-term analysis of the differential peaks revealed enrichment of several terms related to differentiation (lymphocyte differentiation, cell fate commitment) in peaks gained upon treatment, while peaks lost upon treatment were enriched in lymphocyte proliferation (Extended Data Fig. 9e,f). As we detected fewer genes expressed in the treated cells without a concomitant decrease in accessible regions, we hypothesized that more enhancers were connected to individual genes. Using GREAT to assign genes to enhancers, we noted that following treatment the average number of enhancers per gene was increased (11 in PVD vs 10 in DMSO group, $p = 0.00012$ by t-test), with a greater number of genes connected to ten or more enhancers (Fig. 6k). Gene set enrichment analysis of the top genes gaining enhancers revealed that these were involved in the regulation of cellular processes, such as response to cytokines and cell motility (Extended Data Fig. 9g), suggesting that enhancer rewiring supports the expression of a subset of genes needed for persistence under drug pressure. We validated these results using the ABC model for gene-region annotation (Extended Data Fig. 9h,i).

Motif enrichment analysis in accessible regions showed enrichment of several IRF factors upon treatment, consistent with increases in regulon activities (Fig. 6l). Rewired TFs with differentially accessible motifs also included the IRF factors IRF2, IRF9 and IRF1 (Extended Data Fig. 9j). Collectively, these results suggest that treatment directly modulates the transcriptional landscape.

Immunotherapeutic targets in reprogrammed cancer cells.

Having identified deregulated expression of surface markers in MM, we hypothesized that these might represent attractive therapeutic targets, since surface markers are easily accessible and can be targeted by immunotherapy. Therefore, we assessed the expression of

the surface markers CD33, CD4, CD5, and CD20 (Fig. 7a) in MM cells at screening, as examples of CD markers not normally expressed in this lineage and detected up-regulation in individual patients.

To further assess whether altered differentiation states are reflected in distinct surface marker profiles, we performed deconvolution of cell types based on mRNA expression of surface markers. While normal plasma cells manifest a surface marker profile closest to plasma cells and other B cell subtypes, in MM we identified signatures of different lineages, such as endothelial progenitors and common myeloid progenitors (Fig. 7b). This finding suggests that the lineage infidelity that we identified in MM cells is associated with widespread changes in expression of surface markers.

Notably, we observed a further shift in surface marker expression with treatment, arguing that treatment itself may induce expression of different surface molecules (Fig. 7b). Interestingly, we identified the surface marker CXCR4 among the top up-regulated genes following treatment (Extended Data Fig. 10a). CXCR4 has been implicated in MM for playing a role in disease progression and inducing an EMT-like phenotype³⁵. CXCR4 thus represents an attractive therapeutic target, particularly as CXCR4 antagonists are already tested in clinical trials for relapsed myeloma³⁶ and are in use for other clinical applications³⁷. We validated surface expression of CXCR4 following treatment by flow cytometry (Fig. 7c,d).

To assess its regulation upon treatment, we investigated chromatin accessibility in the scATAC data and detected numerous intergenic enhancers linked to CXCR4 (Fig. 7e). We observed significantly greater accessibility in an enhancer peak associated with CXCR4 following treatment, suggesting differential epigenetic activation of CXCR4. As it has been established that IRF4 regulates CXCR4 expression in pre-B cells³⁸, we compared IRF4 ChIP-Seq peaks with our ATAC-Seq profiles. We observed IRF4 binding at the promoter and enhancer regions of CXCR4, arguing that IRF4 regulates CXCR4 expression in MM. We therefore investigated chromatin accessibility in MMCL MOLP2 after treatment with PVD and found the IRF4 motif to be enriched in ATAC peaks following treatment (Fig. 7f, Extended Data Fig. 10b). Consistent with this observation, we detected a significant increase in IRF4 regulon activity following treatment ($p=1.3e-05$ by t-test; Fig. 7g). These data argue that IRF4 regulates CXCR4 and is activated following treatment, thus resulting in increased CXCR4 expression.

To determine if CXCR4 expression can be induced directly by treatment, we treated MMCLs *in vitro* with pomalidomide (P), bortezomib (V) and dexamethasone (D). We observed up-regulation of CXCR4 in multiple MMCLs (Fig. 7h,i), but only minimal cell death (Extended Data Fig. 10c–e) upon treatment. Quantitative analysis of CXCR4 surface protein expression revealed dexamethasone as the drug having the greatest effect on this phenotype (Fig. 7i, Extended Data Fig. 10f,g). Drug removal resulted in down-regulation back to starting levels (Fig. 7j), suggesting that this phenotype is indeed induced by drug treatment and is reversible. To investigate whether we could exploit this finding for therapeutic targeting, we treated MMCLs with the CXCR4 inhibitors BKT140 and Plerixafor (Fig. 7k,l, Extended Data Fig. 10h,i). We observed greater cell death when

adding BKT140 and Plerixafor following pre-treatment with PVD, arguing that we can induce phenotypic changes that provide the rationale for targeted therapeutic strategies. Collectively, these results demonstrate that drug treatment can indeed modulate surface marker expression and can induce expression of potential therapeutic targets.

Discussion

In conclusion, delineating heterogeneity by defining cellular states through single cell transcriptional and epigenetic profiling allowed us to gain important insights into the gene regulatory network underlying MM biology and to identify potential therapeutic targets.

We detected widespread transcriptional reprogramming in myeloma cells, with expression of genes not normally detected in this cell type. Our analyses demonstrate greater developmental potential and expression of markers associated with immaturity. This is of particular interest as plasma cells, the cell of origin for MM³⁹, are terminally differentiated and this process is generally considered irreversible. Our data suggest that while MM cells do not lose their plasma cell identity, they (re-) acquire expression of gene signatures from earlier developmental stages or entirely different lineages. We show that differential regulon usage and transcriptional rewiring underlie these alternate differentiation states and identify differential transcription factor binding motif accessibility as a possible mechanism.

We detected widespread changes in chromatin accessibility, with de-repression of heterochromatin regions and increased accessibility in enhancer regions. These findings are consistent with previous reports demonstrating that enhancer activation is a key feature of myeloma^{28,40}. We also find that a subset of genes enriched for regulation of differentiation and cell death is associated with many more enhancer regions in MM, reminiscent of previous suggestions that establishment of superenhancers is a key feature of MM biology^{27,28}.

We argue that permissive chromatin states in myeloma lower epigenetic barriers, allowing for promiscuous sampling of alternate cell states⁴¹. This plasticity allows MM cells to aberrantly activate alternate gene regulatory programs, resulting in widespread transcriptional reprogramming, and dedifferentiation in this malignancy derived from terminally differentiated plasma cells. Importantly, we show that this plasticity can also lead to therapeutic opportunities, since it results in expression of surface markers not normally expressed in this lineage. Using CXCR4 as a proof of principle, we further show that surface marker expression can be modulated by treatment and exploited for immunotherapeutic targeting, and we identify increased IRF4 activity as a possible mechanism driving CXCR4 upregulation. Interestingly, we observe de-differentiation independent of genotype, arguing that there are convergent cell states dictated by de-repression of chromatin that supersede genetic background and heterogeneity. Targeting MM cell states, rather than distinct genotypes, may thus represent a way to overcome the extensive genetic heterogeneity of the disease. We propose that identifying gene-regulatory networks and the associated complement of surface proteins in cancer can reveal actionable targets for immunotherapies.

Materials and methods

Patient samples

Banked blood or bone marrow samples from patients with relapsed/ refractory multiple myeloma (Supplementary Table 1), who had signed consent and were enrolled on a phase II clinical trial investigating with elotuzumab, pomalidomide, bortezomib and dexamethasone on a clinical trial (Elo-PVD; [NCT02718833](#)) were used in this study. The research was determined not human subject research by DF/HCC IRB (19–511). Samples from 8 patients (5M, 3F) were analyzed, median age at registration was 68.5 years (range 45–80). Myeloma cells were enriched using the EasySep Human CD138 positive selection kit (STEMCELL Technologies). Bone marrow from normal donors was obtained from AllCells.

Single cell sorting

Bone marrow cells were stained with antibodies against either CD38 FITC (multi-epitope, Cytognos, 1:200), CD138 PE (44F9, Miltenyi, 1:100), SLAMF7 APC (162.1, Biolegend, 1:100), BCMA PE-Cy7 (19F2, Biolegend, 1:100), or with antibodies against CD45 FITC (HI30, eBioscience, 1:100), CD3 PerCP-Cy5.5 (OKT3, eBioscience, 1:100), CD14 APC-Cy7 (MoP9, BD Biosciences, 1:100), CD19 PE (HIB19, BioLegend, 1:100). Cells were stained with DAPI (1 µg/mL, Sigma-Aldrich) to exclude dead cells. Single cells were sorted into 96-well plates using a Sony SH800 sorter (Extended Data Fig. 1). Immediately after sorting, plates were spun down, and stored at –80 °C until further processing.

Single cell RNA-Seq library preparation

Full-length single cell RNA-Seq libraries were prepared using the SmartSeq2 protocol¹⁰. Briefly, single cells were sorted into 96-well plates with each well containing 10 µl of TCL buffer (Qiagen) and 1% 2-mercaptoethanol. RNA purification was performed using RNAClean XP beads (Beckman Coulter) prior to cDNA synthesis. RNA was reverse-transcribed using Maxima RNase H-minus (ThermoFisher) in the presence of biotinylated oligo-dT30VN (/5BiosG/AAGCAGTGGTATCAACGCAGAGTACT(30)VN), biotinylated template-switching oligonucleotides (/5BiosG/AAGCAGTGGTATCAACGCAGAGTACATrGrG+G), and betaine. Amplification of cDNA was performed using KAPA HIFI Hotstart ReadyMix (Kappa Biosystems) and biotinylated ISPCR primers (/5BiosG/AAGCAGTGGTATCAACGCAGAGT) with the following amplification protocol: 98°C for 3 min, 24 cycles of [98°C for 15 sec, 67°C for 20 sec, 72°C for 6 min], 72°C for 5 min. A total of 0.15 ng of the amplified cDNA was subjected to tagmentation using the Nextera XT kit (Illumina) and amplified with Nextera indexes. Pooled libraries were paired-end sequenced using a 75 cycle kit on a NextSeq 500 (Illumina) with an average sequencing depth of 1 million reads per cell.

Single cell ATAC-Seq library preparation

2,000–5,000 plasma cells were bulk sorted as described above and single cell ATAC-Seq libraries were prepared as described by Chen et al.⁶⁶. Libraries were paired-end sequenced

using a 75 cycle kit on a NextSeq 500 (Illumina) sequencer with an average sequencing depth of 0.5–1 million reads per cell.

Multiple myeloma cell lines

Human multiple myeloma cell lines (MMCL) were obtained from ATCC (MM1.S, CRL-2974) or DSMZ (MOLP-2, ACC 607; OPM-2 ACC 50). The cell lines MOLP2, OPM2, MM1.S were cultured in RPMI-1640 (Gibco) media with 10% of heat-inactivated fetal bovine serum (MM1.S and OPM2) or 20% (MOLP2), 100 μ M penicillin/streptomycin (Invitrogen), 2 mM L-glutamine (Sigma), 1 mM sodium pyruvate (Sigma), 1x NEAA (Sigma), 20 mM HEPES (Sigma) and 0.5 mM 2-mercaptoethanol. Cell line identity was confirmed by short tandem repeat profiling and cultures were tested for mycoplasma every 3 months.

In vitro inhibitor treatment

MM cell lines were treated for 72h with pomalidomide (Selleck), bortezomib (Selleck), and dexamethasone (MilliporeSigma) individually or in combination (half IC₁₀ concentration for bortezomib and half IC₁₀ for MM1S for dexamethasone and pomalidomide, half IC₅₀ for OPM2 and MOLP2) at the indicated concentrations. To investigate the effects of drug removal on CXCR4 expression, cells were treated for 72h, washed, and incubated in RPMI-1640 media without drug for a further 72h. Cells were stained with PE/Cy7 anti-human CXCR4 (12G5, BioLegend, 1:100) and FITC Annexin V (BioLegend) and were analyzed using a BD Accuri C6 Flow Cytometer. The mean fluorescence index of each sample was calculated using FlowJo software v10.

For bulk ATAC sequencing, MMCLs were treated with inhibitors for 72h and cells were ficolled. 50,000 cells were pelleted and ATAC sequencing was performed using the Omni-ATAC protocol as described by Corces et al.⁴².

To assess sensitivity to CXCR4 inhibitors following pre-treatment with PVD, CXCR4 inhibitors plerixafor (Selleck) and BKT140 (Bachem) were added at IC₅₀ concentrations after 72h. Cell viability was measured after a further 72h using the CellTiterGlo Luminescent Viability Assay (Promega).

STATISTICAL ANALYSIS

Processing of single cell RNA-Seq data

Sequencing reads were trimmed using trimmomatic and aligned to the human genome (version hg19) using STAR aligner with following parameters '-- twopassMode Basic --alignIntronMax 100000 --alignMatesGapMax 100000 --alignSJDBoverhangMin 10 --alignSJstitchMismatchNmax 5 -1 5 5'^{43,44}. HTSeq and RSEM were used to obtain raw counts and normalized TPM values from the aligned bam files^{45,46}.

Quality control and filtering of single cell RNA data

To remove low quality cells from our dataset, we used the following parameters - distribution of library size (i), number of detected genes (ii), percentage of counts mapping

to mitochondrial genes (iii), and percentage of counts mapping to house-keeping genes (HKGs) (iv) per cell (Extended Data Fig. 2a). An outlier cutoff of 3 median absolute deviations (MADs) was chosen and cells with MADs > 3 were flagged as poor quality. We further used the ‘mvoutlier’ package to identify poor quality cells without predefined cut-offs. Cells identified as outliers by both methods were removed from the dataset (Extended Data Fig. 2b). We investigated the contribution of various technical factors, including individual, genes detected, sequencing run and library size, to the total variation observed in the dataset, and noted that the contribution of these variables was low (Extended Data Fig. 2e).

Clustering of single cell RNA-Seq profiles

Clustering of high-quality cells was performed using the multilevel graph-based clustering algorithm within PAGODA2¹¹. Reference-based cell type annotation of the single cell profiles was performed using the R package ‘SingleR’⁴⁷ and a published dataset of 21 immune cell populations from the BLUEPRINT consortium¹² as reference. Marker genes for each of the MM clusters (Extended Data Fig. 3h) were determined using the *findMarker* function in the *scran* package⁴⁸.

Classification of malignant cells

Copy number variations were inferred from the single cell RNA-Seq (scRNA-Seq) expression profiles using ‘*InferCNV*’^{49,50} and ‘*CONICSmat*’⁵¹. Plasma cells from normal donors served as controls.

The CDR3 region was reconstructed for each individual cell using ‘*MiXCR*’ package⁵². The raw reads were aligned to the B cell VDJ receptor datasets. This was followed by steps of partial assembly, with the poor-quality unmapped reads used to extend the alignment. Identical sequences were then grouped into clones after correcting for PCR duplicates and sequencing errors. The amino acid sequence from the CDR3 region of the light chain was used to track the clonality of the plasma cells.

A random-forest classifier was built in R defining patient-specific clusters showing enrichment for copy-number aberrations as malignant. We randomly split the dataset into non-overlapping training and test sets, using 30% of cells for testing. The optimum parameters – *mtry* (62) and *ntrees* (500) for randomforest were determined using *tuneRF* function. The relative importance of individual genes in classification was evaluated using *varImpPlot* function.

A composite malignancy score was calculated by scoring relative expression of positive RF marker genes and adding a value of +1 if i) CNVs called by clinical cytogenetics were detected and ii) the dominant CDR3 sequence was detected.

Molecular classification of myeloma cells on a single cell level

For molecular classification of myeloma cells, we used established gene expression classifiers from a bulk transcriptome study¹⁴. Scores for each subgroup were calculated

by taking the average relative expression of the positive marker genes and subtracting the average relative expression of the negative markers.

Defining heterogeneity programs

Myeloma cells at screening were analyzed using NMF (version 0.23.0)⁵³. The gene expression count matrix was transferred to log2 counts-per-million (logCPM) by edgeR_3.32.0 package^{54,55}. As previously described⁵⁶, all negative values were replaced with zero. The top 50 genes based upon the highest NMF scores were selected as signature genes for each program (Fig. 2e). The connectivity and stability of the obtained programs, assessed by calculation of consensus matrix, was visualized in a heatmap (Fig. 2d). To classify the identified programs, we used the functional enrichment analysis of signature genes with Molecular Signatures Database (MsigDB) hallmarks and ontology⁵⁷⁻⁵⁹(H, C5:BP) gene sets considering FDR-adjusted $P < 0.05$ as significant. We scored the identified signature genes in myeloma cells at screening using *scanpy*⁶⁰ (Fig. 2g).

Comparison to published RNA-Seq datasets

Gene expression data from bulk sorted populations of immune cells from the BLUEPRINT dataset¹² were used to assess dedifferentiation and expression of genes from other cell types. The signatures were scored using *scanpy*⁶⁰. Expression of markers from different lineages was validated using single cell RNA-Seq derived signatures from the Human Cell Atlas¹⁶.

Defining markers of immaturity and differentiation

Markers for different progenitor signatures were derived from signatures defined in the xCell study (Supplementary Table 5)⁶¹. Clustering was performed on the relative expression of genes in progenitor signatures (CLP, CMP, MPP, MEP) along with B cell and plasma cell genes in myeloma and normal donor plasma cells.

RNA velocity projections were plotted onto 2-dimensional clustering to reveal immature root and endpoint states²¹. We assessed expression of marker genes for G1, S and G2M phase to predict the cell-cycle phase using *Seurat*⁶². The R package *CytoTRACE* (v0.1.0) was used to determine developmental potential¹⁸. The top100 genes correlated or anti-correlated with CytoTRACE values (Supplementary Table 6) were chosen as markers for immaturity and differentiation, respectively. Gene set enrichment analysis was performed using MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). Signaling activity was predicted based on perturbation experiments using *PROGENy*⁶³.

Entropy was calculated using the R package *LandSCENT*²⁰ with a previously defined protein-protein interaction (PPI) network provided with the *LandSCENT* package available on github under the filename “net13Jun12.m.RData”. For validation, we used an updated version of the PPI network also provided on github under the filename “net17Jan2016.m.RData”, which contains a greater number of protein-protein interactions (11751 vs 8434 columns, Extended Data Fig. 4j).

Construction of regulon network

The bulk RNA-Seq data from the BLUEPRINT consortium¹² was used to construct a reference network. Only the high confidence regulon targets with Genie3Weight ≥ 0.01 were filtered from the file - `2.5_regulonTargetsInfo.Rds` (Supplementary Table 9) generated by SCENIC²². The modules in the constructed network were detected using the ModuLLand⁶⁴ plugin in cytoscape⁶⁵. Similarly, the network was constructed for the disease state by using the scRNA-Seq data generated from our study from malignant and normal plasma cells. We compared the regulon network layout of the plasma cells in the BLUEPRINT with that of the normal-donor plasma and myeloma cells at screening.

While for the BLUEPRINT plasma-cell type we averaged the AUC values over replicates for each regulon, for the scRNA-Seq data generated from our study we averaged the AUC values for each regulon for MM or ND plasma cells. Finally, for each cell type/group we centered the averaged AUC values by the mean over regulons and scaled by the standard deviation (Supplementary Table 7,8). Whenever a regulon transcription factor produced another (extended) regulon with a different number of target genes, we retained only the one having high-confidence target genes, identified by having the same or close number of target genes as shown in the high-confidence regulon targets data. To project the scaled regulon activity onto the BLUEPRINT and ELO networks, we integrated the scaled regulon activity data to their corresponding regulon target data.

Both the normal hematopoietic and disease networks were compared in order to analyze the rewiring of regulon networks using the DyNet algorithm²⁵ (Supplementary Table 10). Lineage transcription factors were identified as transcription factors present in the B cell or plasma cell modules with selected transcription factors added from the literature (Supplementary Table 11).

Expression of surface markers

Gene expression data from bulk sorted populations of immune cells from the BLUEPRINT dataset¹² were used to score enrichment of surface markers from other cell types in MM cells. Only CD markers were selected, and their enrichment was scored using the R package 'TissueEnrich'.

Dependency scores

Dependency data are based on CERES scores from the Depmap dataset (DepMap Public 19Q1) and lineage-specific dependencies were determined using the R package HDCRC2019⁶⁶. A gaussian mixture model was calculated with 2 components and a cutoff of $FDR < 0.05$ and $estimate > 1$ was chosen for identifying lineage-specific essential genes.

Detection of alternative splicing

We quantified exon inclusion/exclusion ratio (percent of spliced in, psi) using BRIE2 (Bayesian regression for isoform estimate, v2, Python package developed under the Python2 environment)⁶⁷. 7,516 predefined annotations for exon skipping splicing of human data were downloaded from <https://sourceforge.net/projects/brie-rna/files/annotation>. The produced psi

values were used for projection of selected events and downstream analysis by creating a psi-based UMAP.

Poor-quality genes were filtered out using the defined arguments in *brie-quant* mode (*minCount*, *minUniqCount*, *minCell* and *minMIF*) choosing a minimum of 50 counts, a minimum of 10 unique counts across all cells, a minimum of 30 cells with unique counts and a minimum of 0.001 minor isoform frequency in unique counts, respectively.

Single cell samples in each experimental condition were selected randomly to generate Sashimi plots using MISO (Mixture of Isoforms; version 0.5.4)^{68,69}. Skipped exon events (GFF annotation files) were downloaded from <https://miso.readthedocs.io/en/fastmiso/index.html#event-annotation>.

Processing and analysis of scATAC-Seq data

Single cell chromatin accessibility profiles were processed with ArchR⁷⁰. Quality control filtering was performed and only high-quality cells with a TSS enrichment score greater than 4 and greater than 1000 unique nuclear fragments were retained. Doublet filtering resulted in a further removal of 11 cells (1% of cells). QC measures and sample statistics are shown in Extended Data Fig. 6a–d, Supplementary Table 2.

We next performed dimensionality reduction using Iterative Latent Semantic Indexing (LSI) and clustering using a graph-based approach implemented in Seurat with the *FindClusters* function. We performed batch effect correction using *Harmony* prior to clustering again (Extended Data Fig. 6g,h). To define cluster identities, we performed integration with our scRNA-Seq dataset using the *FindTransferAnchors* function implemented in Seurat.

For pseudotime ordering of single cell ATAC profiles, we defined the patient with the highest CytoTRACE score (MM6) as the starting point and the normal donor plasma cells as the end point of the trajectory (Fig. 5e). To predict cis-regulatory interactions we defined peak-to-gene linkage in ArchR. A correlation cutoff of 0.45 and a resolution of 1 were selected.

Differential motif accessibility was investigated using the R package *chromVAR*³¹. We identified a set of background peaks that were matched in GC-content and accessibility and calculated bias-corrected deviations and z-scores. Motifs were imported from the JASPAR 2018 database and differential motif accessibility was determined specifying MM and ND plasma cells as groups for comparison based on z-scores.

For peak annotation in aggregated scATAC data, samples from individual patients where greater than 50 cells were retained were collapsed, downsampled to obtain comparable numbers of reads and analyzed using the *nfcore/atacseq* pipeline v1.0.0 with the *--narrowPeak* parameter⁷¹ (Extended Data Fig. 7a–c). To annotate chromatin states, we obtained ChromHMM annotations for GM12878 cells from ENCODE. GM12878 ChromHMM states were intersected with the ATAC-Seq peaks using *bedtools intersect*.

For gene-enhancer relationships, we used annotations from GREAT with default settings⁹ where GREAT assigns each gene a regulatory domain consisting of a basal domain that

extends 5 kb upstream and 1 kb downstream from its transcription start site, and an extension upstream and downstream up to the basal regulatory domain of the nearest gene within 1 Mb. GREAT further refines the assignment using curated experimentally determined regulatory domains.

We further implemented the Activity-by-Contact model to predict enhancer-gene interactions³⁰. To this end, publicly available H3K27 acetylation ChIP-Seq data were downloaded from the European Nucleotide Archive (Accession Number PRJEB25605²⁸ for patient samples, PRJNA608682⁷² for normal plasma cells) and peaks were called using the nfcore/chipseq pipeline v1.0.0 (<https://github.com/nf-core/chipseq>) with default settings. CUT&TAG data for H3K27ac were generated for MOLP2 as previously described⁷³. Expression data was provided as average logcounts from single-cell RNA-Seq. The ABC score was calculated using average HiC data provided by the package and default options. Elements with an ABC score of below 0.02 were discarded.

The R package ‘Cicero’ was used to define co-accessibility of genomic regions based on the scATAC data⁷⁴. A co-accessibility cutoff of 0.25 was chosen and only connections with the respective TSS were selected using the ‘viewpoint’ feature. The R package ‘GVIZ’ was used for visualization. For comparison, double elite GeneHancer tracks were imported from UCSC and enhancers linked to CXCR4 by GREAT are shown. For visualization of the CXCR4 locus, publicly available IRF4 Chip-Seq data were downloaded from the European Nucleotide Archive (Accession Number PRJEB25605) for comparison and bigwig files were generated using the nfcore/chipseq pipeline v1.0.0⁷¹.

Bed files for specific subsets of peaks were given as input to deeptools (v3.0.2) with normalized bigwig files. A matrix was constructed using the ‘computeMatrix’ function with the following parameters: ‘--referencePoint center -a 2000 -b 2000 --missingDataAsZero --skipZeros’. Regions +/- 2 kb around the peak center were selected for visualization and heatmaps were sorted by the ND sample.

Motif enrichment was assessed using the Differential ATAC-Seq toolkit (DAStk) using the HOCOMOCO v11 motif library. Differential MD scores were determined with a p-value cutoff of 0.001. Motif enrichment over background in samples treated with PVD or DMSO was investigated with HOMER using the *findMotifsGenome.pl* command.

Statistics & Reproducibility

Results were plotted and quantified using R-based packages or GraphPad Prism 8. Significance was calculated using the indicated statistical tests. No statistical method was used to predetermine sample size. No data were excluded from the analyses. For scRNA and scATAC experiments, low-quality cells were filtered and excluded from downstream analyses, as described. No attempts of replication failed. The experiments were not randomized. The investigators were not formally blinded to allocation during experiments and outcome assessment.

Code availability

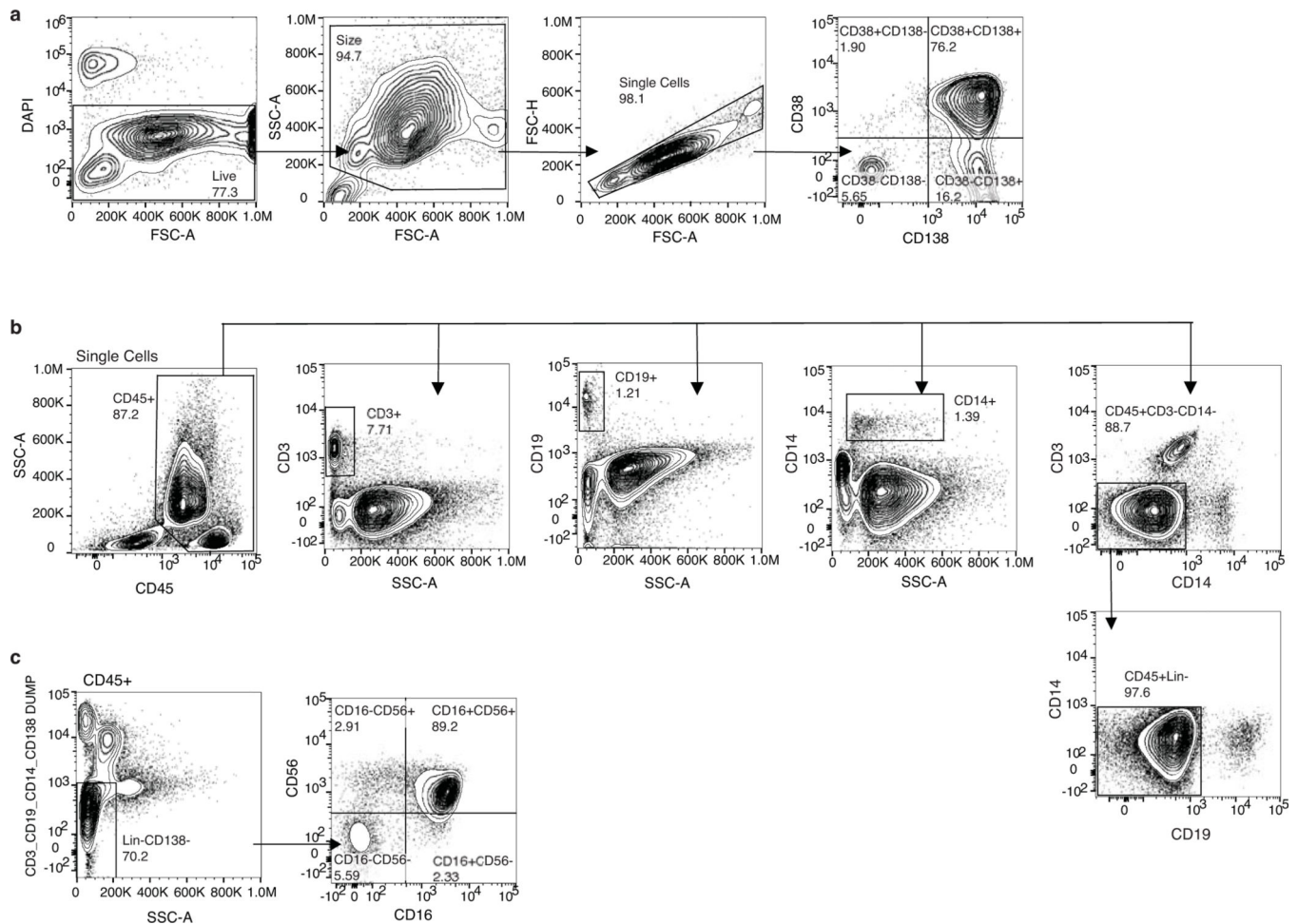
The following tools are available on github: trimmomatic, STAR aligner, HTSeq, RSEM and the nfcore/atacseq pipeline v1.0.0, and for downstream analysis R (v3.6.1 or 4.0), python (v2.7.2 or 3.6.0) and the following packages: mvoutlier (v2.1.1), PAGODA2 (v0.1.1), SingleR (v1.0.1), scran (1.12.1), InferCNV (1.4.0), CONICSmatrix (v0.0.0.1), MiXCR, randomFOREST (v4.6–14), SCANPY (v1.4.6), CYTOTRACE (v0.1.0), PROGENY (1.6.0), LandSCENT (0.99.3), SCENIC (1.1.2–2), Seurat (v3.2.1), velocity, scvelo (v0.2.0), TissueEnrich (v1.4.1), Cicero (v1.2.0), chromVAR (v1.6.0), GVIZ (v1.28.3), deeptools (v3.0.2), ArchR (v0.9.5), edgeR (v3.32.0), NMF (v0.23.0), BRIE2 (v2) and MISO (v0.5.4); also FlowJo (v10), cytoscape (v3.8.0), GREAT (<http://great.stanford.edu/public/html/>), GSEA (<https://www.gsea-msigdb.org/gsea/index.jsp>).

Data availability

The data collected in this paper include next-generation sequencing data: The NGS data consist of 1) single cell RNA sequencing datasets from primary myeloma and CD45+ immune cells and normal donor bone marrow plasma cells and CD45+ immune cells, and 2) single cell RNA sequencing datasets from multiple myeloma cell line MOLP2 treated with PVD or DMSO; 3) single-cell ATAC sequencing datasets were generated from primary myeloma and normal donor plasma cells; 4) bulk ATAC-sequencing datasets from myeloma cell line MOLP2 treated with PVD or DMSO and 5) CUT&Tag data for H3K27ac from myeloma cell line MOLP2 treated with PVD or DMSO. FASTQ files were generated following Illumina sequencing and further analyzed as described below and in the Methods section. The scRNA and ATAC-sequencing dataset and CUT&Tag data generated in this study have been deposited in GEO under the accession number GSE162337. The cytoscape files for the final gene regulatory networks are available for download at the Boad Institute Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1511/dynamic-transcriptional-reprogramming-leads-to-novel-immunotherapeutic-vulnerabilities-in-myeloma).

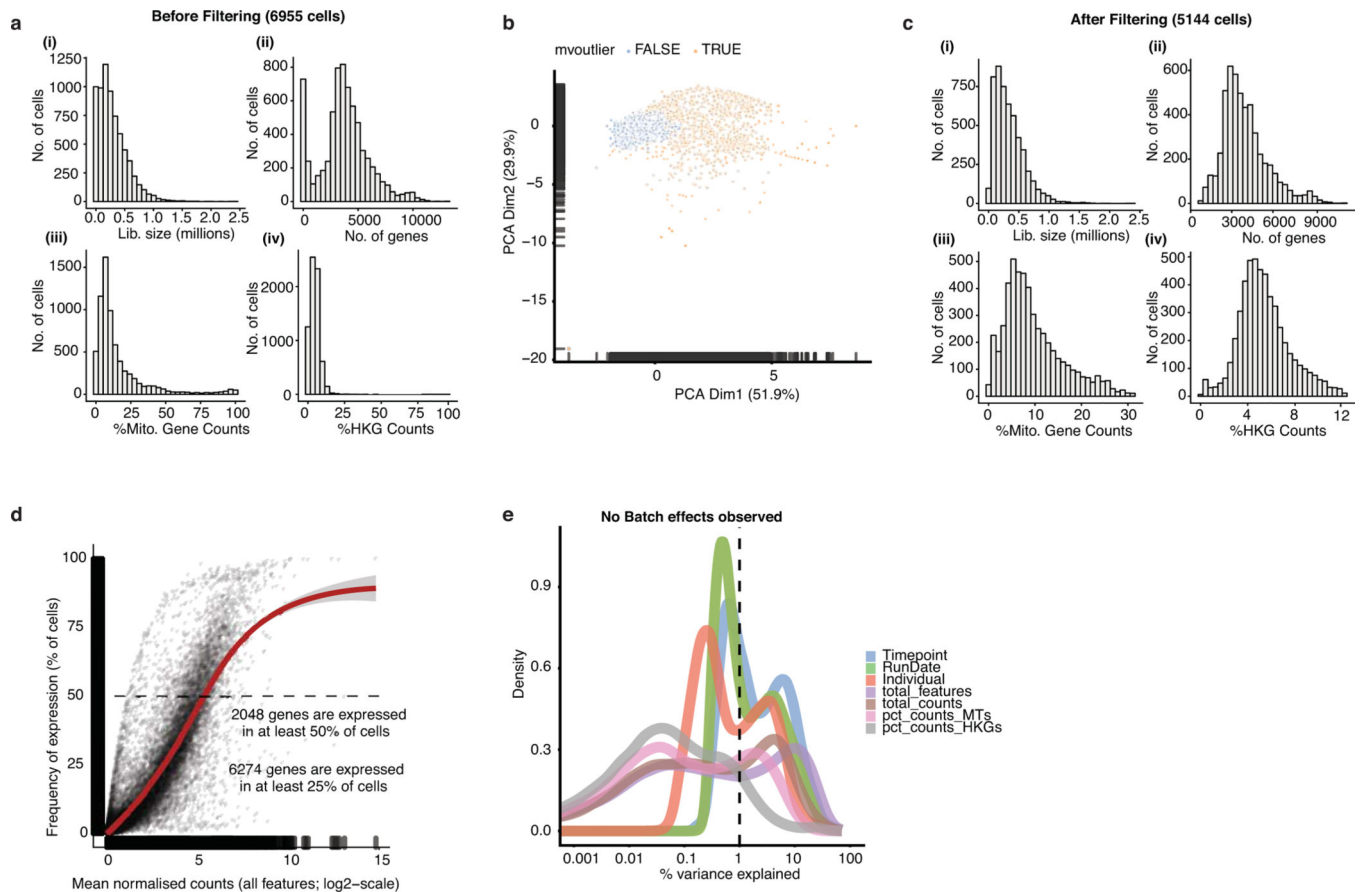
Previously published IRF4 Chip-Seq data that were re-analyzed here are available under accession code PRJEB25605 from the European Nucleotide Archive. Previously published H3K27ac ChIP-Seq data that were re-analyzed here are available from the European Nucleotide Archive under accession codes PRJEB2560542 (for patient samples), and PRJNA608682103 (for normal plasma cells). Additionally, we used the following publicly available datasets: the dependency scores available through the depmap download portal (<https://depmap.org/portal/download/>); the BLUEPRINT dataset¹², JASPAR motif database. Source data are provided with this study. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

Extended Data



Extended Data Fig. 1. Sorting strategy.

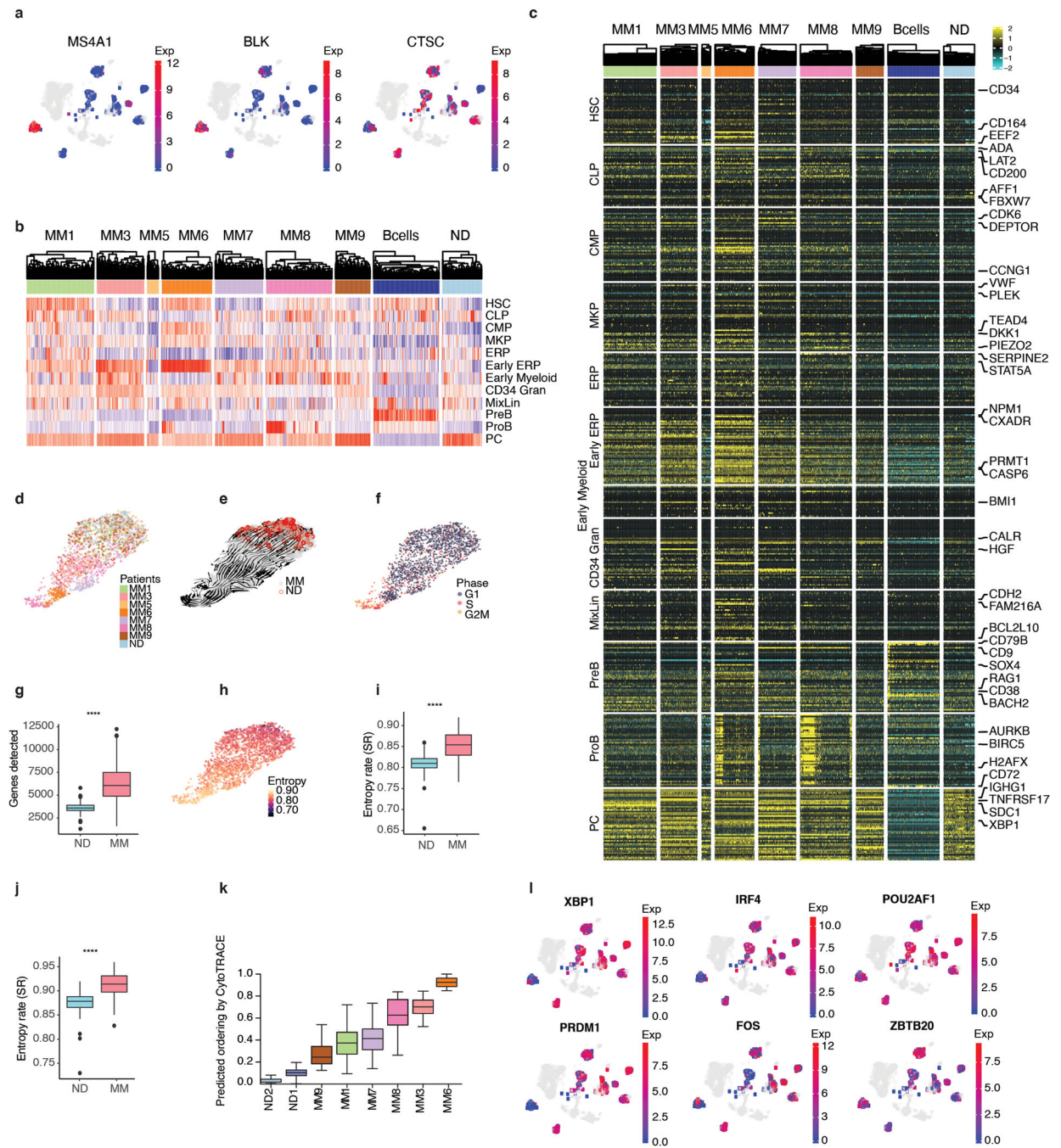
a) Sorting strategy for myeloma cells and normal donor plasma cells with representative flow cytometry plots. CD38+CD138+ cells were sorted after EasySep enrichment of CD138+ cells from bone marrow or peripheral blood. b) Sorting strategy for immune cell subsets. CD3+, CD19+, CD14+ and CD45+Lin- cells were sorted following exclusion of dead cells and doublets. c) Sorting strategy for NK cells. CD56+CD16+ cells (Q1–3) were sorted after exclusion of dead cells, doublets, CD3+ cells and CD138+ cells.



Extended Data Fig. 2. Quality assessment and filtering of single cell RNA data.

a) Distribution of library size (i), number of detected genes (ii), percentage of counts mapping to mitochondrial genes (iii), and percentage of counts mapping to house-keeping genes (HKGs) (iv) per cell. b) Scatter plot depicting principal component analysis using the top two dimensions. The PCA was performed on the four features depicted in a) for all cells in the unfiltered dataset. The outliers, highlighted in orange, were identified using the mvoutlier package. c) The distribution of features shown in a) after filtering out cells identified as outliers. d) Scatter plot depicting expression frequency vs mean read counts per gene in the filtered dataset. e) Density plot showing the contribution of various technical factors (timepoint, run date, individual, total features, total counts, percentage of counts mapping to mitochondrial genes, and percentage of counts mapping to house-keeping genes (HKGs)) to the total variation observed in the dataset.

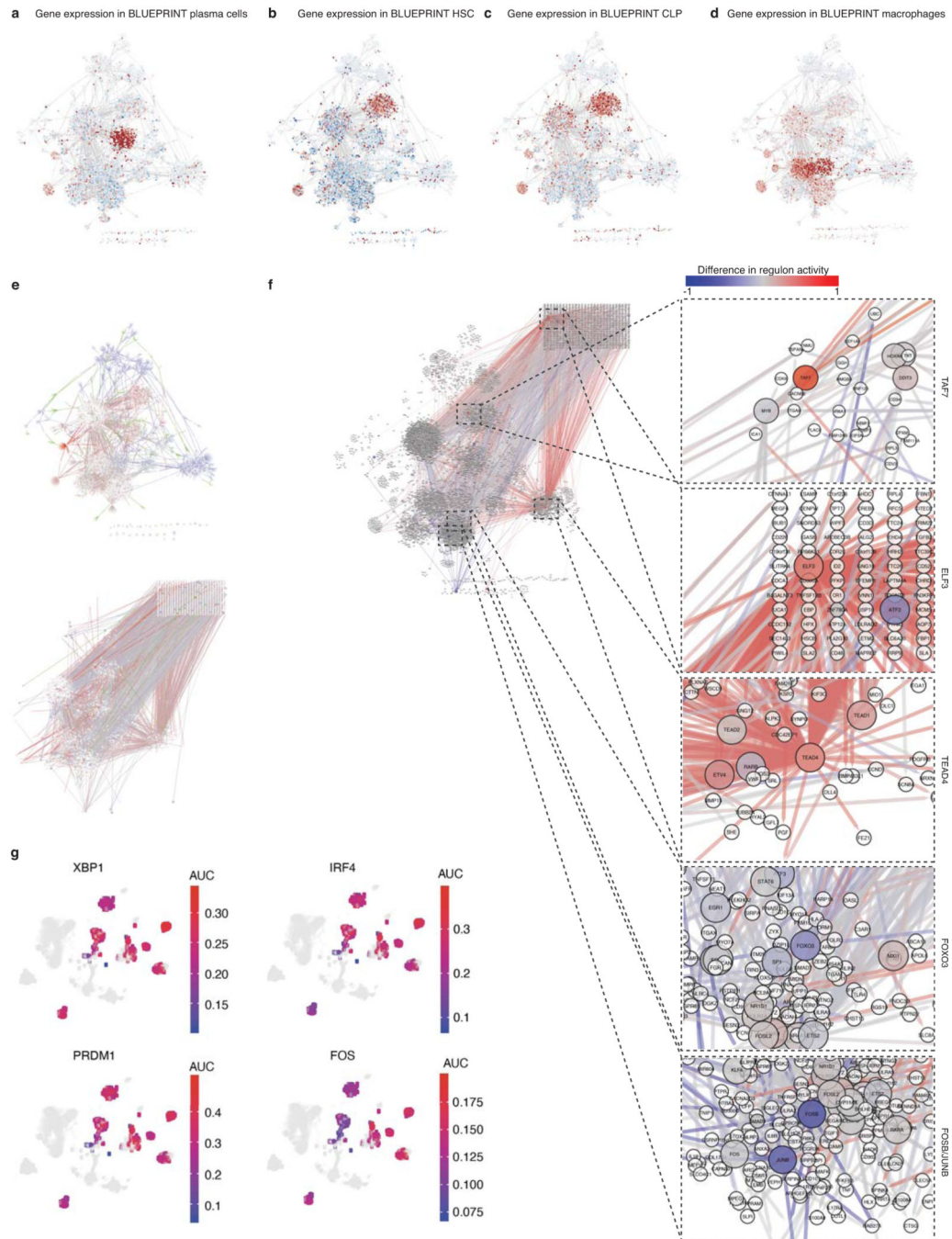
best discriminate between normal and malignant cells, and composite malignancy score for patient MM5. g) tSNE plots showing PAGODA2 clustering highlighting cells from patient MM5 with malignancy scores ≥ 1 or < 1 . h) Heatmap with relative expression of marker genes for individual patients in single myeloma and normal plasma cells.



Extended Data Fig. 4. Transcriptional diversity is increased in MM.

a) tSNE plots showing expression of selected genes not normally expressed in plasma cells (log₂-transformed counts). Depicted are CD20 (*MS4A1*), the B Lymphoid Tyrosine Kinase

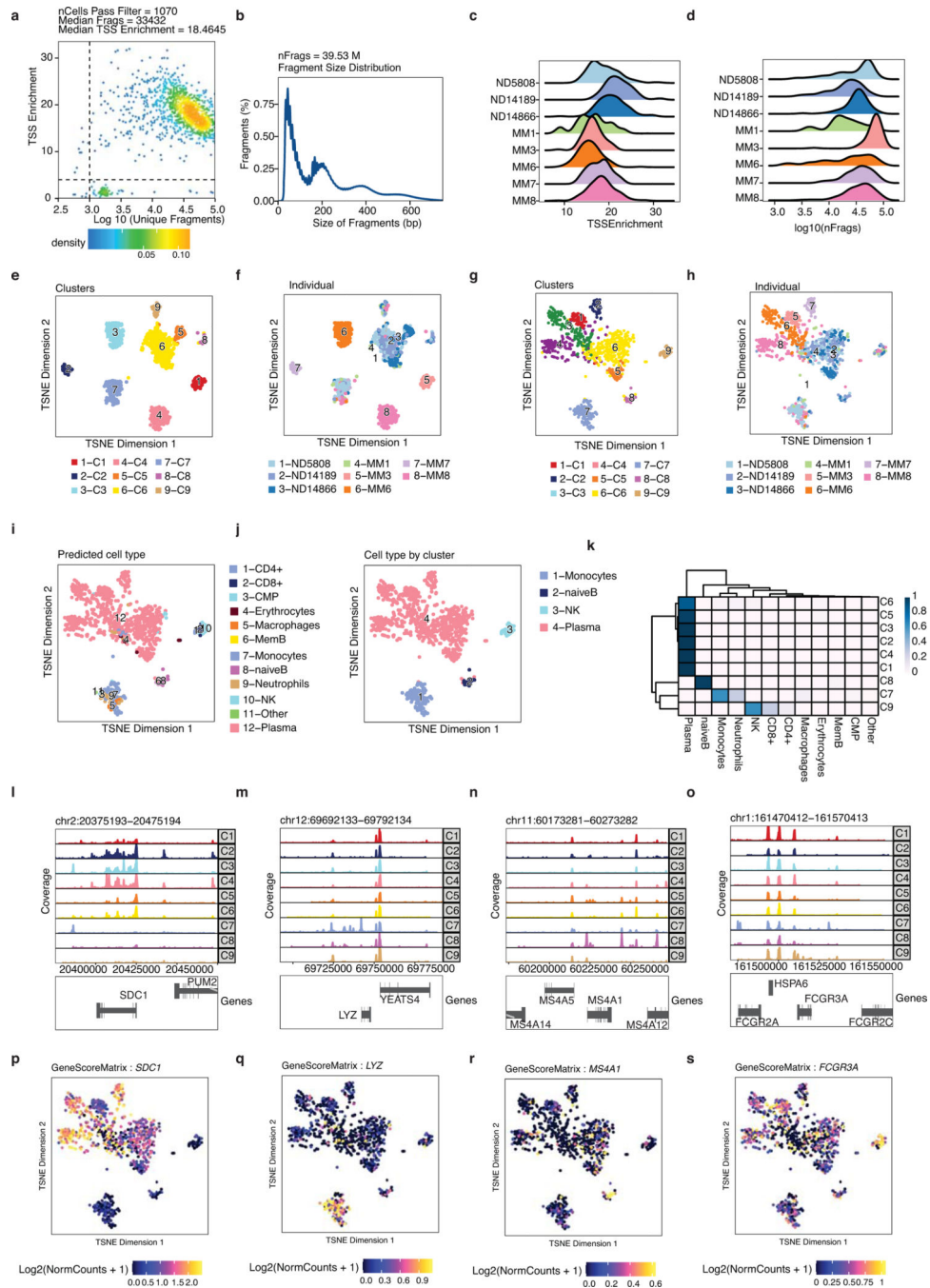
BLK, usually expressed in earlier stages of B cell development and the myeloid restricted serine protease C (*CTSC*). b) Heatmap with lineage scores from single cell RNA-Seq derived datasets from the Human Cell Atlas in single myeloma cells from patients, single plasma cells from normal donors or single B cells. HSC, hematopoietic stem cell; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; MKP, MK progenitor; ERP, ER progenitor; CD34 Gran, CD34+ Granulocyte progenitor, MixLin, mixed lineage progenitor; PreB, Pre-B cell; ProB, Pro-B cell; PC, plasma cell. c) Heatmap showing relative expression of individual genes from different lineages from the Human Cell Atlas dataset in single myeloma cells, single normal donor plasma cells or single B cells. d) UMAP colored by patient. The enlarged heatmaps in b-c are provided in Supplementary Information. e) RNA velocity estimates of single myeloma cells and normal donor plasma cells projected onto two-dimensional UMAP. Normal donor plasma cells are indicated in red. f) Cells colored based on cell cycle phase. g) Number of genes detected per cell in single myeloma vs normal donor plasma cells. Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. n= 1,162 cells. $p < 2.2e-16$ by two-sided Wilcoxon-test. h,i) Entropy is increased in myeloma vs normal donor plasma cells. Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. n= 1,162 cells. $p < 2.2e-16$ by two-sided Wilcoxon-test. j) Validation of entropy using an alternative protein-protein interaction network. Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. n= 1,162 cells. $p < 2.2e-16$ by two-sided Wilcoxon-test. k) Predicted ordering by CytoTRACE, which orders MM cells based on their developmental potential from most mature (lowest values) to most immature (highest values). Boxplots show the median and interquartile range, whiskers extend from min to max. l) tSNE plots showing expression of plasma cell lineage transcription factors *XBPI*, *IRF4*, *PRDM1*, *FOS*, *POU2AF1* and *ZBTB20* (log2-transformed counts) in single myeloma cells, normal donor plasma cells and B cells.



Extended Data Fig. 5. Gene regulatory network activity in different cell types.

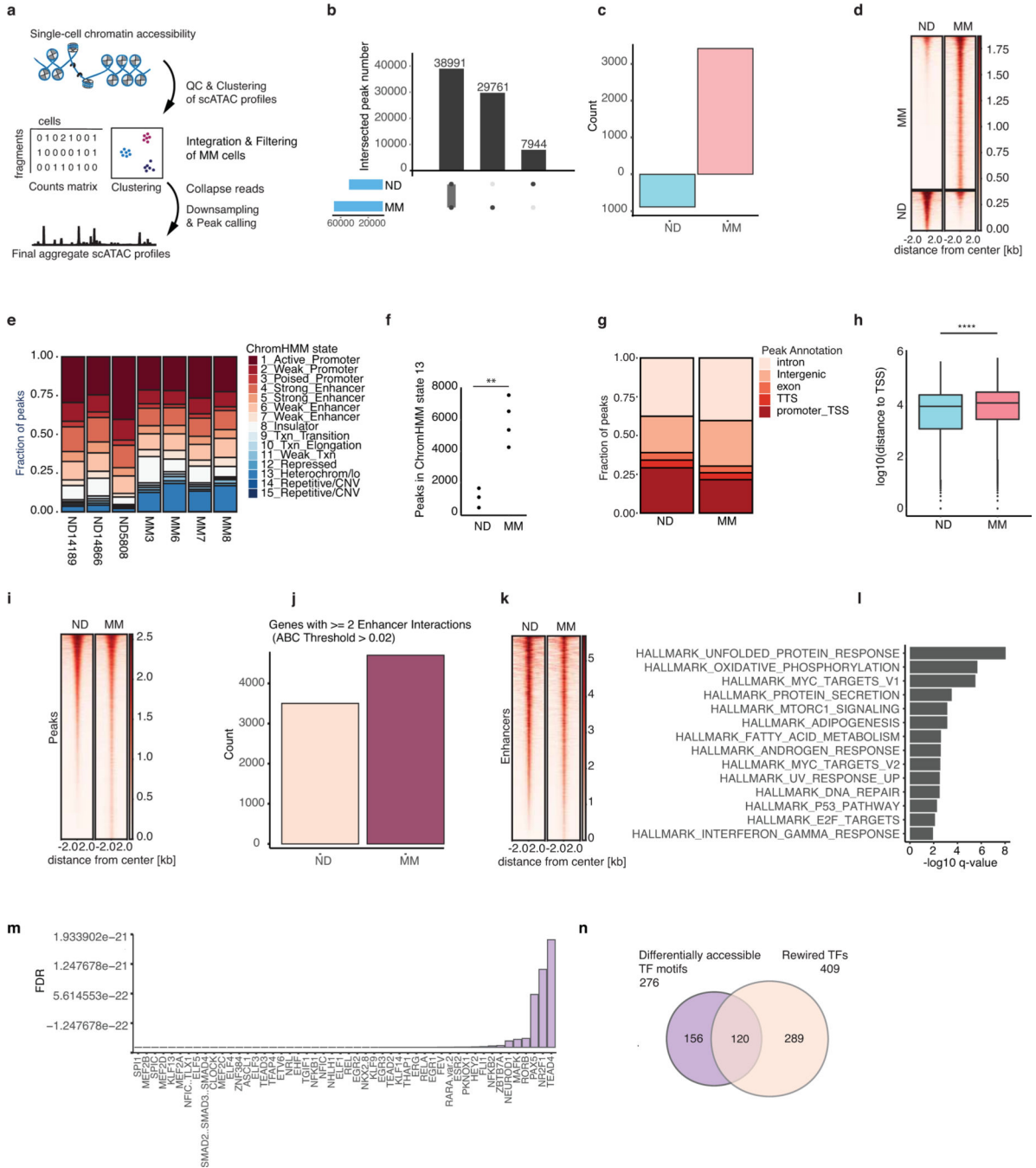
a-d) Gene regulatory network activity for different cell types was determined from the BLUEPRINT dataset. a) Gene expression in normal plasma cells. b) Gene expression in hematopoietic stem cells (HSCs). c) Gene expression in CLP (common lymphoid progenitor) population. d) Gene expression in macrophages. e) Network layouts for normal hematopoiesis based on the BLUEPRINT dataset (top) and based on our single-cell RNA data (bottom) illustrating extensive rewiring, gain of new connections and changes in relative activity. Edges are colored based on regulon activity where high activity is indicated

in red, low activity in blue. Target genes are depicted in white, transcription factors which are not among regulons are shown in green. f) Changes in regulon activity in myeloma compared to normal donor plasma cells are projected onto the network. Transcription factors with the largest difference in regulon activity in myeloma compared to normal donor plasma cells are highlighted in insets. g) tSNE plots showing regulon activity (area under the curve, AUC) of plasma cell lineage transcription factors XBP1, IRF4, PRDM1, and FOS in single myeloma and normal plasma cells.



Extended Data Fig. 6. Quality assessment and filtering of single cell ATAC data.

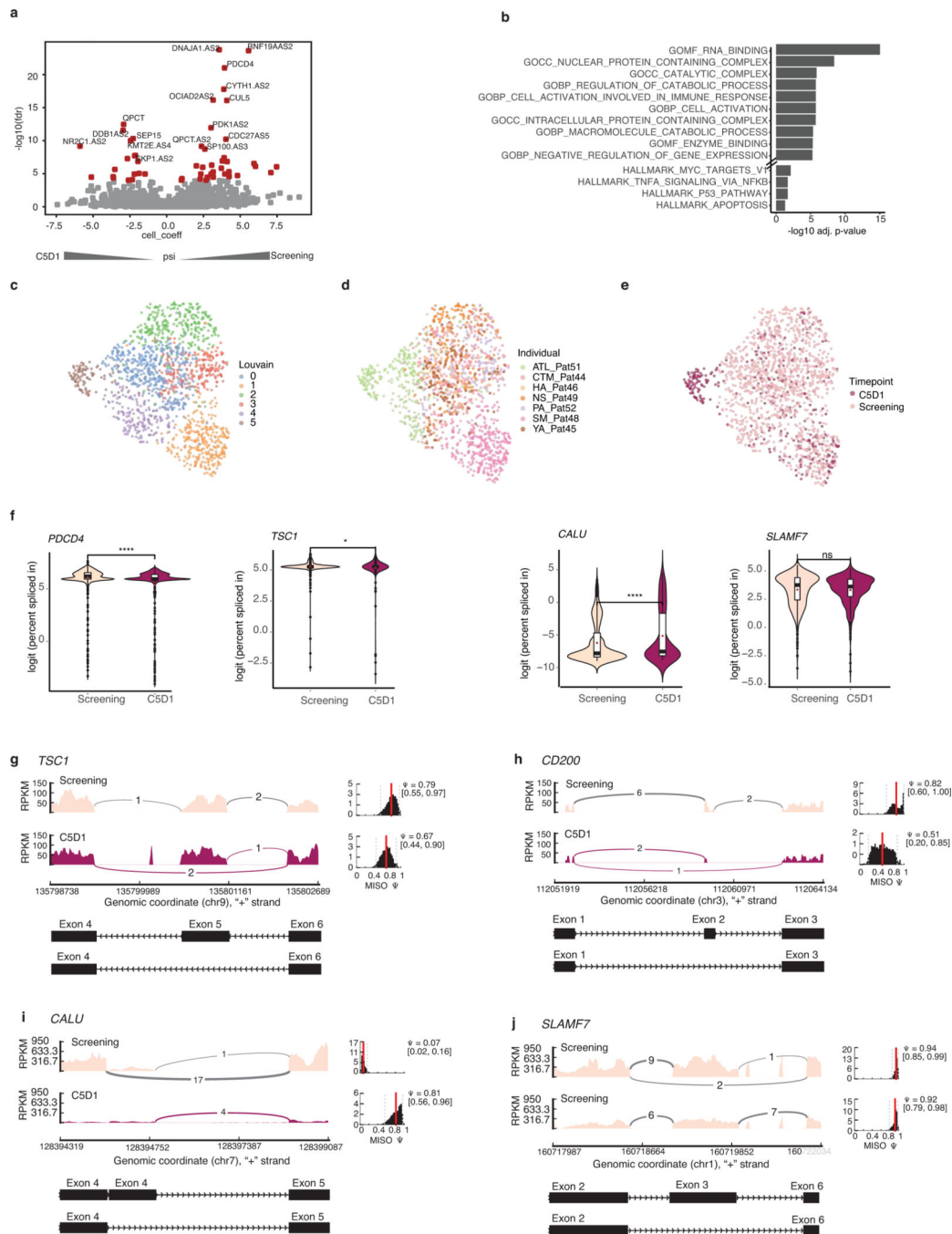
a) Filtering of single cell ATAC profiles based on TSS enrichment and number of unique nuclear fragments. b) Fragment length distribution of filtered scATAC profiles showing characteristic distribution with nucleosome-free region and mononucleosome peaks. c) TSS enrichment after filtering per sample. d) Number of unique fragments per sample. e,f) tSNE plot colored by clusters (e) or individual (f). g,h) tSNE plots following batch effect removal by Harmony colored by clusters (g) or individual (h). i-k) Defining cluster identities following integration with scRNA-seq data. i) tSNE plot colored by predicted cell type identities. j) tSNE plot showing cell type identities by cluster. k) Heatmap showing confusion matrix for predicted cell type identities by cluster. l-o) Tracks with aggregated ATAC profiles for each cluster for marker genes *SDC1*, *LYZ*, *MS4A1* and *FCGR3A*, respectively. p-s) tSNE plots colored by gene scores for marker genes *SDC1*, *LYZ*, *MS4A1* and *FCGR3A*, respectively.



Extended Data Fig. 7. Annotation of peaks from aggregated scATAC data.

a) Processing of scATAC profiles. b) Intersection of peaks in aggregate scATAC samples showing the overlap of peaks between MM and normal donor (ND) samples. c) Number of peaks significantly different by DESeq2 enriched in ND or MM (FDR = 0.1, absolute log2FC = 1). d) Heatmap showing differentially accessible peaks from c) in ND and MM, sorted by the ND sample. e) ChromHMM state annotation in aggregate scATAC samples. Depicted is the fraction of peaks in each of the indicated states. f) Number of peaks in ChromHMM state 13 corresponding to heterochromatin. **p = 0.0029 by two-sided

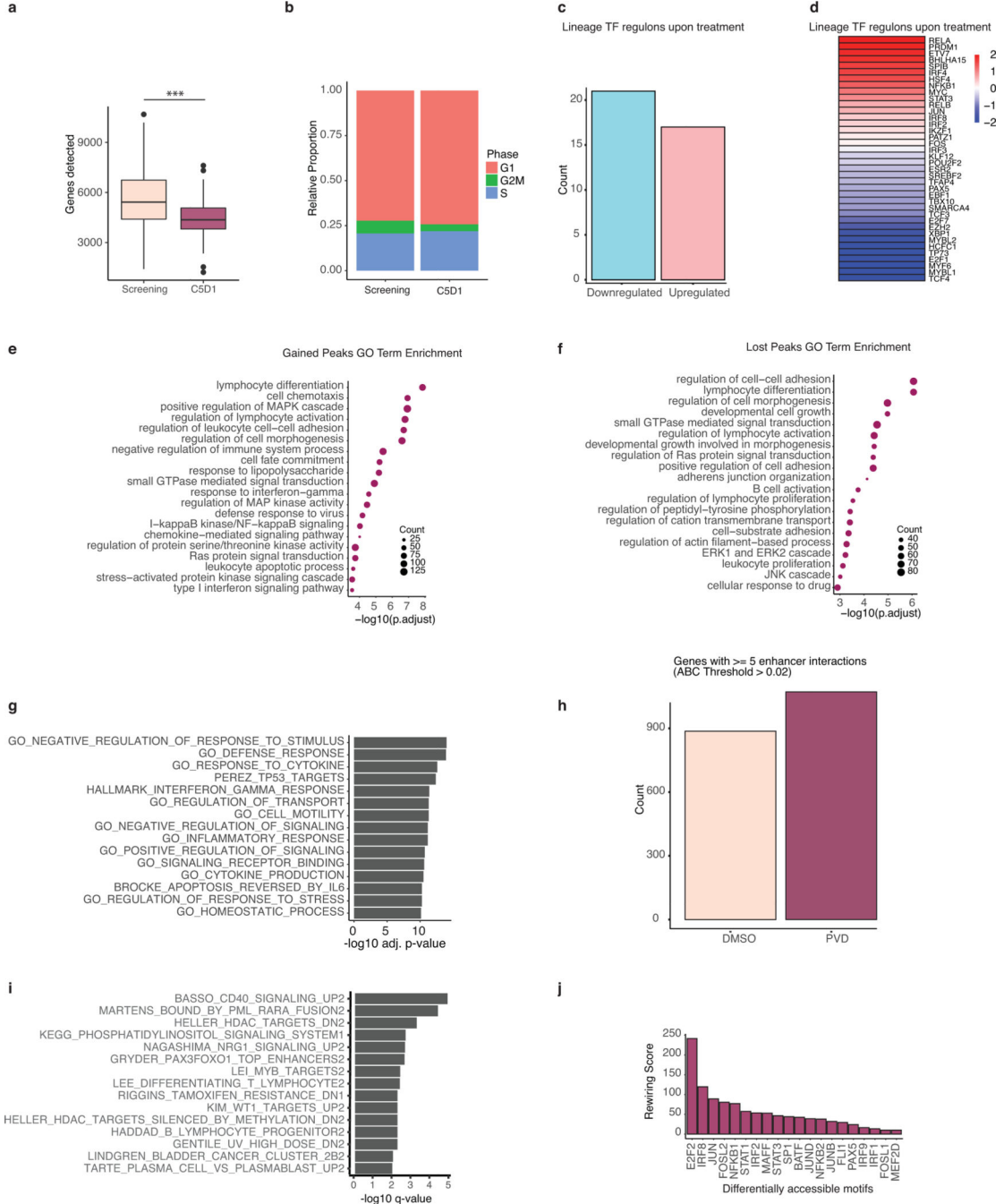
t-test. g) Fraction of peaks in indicated genomic regions. h) Boxplot comparing distance to TSS in ND vs MM. **** $p < 2.2e-16$ by two-sided Wilcoxon test. Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. n=46,935 peaks and 68,752 peaks, respectively. i) Heatmap showing accessibility of enhancers (n=15,748) associated with the top multi-enhancer genes, sorted by the ND sample. j) Barplot showing genes with 2 enhancer interactions by ABC model. k) Heatmap showing multi-enhancer genes by ABC model (n=16,635), sorted by the ND sample. l) Barplot showing gene set enrichment analysis for multi-enhancer genes in MM by ABC model. m) Barplot showing differential motif enrichment analysis of the single cell ATAC-Seq dataset comparing myeloma cells with normal donor plasma cells. Shown are the top 40 differentially enriched transcription factor motifs ordered by FDR. n) Venn diagram showing overlap of differentially enriched motifs and rewired transcription factors determined by DyNet algorithm.



Extended Data Fig. 8. Alternative splicing following treatment in MM.

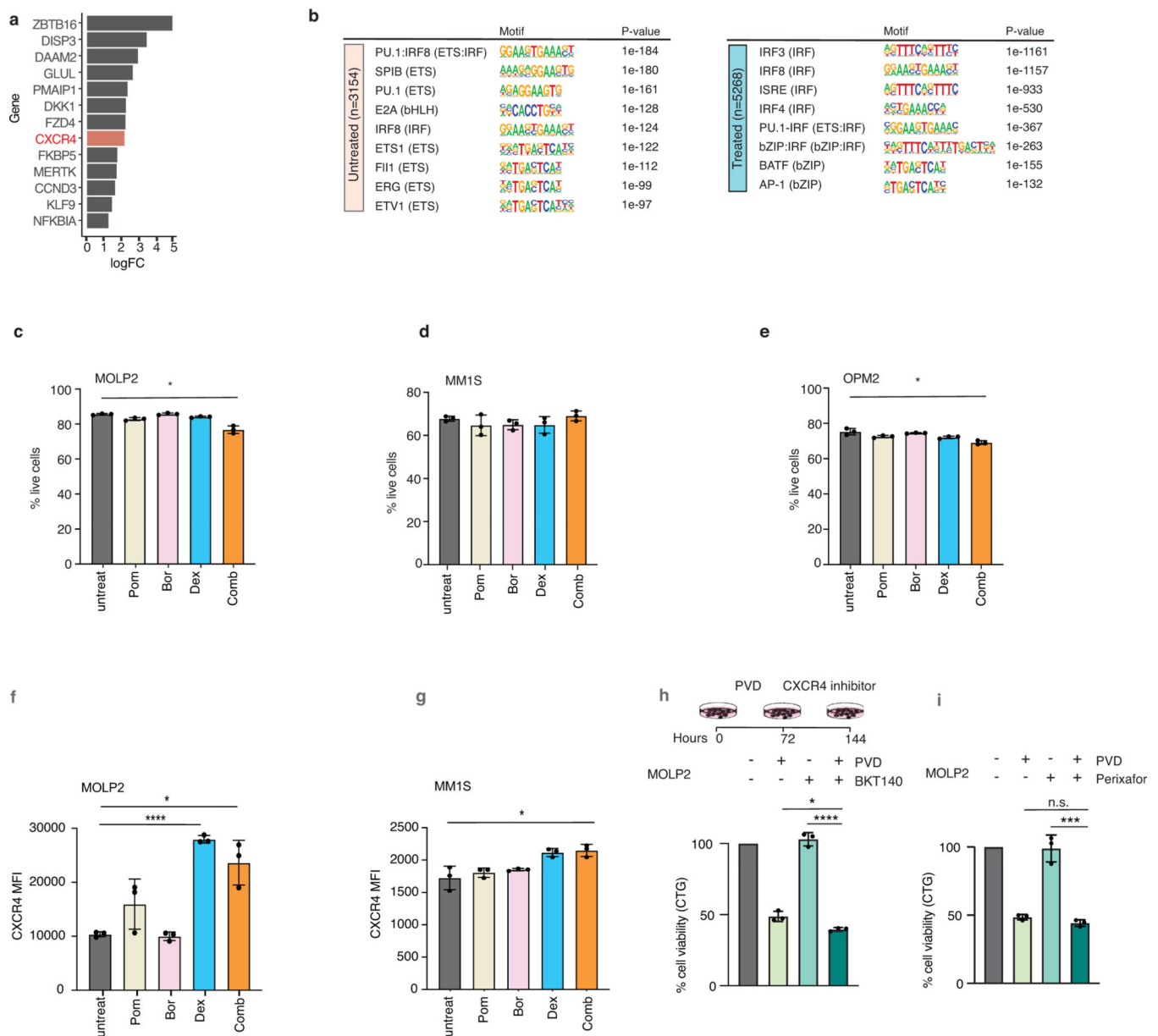
a) Quantification of exon inclusion/exclusion ratio (percent of spliced in, psi). Volcano plot showing differential splicing at C5D1 vs screening timepoints with FDR < 0.05. b) Barplot showing gene set enrichment analysis of differentially spliced transcripts. c) UMAP colored by Louvain clusters based on calculated psi (percent of spliced in) values. d) psi-based UMAP colored by individual. e) psi-based UMAP colored by timepoint. f) Violin plots showing the single cell distribution of logit (percent spliced in) values at screening and C5D1 for selected differentially spliced transcripts. Boxplots show the median and

interquartile range, whiskers extend to 1.5x the interquartile range. n= 1,374 cells. ****
 p 0.0001 and * p 0.05 by two-sided Wilcoxon test. g-j) Miso plots visualizing splice junctions and potential exon skipping events in differentially spliced transcripts for *TSC1* (g), *CD200* (h) and *CALU* (i) as well as *SLAMF7* (j). Splice probabilities are shown on the right.



Extended Data Fig. 9. Transcriptional diversity following treatment.

a) Number of genes expressed in primary MM cells before and after treatment ($p < 2.2 \times 10^{-16}$ by two-sided t-test). Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. $n = 1,374$ cells. b) Bar graph showing relative proportion of cells in each cell cycle phase at screening and C5D1. c) Shown are the lineage TF regulons downregulated and upregulated upon treatment. d) Change in regulon activity of lineage TFs upon treatment. e,f) GO-term enrichment of ATAC-Seq peaks gained (e) and lost (f) in MOLP2 cells following 72h treatment with PVD. g) Gene set enrichment analysis of the top 500 genes gaining enhancers following treatment with PVD. h) Genes with 5 enhancer interactions by ABC model. i) Gene set enrichment analysis of genes gaining enhancer interactions following treatment with PVD by ABC model. j) Top rewired TFs with differentially accessible motifs upon treatment.



Extended Data Fig. 10. Surface marker expression in MM.

a) Genes upregulated upon treatment. Shown is a barplot with log₂ fold change values compared to screening timepoint. b) Motif enrichment in differential ATAC-Seq peaks in MMCL MOLP2 after PVD treatment (right), comparing to untreated (left). p values are calculated using binomial test. c-e) Live cell counts following 72h of treatment with pomalidomide (Pom), bortezomib (Bor), dexamethasone (Dex) and combination of all three drugs (PVD) in MM cell lines MOLP2 (c), MM1.S (d) and OPM2 (e) as a percentage of total cell numbers. f,g) Quantification of CXCR4 surface levels by flow cytometry following treatment with pomalidomide (Pom), bortezomib (Bor), dexamethasone (Dex) and combination of all three drugs (PVD) in myeloma cell lines MOLP2 (f) and MM1.S (g). c-g) Significance was assessed using a two-sided t-test with Welch's correction, with * p < 0.05 and **** p < 0.0001. Data are presented as mean ± SD, n=3 replicates. h,i) Cell viability following treatment with CXCR4 inhibitors BKT140 (h) and Plerixafor (i) following pre-treatment with PVD in MMCL MOLP2. Data are presented as mean ± SD, n=3 replicates. n.s. p>0.05, * p < 0.05, *** p < 0.001 and **** p < 0.0001 by two-sided t-test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank Dr. Anthony Letai for discussion and critical comments on the manuscript. This project was supported by grants from the National Cancer Institute (K08CA191026) (J.G.L.), (K08CA191091) (B.K.), the Anna Fuller Fund (J.G.L.), and a Scholar Award by the V Foundation for Cancer Research (J.G.L.). K.C.A. is an American Cancer Society Clinical Research Professor. J.F. acknowledges support from the German Research Foundation (DFG) and the American Association for Cancer Research (AACR).

References

- Sharma SV et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* 141, 69–80, doi:10.1016/j.cell.2010.02.027 (2010). [PubMed: 20371346]
- Hangauer MJ et al. Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. *Nature* 551, 247–250, doi:10.1038/nature24297 (2017). [PubMed: 29088702]
- Shaffer SM et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546, 431–435, doi:10.1038/nature22794 (2017). [PubMed: 28607484]
- Risom T. et al. Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. *Nat Commun* 9, 3815, doi:10.1038/s41467-018-05729-w (2018). [PubMed: 30232459]
- Verfaillie A. et al. Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat Commun* 6, 6683, doi:10.1038/ncomms7683 (2015). [PubMed: 25865119]
- Wouters J. et al. Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. *Nat Cell Biol* 22, 986–998, doi:10.1038/s41556-020-0547-3 (2020). [PubMed: 32753671]
- LaFave LM et al. Epigenomic State Transitions Characterize Tumor Progression in Mouse Lung Adenocarcinoma. *Cancer Cell* 38, 212–228 e213, doi:10.1016/j.ccell.2020.06.006 (2020). [PubMed: 32707078]

8. Paiva B. et al. Differentiation stage of myeloma plasma cells: biological and clinical significance on behalf of GEM/PETHEMA cooperative study groups Europe PMC Funders Group. *Leukemia* 31, 382–392, doi:10.1038/leu.2016.211 (2017). [PubMed: 27479184]
9. Joseph D, Gonsky JP & Blain SW Macrophage Inhibitory Factor-1 (MIF-1) controls the plasticity of multiple myeloma tumor cells. *PLoS One* 13, e0206368, doi:10.1371/journal.pone.0206368 (2018).
10. Picelli S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* 9, 171–181, doi:10.1038/nprot.2014.006 (2014). [PubMed: 24385147]
11. Fan J. et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat Methods* 13, 241–244, doi:10.1038/nmeth.3734 (2016). [PubMed: 26780092]
12. Fernández JM et al. The BLUEPRINT Data Analysis Portal. *Cell Syst* 3, 491–495.e495, doi:10.1016/j.cels.2016.10.021 (2016). [PubMed: 27863955]
13. Ledergor G. et al. Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nat Med* 24, 1867–1876, doi:10.1038/s41591-018-0269-2 (2018). [PubMed: 30523328]
14. Broyl A. et al. Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. *Blood* 116, 2543–2553, doi:10.1182/blood-2009-12-261032 (2010). [PubMed: 20574050]
15. Kinker GS et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet* 52, 1208–1218, doi:10.1038/s41588-020-00726-6 (2020). [PubMed: 33128048]
16. Hay SB, Ferchen K, Chetal K, Grimes HL & Salomonis N. The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp Hematol* 68, 51–61, doi:10.1016/j.exphem.2018.09.004 (2018). [PubMed: 30243574]
17. La Manno G. et al. RNA velocity of single cells. *Nature* 560, 494–498, doi:10.1038/s41586-018-0414-6 (2018). [PubMed: 30089906]
18. Gulati GS et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 367, 405–411, doi:10.1126/science.aax0249 (2020). [PubMed: 31974247]
19. Teschendorff AE & Enver T. Single-cell entropy for accurate estimation of differentiation potency from a cell’s transcriptome. *Nat Commun* 8, 15599, doi:10.1038/ncomms15599 (2017). [PubMed: 28569836]
20. Chen W. et al. Single-cell landscape in mammary epithelium reveals bipotent-like cells associated with breast cancer risk and outcome. *Commun Biol* 2, 306, doi:10.1038/s42003-019-0554-8 (2019). [PubMed: 31428694]
21. Setz CS et al. PI3K-Mediated Blimp-1 Activation Controls B Cell Selection and Homeostasis. *Cell Rep* 24, 391–405, doi:10.1016/j.celrep.2018.06.035 (2018). [PubMed: 29996100]
22. Aibar S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 14, 1083–1086, doi:10.1038/nmeth.4463 (2017). [PubMed: 28991892]
23. Kas K. et al. ESE-3, a novel member of an epithelium-specific ets transcription factor subfamily, demonstrates different target gene specificity from ESE-1. *J Biol Chem* 275, 2986–2998, doi:10.1074/jbc.275.4.2986 (2000). [PubMed: 10644770]
24. Joshi K, Chain BM, Peggs KS & Quezada SA TEAD transcription factors are required for normal primary myoblast differentiation in vitro and muscle regeneration in vivo. *PLoS Genet* 13, e1006600, doi:10.1371/journal.pgen.1006600 (2017). [PubMed: 28178271]
25. Goenawan IH, Bryan K. & Lynn DJ DyNet: visualization and analysis of dynamic molecular interaction networks. *Bioinformatics* 32, 2713–2715, doi:10.1093/bioinformatics/btw187 (2016). [PubMed: 27153624]
26. Cuitiño MC et al. Two Distinct E2F Transcriptional Modules Drive Cell Cycles and Differentiation. *Cell Rep* 27, 3547–3560.e3545, doi:10.1016/j.celrep.2019.05.004 (2019).
27. Fulciniti M. et al. Non-overlapping Control of Transcriptome by Promoter- and Super-Enhancer-Associated Dependencies in Multiple Myeloma. *Cell Rep* 25, 3693–3705.e3696, doi:10.1016/j.celrep.2018.12.016 (2018). [PubMed: 30590042]

28. Jin Y. et al. Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. *Blood* 131, 2138–2150, doi:10.1182/blood-2017-09-808063 (2018). [PubMed: 29519805]
29. McLean CY et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28, 495–501, doi:10.1038/nbt.1630 (2010). [PubMed: 20436461]
30. Fulco CP et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* 51, 1664–1669, doi:10.1038/s41588-019-0538-0 (2019). [PubMed: 31784727]
31. Schep AN, Wu B, Buenrostro JD & Greenleaf WJ chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* 14, 975–978, doi:10.1038/nmeth.4401 (2017). [PubMed: 28825706]
32. Salomonis N. et al. Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc Natl Acad Sci U S A* 107, 10514–10519, doi:10.1073/pnas.0912260107 (2010). [PubMed: 20498046]
33. Lee JH, Kwon EJ & Kim DH Calumenin has a role in the alleviation of ER stress in neonatal rat cardiomyocytes. *Biochem Biophys Res Commun* 439, 327–332, doi:10.1016/j.bbrc.2013.08.087 (2013). [PubMed: 24012670]
34. Campbell KS, Cohen AD & Pazina T. Mechanisms of NK Cell Activation and Clinical Activity of the Therapeutic SLAMF7 Antibody, Elotuzumab in Multiple Myeloma. *Front Immunol* 9, 2551, doi:10.3389/fimmu.2018.02551 (2018). [PubMed: 30455698]
35. Roccaro AM et al. CXCR4 regulates extra-medullary myeloma through epithelial-mesenchymal-transition-like transcriptional activation. *Cell Reports* 12, 622–635, doi:10.1016/j.celrep.2015.06.059 (2015). [PubMed: 26190113]
36. Ghobrial IM et al. A Phase Ib/II Trial of the First-in-Class Anti-CXCR4 Antibody Ulocuplumab in Combination with Lenalidomide or Bortezomib Plus Dexamethasone in Relapsed Multiple Myeloma. *Clin Cancer Res* 26, 344–353, doi:10.1158/1078-0432.Ccr-19-0647 (2020). [PubMed: 31672767]
37. DiPersio JF et al. Plerixafor and G-CSF versus placebo and G-CSF to mobilize hematopoietic stem cells for autologous stem cell transplantation in patients with multiple myeloma. *Blood* 113, 5720–5726, doi:10.1182/blood-2008-08-174946 (2009). [PubMed: 19363221]
38. Mandal M. et al. CXCR4 signaling directs Igk recombination and the molecular mechanisms of late B lymphopoiesis. *Nat Immunol* 20, 1393–1403, doi:10.1038/s41590-019-0468-0 (2019). [PubMed: 31477919]
39. Barwick BG, Gupta VA, Vertino PM & Boise LH Cell of Origin and Genetic Alterations in the Pathogenesis of Multiple Myeloma. *Front Immunol* 10, 1121, doi:10.3389/fimmu.2019.01121 (2019). [PubMed: 31231360]
40. Ordonez R. et al. Chromatin activation as a unifying principle underlying pathogenic mechanisms in multiple myeloma. *Genome Res* 30, 1217–1227, doi:10.1101/gr.265520.120 (2020). [PubMed: 32820006]
41. Flavahan WA, Gaskell E. & Bernstein BE Epigenetic plasticity and the hallmarks of cancer. *Science* 357, doi:10.1126/science.aal2380 (2017).
42. Corces MR et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 14, 959–962, doi:10.1038/nmeth.4396 (2017). [PubMed: 28846090]
43. Bolger AM, Lohse M. & Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120, doi:10.1093/bioinformatics/btu170 (2014). [PubMed: 24695404]
44. Dobin A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21, doi:10.1093/bioinformatics/bts635 (2013). [PubMed: 23104886]
45. Li B. & Dewey CN RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323, doi:10.1186/1471-2105-12-323 (2011). [PubMed: 21816040]
46. Anders S, Pyl PT & Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169, doi:10.1093/bioinformatics/btu638 (2015). [PubMed: 25260700]

47. Aran D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 20, 163–172, doi:10.1038/s41590-018-0276-y (2019). [PubMed: 30643263]
48. Lun AT, McCarthy DJ & Marioni JC A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* 5, 2122, doi:10.12688/f1000research.9501.2 (2016). [PubMed: 27909575]
49. Patel AP et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401, doi:10.1126/science.1254257 (2014). [PubMed: 24925914]
50. Tirosh I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (New York, N.Y.)* 352, 189–196, doi:10.1126/science.aad0501 (2016).
51. Muller S, Cho A, Liu SJ, Lim DA & Diaz A. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. *Bioinformatics* 34, 3217–3219, doi:10.1093/bioinformatics/bty316 (2018). [PubMed: 29897414]
52. Bolotin DA et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nature Methods* 12, 380–381, doi:10.1038/nmeth.3364 (2015). [PubMed: 25924071]
53. Gaujoux R. & Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 11, 367, doi:10.1186/1471-2105-11-367 (2010). [PubMed: 20598126]
54. Robinson MD, McCarthy DJ & Smyth GK edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140, doi:10.1093/bioinformatics/btp616 (2010). [PubMed: 19910308]
55. McCarthy DJ, Chen Y. & Smyth GK Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40, 4288–4297, doi:10.1093/nar/gks042 (2012). [PubMed: 22287627]
56. Puram SV et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* 171, 1611–1624.e1624, doi:10.1016/j.cell.2017.10.044 (2017). [PubMed: 29198524]
57. Subramanian A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545–15550, doi:10.1073/pnas.0506580102 (2005). [PubMed: 16199517]
58. Liberzon A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740, doi:10.1093/bioinformatics/btr260 (2011). [PubMed: 21546393]
59. Liberzon A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 1, 417–425, doi:10.1016/j.cels.2015.12.004 (2015). [PubMed: 26771021]
60. Wolf FA, Angerer P. & Theis FJ SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19, 15, doi:10.1186/s13059-017-1382-0 (2018). [PubMed: 29409532]
61. Aran D, Hu Z. & Butte AJ xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology* 18, 1–14, doi:10.1186/s13059-017-1349-1 (2017). [PubMed: 28077169]
62. 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, doi:10.1038/nbt.4096 (2018). [PubMed: 29608179]
63. Schubert M. et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun* 9, 20, doi:10.1038/s41467-017-02391-6 (2018). [PubMed: 29295995]
64. Anand P. et al. Single-cell RNA-seq reveals developmental plasticity with coexisting oncogenic states and immune evasion programs in ETP-ALL. *Blood* 137, 2463–2480, doi:10.1182/blood.2019004547 (2021). [PubMed: 33227818]
65. Shannon P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498–2504, doi:10.1101/gr.1239303 (2003). [PubMed: 14597658]
66. Rauscher B, Henkel L, Heigwer F. & Boutros M. Lineage specific core-regulatory circuits determine gene essentiality in cancer cells. *bioRxiv preprint*, doi:doi: 10.1101/609552 (2019).
67. Huang Y. & Sanguinetti G. in *bioRxiv* (2020).
68. Katz Y, Wang ET, Airolidi EM & Burge CB Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* 7, 1009–1015, doi:10.1038/nmeth.1528 (2010). [PubMed: 21057496]

69. Wang ET et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476, doi:10.1038/nature07509 (2008). [PubMed: 18978772]
70. Granja JM et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* 53, 403–411, doi:10.1038/s41588-021-00790-6 (2021). [PubMed: 33633365]
71. Ewels PA et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* 38, 276–278, doi:10.1038/s41587-020-0439-x (2020). [PubMed: 32055031]
72. Jia Y. et al. Myeloma-specific superenhancers affect genes of biological and clinical relevance in myeloma. *Blood Cancer J* 11, 32, doi:10.1038/s41408-021-00421-7 (2021). [PubMed: 33579893]
73. Kaya-Okur HS et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10, 1930, doi:10.1038/s41467-019-09982-5 (2019). [PubMed: 31036827]
74. Pliner HA et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* 71, 858–871 e858, doi:10.1016/j.molcel.2018.06.044 (2018). [PubMed: 30078726]

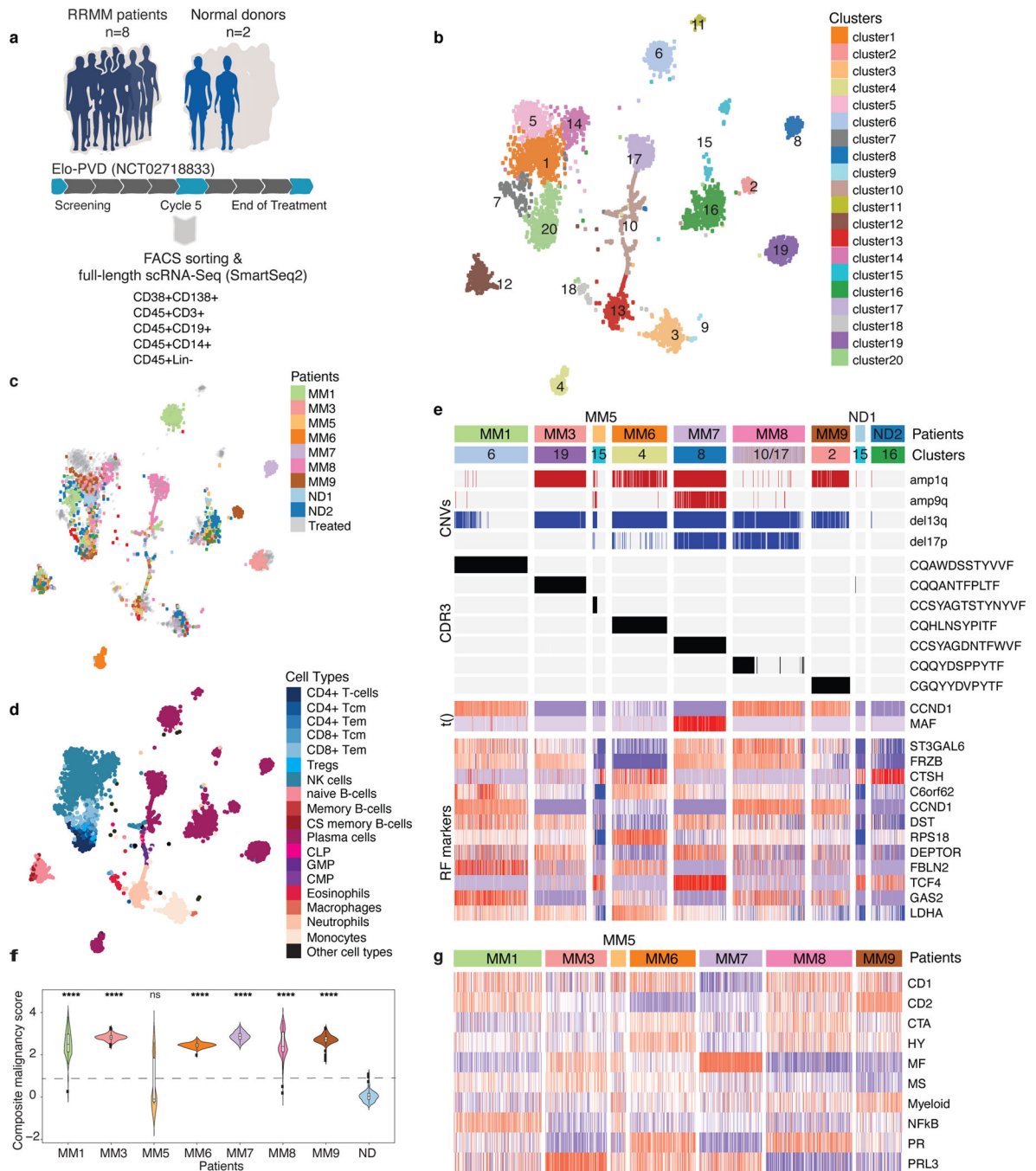


Fig. 1: Co-expression of multiple transcriptional signatures in MM cells.

a) Schematic illustrating experimental setup. b) T-stochastic network embedding (tSNE) of the processed single cell RNA-Seq dataset color-coded by clusters identified by PAGODA2. c) tSNE plot color-coded by individual. d) Cell type annotation based on 21 immune cell populations from the BLUEPRINT dataset. e) Classification of malignant and normal plasma cells based on CNVs, CDR3 sequence, expression of translocation targets, and genes identified by random forest (RF) model that best discriminate between normal and malignant cells. f) Composite malignancy score to distinguish myeloma cells from normal plasma

cells. Cells from normal donors are aggregated to facilitate comparison. Boxplots show median and interquartile range, whiskers extend to 1.5x interquartile range (n= 1,162 cells. ns p>0.05 and **** p = 0.0001 by Wilcoxon test compared to ND). g) Co-expression of multiple established myeloma gene expression classifiers in single myeloma cells.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

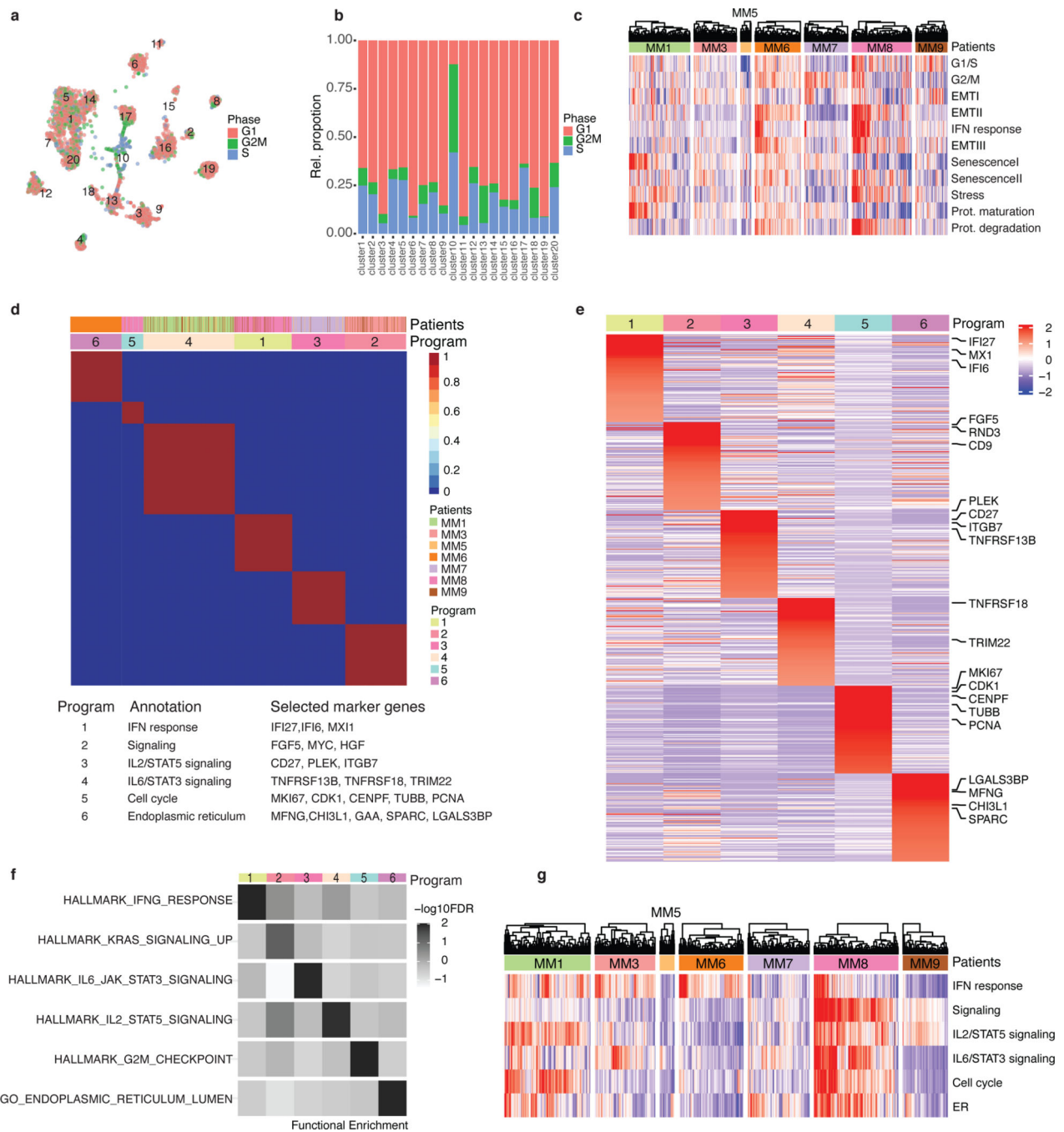


Fig. 2: Transcriptional heterogeneity in single myeloma cells.

a) tSNE plot colored by cell cycle phase predicted by SEURAT. b) Stacked bar plot showing the relative proportion of cells in each cell cycle phase per cluster. c) Heatmap showing expression scores for recurrent heterogeneity programs in MM patients (scores identified across 198 cell lines reflecting 22 cancer types¹⁵). d) Consensus matrix depicting pairwise similarities between NMF programs ordered by hierarchical clustering. Six clusters corresponding to the six identified programs and assignment of patients are indicated on top. Functional annotation and selected marker genes are shown below. e) Heatmap

showing the top 50 genes based upon the highest NMF scores selected as signature genes for each program with selected genes labeled. f) Functional enrichment ($-\log_{10}$ FDR) of heterogeneity programs with six annotated gene sets. g) Heatmap showing expression scores for the six heterogeneity programs identified in d). The enlarged heatmaps in c and g are provided in Supplementary Information.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

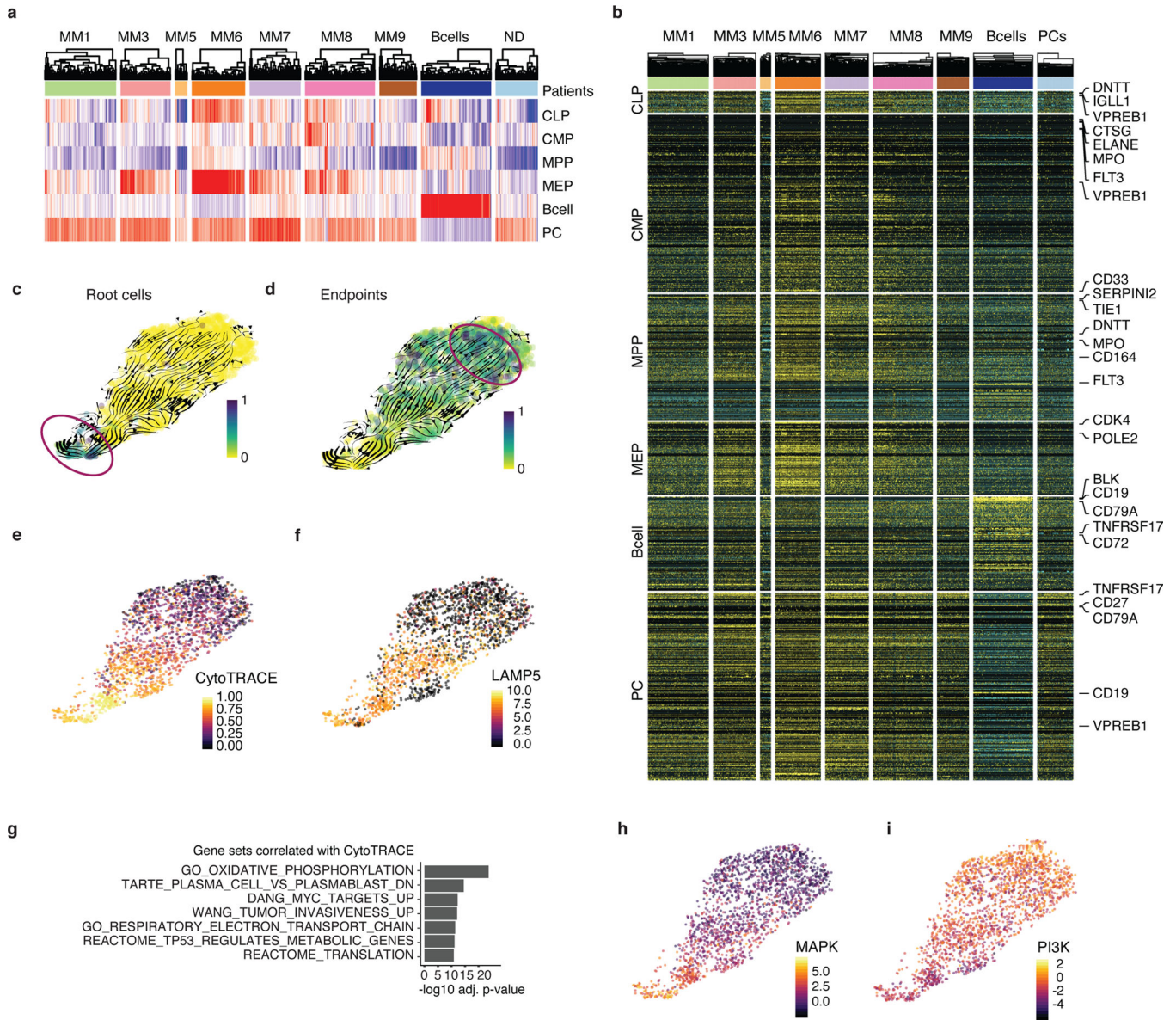


Fig. 3: Lineage infidelity - transcriptional states diverge towards immature states.
a) Heatmap showing lineage scores for selected cell types from the BLUEPRINT dataset in single myeloma cells from patients, plasma cells from normal donors or B cells. CLP, common lymphoid progenitor; CMP, common myeloid progenitor; MPP, multipotent progenitor; MEP, megakaryocyte-erythroid progenitor; PC, plasma cell. b) Heatmap of single myeloma cells from patients or normal donor plasma cells showing expression of genes corresponding to selected cell types from the BLUEPRINT dataset. The enlarged heatmaps in a-b are provided in Supplementary Information. c,d) RNA velocity estimates, root and end states projected onto UMAP. Root and endpoints are circled. e) CytoTRACE values showing distribution of differentiation states from most immature (highest values) to most mature (lowest values). f) Expression of marker gene *LAMP5* correlated with immaturity (log₂-transformed counts). g) Gene set enrichment analysis of genes correlated

with high CytoTRACE (associated with immaturity). h,i) Signaling activity scores across single cells for MAPK (h) and PI3K (i) signaling pathways, projected on UMAP.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

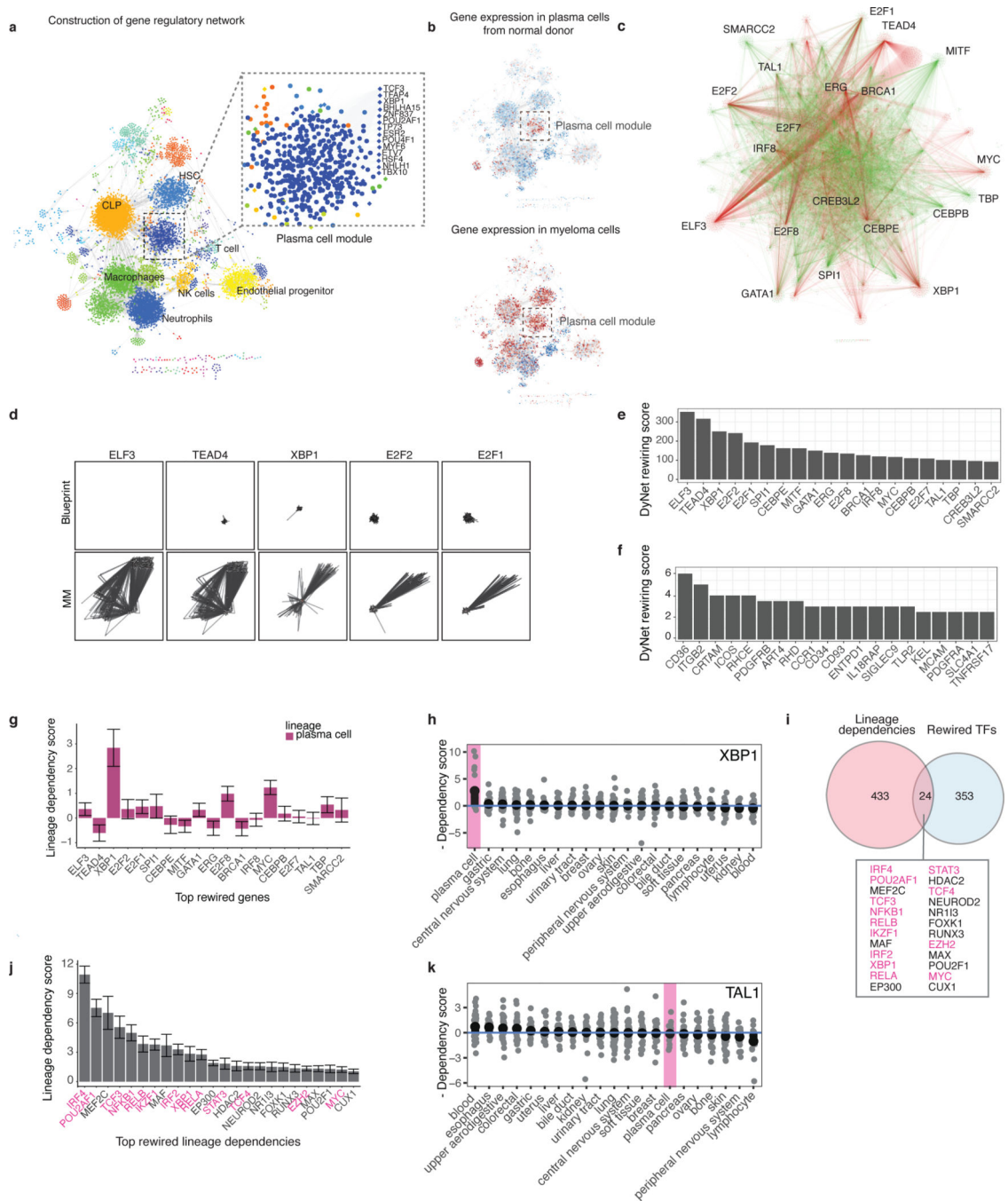


Fig. 4: Differential regulon activity and transcriptional rewiring in MM underlie alternate differentiation states.

a) Gene regulatory network (GRN) constructed based on transcriptional modules present in normal hematopoiesis. Highlighted are plasma cell module and its dominating transcription factors. b) Gene expression in normal donor plasma cells (top) and myeloma cells (bottom) projected onto normal hematopoiesis network. The plasma cell module is highlighted. c) GRN constructed exclusively from myeloma cells (red) overlaid with GRN from normal hematopoietic cells (green). d) Regulons, i.e. transcription factors with their target genes, as part of network layouts in normal hematopoiesis (Blueprint, top) and myeloma (bottom).

Regulons were chosen based on rewiring score (e). e, f) Transcription factors (e) or CD markers (f) ordered by rewiring score determined by the DyNet algorithm. g-k) Dependency data were downloaded from the Cancer Dependency Map (<https://depmap.org/portal/>). g) Lineage dependency score in MM cell lines for top rewired TFs in plasma cells. Data are presented as mean values \pm standard error. n=20 cell lines. h) Scaled dependency scores for 675 cell lines from 21 different lineages ordered by dependency for XBP1. Myeloma cell lines (plasma cell lineage) are highlighted in pink. i) Venn diagram showing overlap between lineage dependency genes and rewired TFs. Shown are TFs in overlap. j) TFs in overlap between lineage dependencies and rewired TFs ordered by lineage dependency score in MM cell lines. Data are presented as mean values \pm standard error. n=20 cell lines. i,j) Lineage TFs are highlighted in pink. k) Scaled dependency scores for 675 cell lines from 21 different lineages ordered by dependency for TAL1. Myeloma cells (plasma cell lineage) are highlighted in pink.

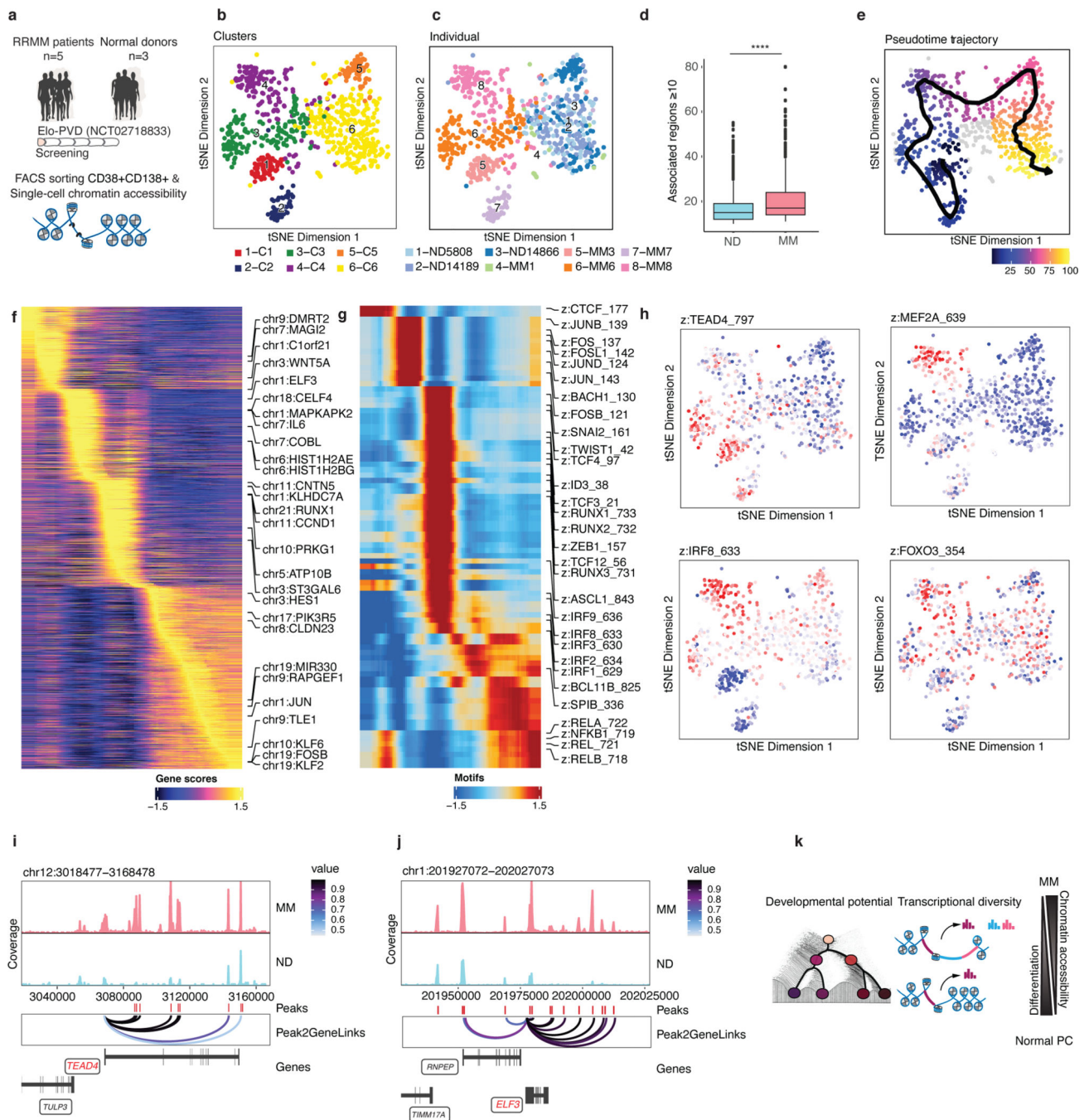


Fig. 5: Transcriptional reprogramming in MM and associated changes in chromatin accessibility and enhancer rewiring.

a) Schematic illustrating experimental setup. b) tSNE plot showing clustering of single myeloma cells colored by clusters based on single cell ATAC (scATAC). c) tSNE plot showing clustering of single myeloma cells colored by patient based on single cell ATAC (scATAC). d) Number of associated enhancers per gene for genes associated with 10 enhancers in normal donor plasma cells or patient-derived myeloma cells based on association by GREAT. Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. n=1,020 and 1,159 genes, respectively. p=5.5e-08 by two-

sided Wilcoxon test. e) Pseudotime trajectory. f) Heatmap showing correlation of gene scores with pseudotime. Pseudotime increases from left to right. g) Heatmap showing correlation of motif accessibility with pseudotime. Pseudotime increases from left to right. h) tSNE plots of single cell ATAC-Seq colored based on bias-corrected z-scores of selected rewired transcription factors that bind differentially accessible motifs. Shown are TEAD4, FOXO3, MEF2A and IRF8. i, j) Peak-to-gene-linkage determined based on integrated single cell ATAC for TEAD4 locus (i) and ELF3 locus (j). k) Model of developmental potential and transcriptional diversity in MM.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

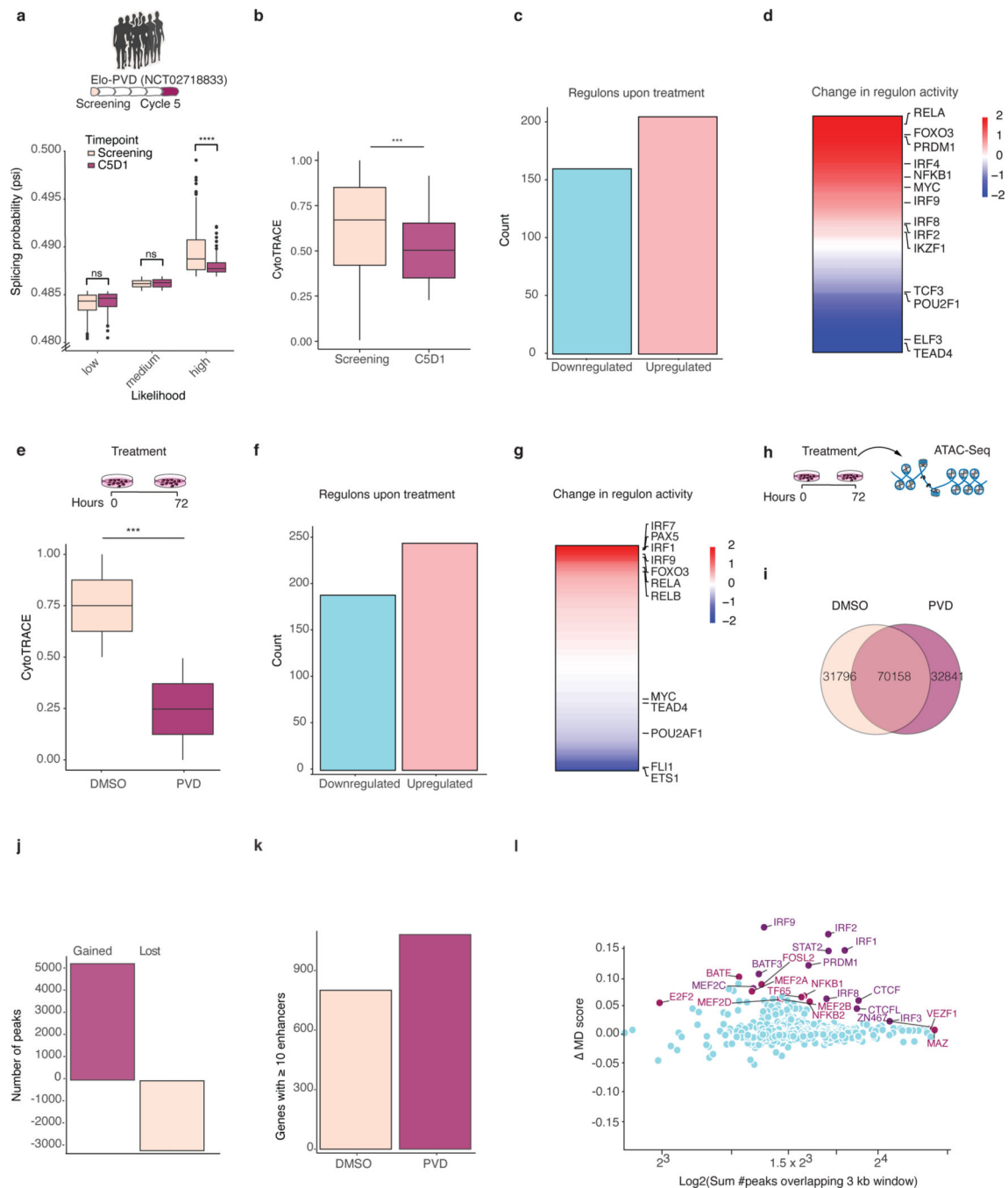


Fig. 6: Treatment reduces developmental potential while increasing regulon activity.

a-d) Comparison of primary MM cells on treatment (C5D1) compared to screening timepoint. a) Exon inclusion/exclusion ratio (percent of spliced in, psi) quantified at screening vs. C5D1 for transcripts with low, medium or high splicing probability. Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. n=1,374 cells. **** p = 0.0001 by two-sided Wilcoxon test. b) CytoTRACE values in MM cells before and after treatment (p<2.2e-16 by two-sided t-test). Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. n= 1,374

cells. c) Number of regulons downregulated (160/365) and upregulated (n= 205/365) upon treatment. d) Change in regulon activity upon treatment. Selected lineage and non-lineage transcription factors are highlighted. e-g) MMCL MOLP2 was treated with pomalidomide, bortezomib and dexamethasone (PVD) for 72h prior to single cell RNA sequencing. e) CytoTRACE values for cells treated with PVD or DMSO ($p < 2.2 \times 10^{-16}$ by two-sided t-test). Boxplots show the median and interquartile range, whiskers extend to 1.5x the interquartile range. n= 192 cells. f) Regulons downregulated and upregulated in cells treated with PVD compared to DMSO. g) Change in regulon activity in cells treated with PVD compared to DMSO. Selected lineage TF regulons are highlighted. h) ATAC sequencing was performed on MOLP2 cells after 72h of treatment with PVD or DMSO. i) Venn diagram illustrating overlap of peaks in cells treated with DMSO or PVD. j) Number of peaks gained and lost after treatment determined by DESeq2. k) Number of genes associated with 10 enhancers in cells treated with DMSO or PVD by GREAT. l) Motif enrichment in MOLP2 cells after 72h of treatment with PVD vs DMSO. Enriched TF motifs are highlighted in red.

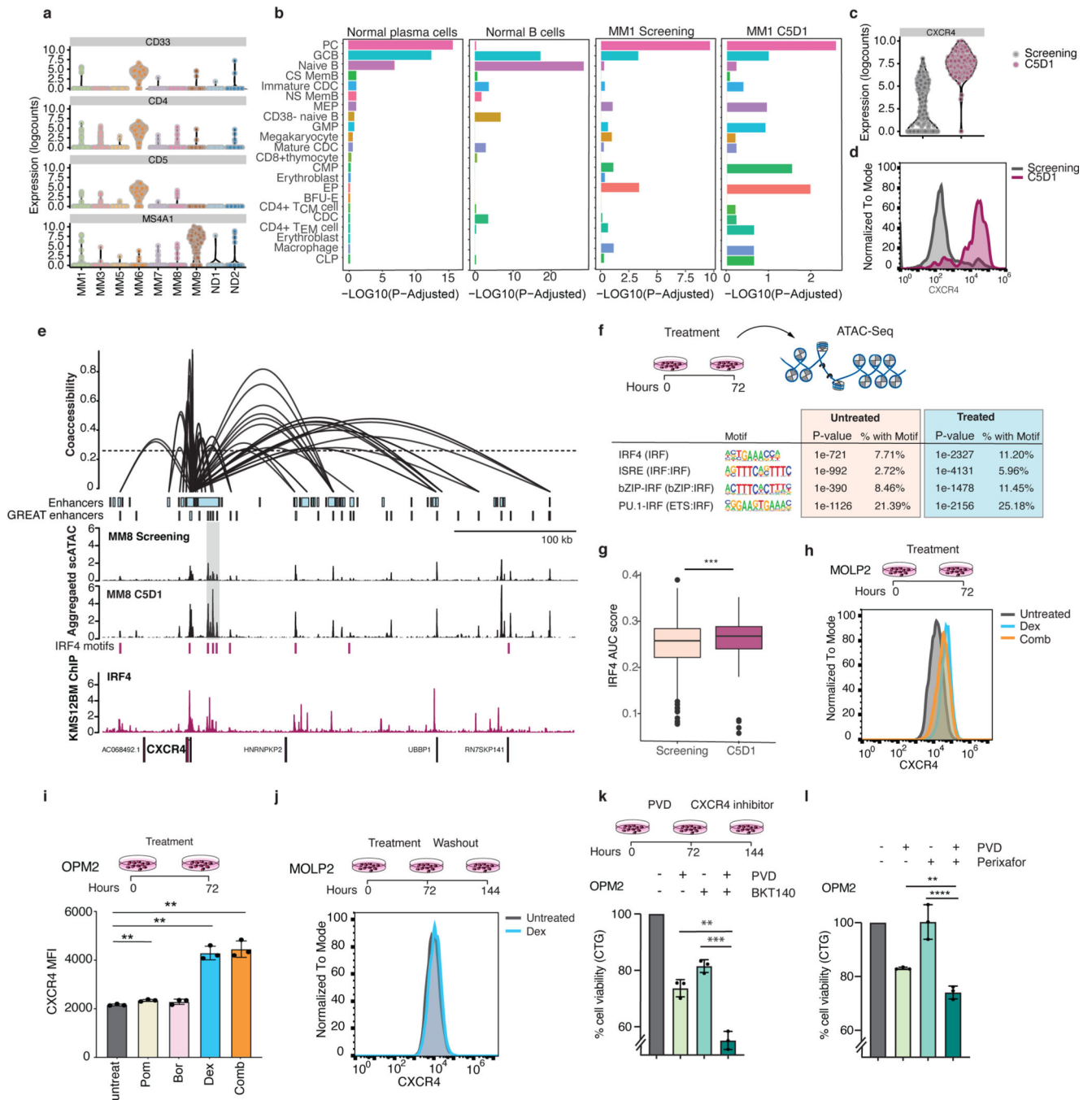


Fig. 7: Inducible and stable expression of putative immunotherapy targets on myeloma cells.
 a) Expression of selected surface markers not found in normal plasma cells in single myeloma (MM) cells or normal donor plasma cells (ND). b) Deconvolution of surface marker expression in normal donor plasma cells, B cells and myeloma cells from patient MM1 at screening and at cycle 5 day 1 (C5D1) based on enrichment of surface marker signatures derived from BLUEPRINT dataset. c,d) *CXCR4* mRNA (c) and surface protein (d) expression in single myeloma cells from patient MM1 before treatment and at C5D1. e) Co-accessibility determined based on single cell ATAC data showing the *CXCR4* locus.

Aggregated scATAC tracks show chromatin accessibility upstream of CXCR4 at screening and C5D1 with differential peaks highlighted in grey. IRF4 motifs in peaks and IRF4 ChIP peaks from KMS12BM are displayed below. f) Motif enrichment of IRF4 and IRF composite elements in ATAC-peaks in MMCL MOLP2 after 72h of treatment. p values calculated using binomial test. g) IRF4 regulon activity in primary MM cells ($p=1.3e-05$ by two-sided t-test). Boxplots represent the median and interquartile range, whiskers extend to 1.5x the interquartile range. $n= 1,374$ cells. h) Treatment of MOLP2 myeloma cells with dexamethasone (Dex) or combination of dexamethasone, pomalidomide, bortezomib (PVD) resulting in upregulation of CXCR4 at the cell surface. i) Quantitative analysis of CXCR4 surface protein expression changes on OPM2 myeloma cells with drug exposure by flow-cytometry (MFI = mean fluorescence intensity). Data are presented as mean \pm SD, $n=3$ independent experiments. Significance was assessed using an unpaired two-sided t-test with Welch's correction; ** $p < 0.01$. j) Drug removal resulting in downregulation of CXCR4 surface expression in MOLP2 cells. k,l) Cell viability following treatment with CXCR4 inhibitors BKT140 (k) and Plerixafor (l) following pre-treatment with PVD in MMCL OPM2. Data are presented as mean \pm SD, $n=3$ independent experiments. ** $p < 0.01$ and *** $p < 0.001$ by unpaired two-sided t-test.