bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Systematically characterizing the roles of E3-ligase family members in inflammatory responses with massively parallel Perturb-seq

Kathryn Geiger-Schuller[1,2*], Basak Eraslan[1,2*], Olena Kuksenko[2], Kushal K. Dey[3], Karthik A. Jagadeesh[2], Pratiksha I. Thakore[1,2], Ozge Karayel[1], Andrea R. Yung[1], Anugraha Rajagopalan[1], Ana M Meireles[1], Karren Dai Yang[4], Liat Amir-Zilberstein[2], Toni Delorey[2], Devan Phillips[1], Raktima Raychowdhury[2], Christine Moussion[1], Alkes L. Price[3,5], Nir Hacohen[2], John G. Doench[2], Caroline Uhler[4,6], Orit Rozenblatt-Rosen[1,2], Aviv Regev[1,2,7,8]

[1]Genentech Research and Early Development, South San Francisco, CA

[2]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge MA

[3]T.H. Chan Harvard School of Public Health, Harvard University, Boston, MA

[4]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA

[5]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA

[6]Broad Institute of MIT and Harvard, Cambridge MA

[7]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA

[8]Lead contact: regev.aviv.sc@gmail.com

Correspondence to: Kathryn Geiger-Schuller (gschuller.kathryn@gmail.com), Basak Eraslan (basak.eraslan.sc@gmail.com), Aviv Regev (regev.aviv.sc@gmail.com)

*These authors contributed equally to this work

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## ABSTRACT

E3 ligases regulate key processes, but many of their roles remain unknown. Using Perturb-seq, we interrogated the function of 1,130 E3 ligases, partners and substrates in the inflammatory response in primary dendritic cells (DCs). Dozens impacted the balance of DC1, DC2, migratory DC and macrophage states and a gradient of DC maturation. Family members grouped into co-functional modules that were enriched for physical interactions and impacted specific programs through substrate transcription factors. E3s and their adaptors co-regulated the same processes, but partnered with different substrate recognition adaptors to impact distinct aspects of the DC life cycle. Genetic interactions were more prevalent within than between modules, and a deep learning model, comβVAE, predicts the outcome of new combinations by leveraging modularity. The E3 regulatory network was associated with heritable variation and aberrant gene expression in immune cells in human inflammatory diseases. Our study provides a general approach to dissect gene function.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## INTRODUCTION

Despite systematic efforts in genetics and genomics, our knowledge of the function of many genes remains limited, especially those from large gene families, where the general molecular function may be inferred from sequence features, but the specific mechanism, biological process, cellular context and physiological impact of individual genes and their combinations often remain partly or completely unknown. Multiple approaches can help decipher individual gene function, including Genome-Wide Association Studies (GWAS) to relate causal genetic variants to quantitative traits[1]; forward genetic screens followed by phenotypic assessment, including cell viability, images or molecular profiles[2]; and guilt-by-association approaches, based on similarity in molecular patterns between a gene of interest and other genes. Despite their power and utility, each of these approaches has some limitations. Genetic association studies are often limited by the modest effect sizes associated with common variants in human populations[3]; correlative approaches provide suggestive associations but not causal relations; and forward genetic screens have typically had to pre-define the phenotype of interest, such as cell viability[4] or a cellular marker[5,6]. Finally, all approaches are challenged at deciphering genetic interactions, due to limited statistical power or experimental scale to test exponentially large numbers of combinations.

Large protein families, such as E3 ubiquitin ligases ("E3s"), are an important example of this challenge. The human genome codes for >600 different E3s responsible for catalyzing the ligation of ubiquitin (Ub) to substrates in almost every biochemical pathway[7], including many immune functions[8]. GWAS have implicated variants in E3 ligase genes in many diseases, including inflammatory and autoimmune diseases[9,10], but characterizing their specific cellular

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

roles remains challenging, as is determining their inter-relationships. In particular, dendritic cells (DCs) play a key role in initiating immune responses, including multiple inflammatory and autoimmune diseases[11,12], and heritable variants in multiple genes in DCs contribute to their aberrant inflammatory signaling in disease, including several axes that may be targeted therapeutically[13]. While previous studies implicate different E3 ligases in the DC inflammatory response to lipopolysaccharide (LPS)[5], relatively little is known about the E3 circuit in these or other primary immune cells, as many studies focus on transformed cancer cell lines.

Recent advances in combining pooled genetic perturbation screens with rich, single cell readouts, especially single cell RNA-seq (scRNA-seq) in Perturb-Seq assays, open opportunities to dissect the function of genes and gene combinations from large gene families[14–24]. In Perturb-Seq screens, the perturbed genes can be partitioned into co-functional modules, based on the similarity of their effects across many genes, controlling co-regulated programs of genes responding similarly across multiple perturbations[16,17]. Moreover, any diversity in cell subsets or processes, such as the cell cycle or differentiation, is naturally captured in the screen[16]. Most Perturb-Seq studies to date, and especially the very few done in primary cells, have analyzed up to a few dozen perturbations[14–22], with a recent notable screen of thousands of genes but only in a transformed cell line[23].

Here, we used Perturb-seq at scale to screen the function of each of 1,130 genes spanning E3 ligases, E3-like proteins and their interacting partners and substrates in the inflammatory response to stimulation with LPS in primary mouse bone marrow derived dendritic cells (BMDCs). The cells in one experiment spanned DC1, DC2, and migratory DC (mDC) subsets, a

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

gradient of DC maturation, and a range of gene programs, allowing us to decipher the role of E3 ligases in multiple contexts simultaneously. A regulatory model distinguished six co-functional modules impacting eleven different programs of co-regulated genes, showing which E3 ligases, adaptors and substrate recognition adaptor proteins regulate each process in DCs, capturing known associations and making many new functional annotations. Computational integration of the regulatory (genetic) model with physical protein-protein interactions and transcription factor (TF)-target genes relations shows that the regulatory model was congruent with physical mechanistic interactions. E3s and their physically interacting partners were enriched in the same co-functional module, and the programs they regulated were enriched for targets of specific TFs, highlighting multiple paths from E3s and complex members through TFs to different DC processes they regulate. Moreover, the circuit was modular, such that Cullin-RING ligases (the largest subfamily; >200 members) and their known adaptors co-regulated the same processes, but combined with different substrate recognition adaptor proteins to control distinct aspects of the DC life cycle. The circuit was also congruent with human disease biology, with both co-functional modules and their regulated programs enriched in heritability for risk immune and inflammatory diseases. Leveraging our large screen's design to also randomly sample combinations of perturbations, we found that intra-module (non-additive) genetic interactions are more prevalent than inter-module ones and then used the modular architecture to devise com$\beta$VAE, a new deep learning model that predicts genetic interactions. Our study offers a general scalable approach to dissect gene function, including physiological functions for dozens of E3s and related genes, congruent physical circuits, principles of modularity in the regulatory and molecular architecture, characterization and prediction of genetic interactions, and an overall

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

model of the inflammatory response to help interpret human genetics signal at unprecedented resolution.

## RESULTS

### A systematic screen of E3 ligases in immune dendritic cells

To study the role and circuitry of members of the large gene family of E3 ligases in inflammatory responses, we curated a comprehensive set of 1,137 genes encoding E3s and related proteins to screen them by Perturb-seq (**Table S1-2**; **STAR Methods**). These included 382 genes with 'E3 activity' designation in the Integrated Annotations for Ubiquitin and Ubiquitin-like Conjugation Database (iUUCD [7]), such as proteins from RING, HECT, U-box, PHD, RBR, and other families; 509 genes with 'E3 adaptor' designation in iUUCD, such as those from DWD, BTB, APC, Cullin, BC-box, F-box, DDB1, and other families; 6 genes with an annotated ubiquitin binding domain and one with a ubiquitin-like domain (Rbbp6) also from iUUCD; and 239 genes based on an NCBI search for 'E3 activity', capturing other enzymes in the ubiquitylation cascade (E1s, E2s), known E3 substrates (*e.g.*, Tp53, Ikbke, and Cebpb), and members of relevant signaling networks regulated by E3 ligases (*e.g.*, TLR and TNF signaling). We synthesized and screened guides targeting 1,130 of the genes (we could not confidently design gRNAs for 7 putative pseudogenes).

We optimized Perturb-seq for large scale screening and used it to screen the 1,130 genes, perturbed by 3,390 targeting guides, profiling 838,201 individual BMDCs, after 3 hours of treatment with LPS (**Figure 1A**). We chose the 3h time point because the DC transcriptional response to LPS has a single wave, peaking around 3hr for mature mRNA at both the population

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and single cell level, and is optimal for observing RNA expression effects[5,25–30], whereas protein level changes occur later[26]. We isolated 54 million cells from the bone marrow of Cas9 transgenic mice[31], treated them with GM-CSF to differentiate them towards BMDCs and, on day 2, transduced them at a planned multiplicity of infection (MOI) of 0.2 with a pooled lentiviral library of 3,390 sgRNAs targeting the 1,130 genes (3 guides per gene) and 330 control guides (165 targeting intergenic regions and 165 non-targeting). We designed a new Perturb-Seq vector (pRDA122; **Figure 1A**) with a capture sequence appended to the 3' terminus of the gRNA scaffold sequence to enable direct capture of CRISPR gRNAs for scRNA-seq (**STAR Methods**), compatible with feature barcoding for droplet-based 3' scRNA-seq[24]. We continued to differentiate the transduced cells for another 7 days, when they are predominantly BMDCs[25,30,32,33], and then treated them with LPS. At 3 hours post-LPS treatment, we sorted mKate2$^{+}$Cas9-2A-EGFP$^{+}$ cells, loaded 2.32 million onto 46 channels using cell hashing and "super-loading"[34,35] (**Figure S1A-D**), and performed scRNA-seq. We obtained profiles from 1,071,671 non-empty droplets (EmptyDrops [36] FDR < 0.01, **STAR Methods**), containing 838,201 single cells and 233,470 multiplets, followed by dedicated PCR to detect the guide RNA in each cell (**STAR Methods**). After QC and guide assignment, we retained 519,535 single cell profiles assigned with one or more gRNA (targeting or control; detected MOI of 1.2) for our main analyses, followed by analysis of 177,871 cells with multiple perturbations with dedicated methods below, as well as applying a compressed sensing approach to all the multiplets in a companion study (Yao et al, bioRxiv 2023). As reference, we also profiled a total of 10,347 unperturbed cells, encompassing both LPS stimulated and unstimulated cells.

**An end-to-end computational pipeline for large Perturb-seq screens**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To analyze large screens, we developed PerturbDecode, for end-to-end, automated analysis in four consecutive pillars (**Figure 1B and S1E, STAR Methods**): (**1**) data QC and preprocessing; (**2**) identification of the effects of perturbations on genes; (**3**) learning the regulatory topology of perturbed and impacted genes for single and/or combinatorial perturbations; and (**4**) relating the regulatory (genetic) topology to physical interactions and human genetics (**STAR Methods**). Briefly, in pre-processing, PerturbDecode addresses hashing and feature barcoding assignment, detects depleted feature barcodes and cells with multiple guides and removes outliers. Next, it efficiently identifies impactful guides and impacted cells[16], by estimating the effect of each guide on each gene with a negative binomial linear regression model, accounting for confounders (**STAR Methods**). It retains impactful guides defined as those with effects that are significantly similar to those of at least one other guide targeting the same gene (*vs*. a background of all guides), and iteratively identifies and retains impacted cells[16]. To determine the impact at the level of perturbed genes, PerturbDecode uses a mixed effects negative binomial linear regression model, with cell subsets inferred by initial clustering as random effects and the feature barcode matrix as the set of fixed effects, correcting for confounders (**STAR Methods**). It retains perturbations affecting a significant (FDR < 0.1) number of genes (**STAR Methods**), clusters the resulting coefficient matrix to generate co-functional modules and co-regulated programs, and decomposes the matrix by independent component analysis (ICA) to infer latent independent processes that could generate the observed perturbation responses. For combinatorial perturbations, it estimates the effect of genetic interactions and predicts the impact of unseen combinations. Finally, it includes multiple *post-hoc* analytics to relate the learned model to molecular circuits (protein-protein and TF-target interactions) and to human genetics data.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Perturb-Seq screen yields impactful perturbations consistently across guides**

Estimating the impact of each of the 3,204 targeting guides (detected in at least 20 cells) on the expression of each of 6,685 genes (expressed in at least 5% of cells) showed that the correlation in effect sizes between cells with guides targeting the same gene was significantly higher than between guides targeting different genes or between targeting and control guides (P-value < $10^{-16}$, Kolmogorov-Smirnov (KS) test, **Figure S1F**). Focusing on concordant guides (**STAR Methods**), we retained 2,263 KO guides targeting 1,031 genes for downstream analysis and learned a model at the targeted gene (rather than guide) level (**STAR Methods**). The average number of genes significantly affected by each perturbed gene (FDR < 0.1, **STAR Methods**) was significantly higher than for controls (36.91 *vs*. 0.78 (non-targeting) and 1.04 (intergenic) on average, P < $2.2*10^{-16}$, one-sided Wilcoxon rank-sum test, **Figure S1G, STAR Methods**). Of the 1,031 targeted genes, 544 were also among the 6,685 analyzed genes: the vast majority had 491 had a nominal negative effect on their own expression, 137 of them significantly (FDR< 0.1, **Figure S1H**), and only four had a significant positive effect. Expressed (detected) E3s affected significantly more genes than undetected ones (P < $10^{-4}$, one-sided non-parametric Wilcoxon test, **Figure S1J**) and the mean expression of the targeted (KO) gene in unperturbed cells was modestly but positively correlated with the number of genes impacted by their perturbation (Spearman's $\rho = 0.22$, P < $1.5*10^{-10}$, **Figure S1I**). Overall these results suggest that CRISPR-induced indels overall caused nonsense mediated decay (NMD) of the respective transcripts, for expressed E3s and family members, as well as with prior observations that some CRISPR-knockout generated indels having poor NMD [5,16] or even (futile) transcriptional compensation[37].

**DC1-, DC2-, migratory DC-, and macrophage-like cells are screened jointly**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

BMDC populations are heterogeneous, and previous studies[29,30,38–40], including our earlier Perturb-Seq screen in this system[16], all highlighted the presence of different subsets, including cells expressing macrophage-like signatures, and "cluster-disrupted" DCs[30,41]. Because Perturb-Seq characterizes any cell diversity *post hoc*, it assesses these multiple phenotypes simultaneously[16].

The 519,535 perturbed single cell profiles partitioned into ten clusters (**Figure 1C-H** and **S2A**, **Table S3** and **S4**), which included multiple DC2-like subsets (high expression of Cd9, Il1b, and Sirpa (but also Irf4 and Il6); **Figure 1C,D,** clusters 1-6; **Figure S2A,B**); migratory DC (mDC)-like cells (high expression of Ccr7, Fscn1, Il4i, Socs2, and Relb (but not Pdl2); cluster 7, **Figure 1C,E** and **S2A,C**); DC1-like cells (high expression of Clec9a, Xcr1, Batf3, Irf8, Tap1, Flt3, and Wdfy4 but also Cd8a and Tlr3, and no expression of pDC marker genes (Tcf4, Tlr7, and Tlr9; additionally, Siglec-H was not among detected transcripts) **Figure 1C,F** and **S2A,D,E**), and macrophage-like cells (high expression of M1 markers Il6, Il1b, and Fpr2 and moderate expression of M2 markers Chil3, Fn1, and Mc1; **Figure 1C** and **S2A,F,G**). The three main DC subsets also followed a gradient of expression from more mature DC-like (most prominent in mDCs) to macrophage-like (more prominent in some DC2s) signatures (**Figure 1H** and **S2H,I**). Each DC2-like cell subset had additional distinguishing markers (**Figure S2A**). Cycling cells (**Figure 1C,G**, Cluster 9, 13%) expressed signatures of either DC2s or macrophages/monocytes (**Figure 1D,G,H** and **S2H,I**). These cycling monocyte-derived macrophages are consistent with previous reports[38,42]. A very small subset (460 cells; 0.08%) expressed a neutrophil signature (**Figure 1C** and **S2A**, cluster 10). DC2-like, DC1-like, mDC-like, and cycling monocyte-derived macrophages were also present in genetically unperturbed cells, with and without LPS

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

stimulation (**Figure S2J-W**), but at different proportions, with the fraction of macrophage-like cycling cells lower in genetically unperturbed LPS stimulated cells than in either unperturbed unstimulated DCs or perturbed stimulated DCs (**Figure S2W**, P-value < $2.2*10^{-16}$, one-sided Fisher's exact test, **STAR Methods**), and the fraction of mDCs in unperturbed cells (stimulated and unstimulated) higher than in perturbed stimulated DCs (**Figure S2W**, P-value < $2.2*10^{-16}$, one-sided Fisher's exact test). The increase in monocyte-derived macrophages in genetically perturbed cells (**Figure S2W**), suggests that the macrophage-like state is typically repressed by LPS stimulation but remains accessible upon perturbation.

**Multiple E3 ligases impact specific DC cell subsets or differentiation**

Sixty-five (65) genes were targeted by two or more guides that were significantly depleted from BMDCs *vs.* the input guide library distribution, suggesting that these genes are essential for BMDC survival and proliferation[5,16] (**Table S5**, **STAR Methods**; P-value<0.05, considering a background of the corresponding change in control guides). Indeed, these were enriched for regulation of cell division (*e.g.*, Aurka, Myc, Plk1, Pou5f1, Prc1, Tle6, Wdr5, Ybx1), including Mdm2 (all three guides depleted), an E3 ligase that ubiquitylates p53 as an active heterodimer with Mdm4 and is essential for cell cycle regulation[43]. Some perturbations affected the proportions of all cycling cells of a specific type (**Figure S2X**), such as enrichment in cycling macrophages of guides targeting the tumor suppressor and E3 ligase substrate Trp53, and enrichment in cycling DC1-like cells of guides targeting the substrate Nf1, a tumor suppressor in myeloid cells that affects growth sensitivity to GM-CSF[44] and regulates proliferation regulators[45]. Because different subsets were enriched for cycling cells (**Figure 1G**), some of these perturbations also shifted subset proportions. For example, consistent with previous

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reports[46], Trp53 KO were enriched in cycling monocyte-derived macrophages, and depleted in non-cycling DC2s; and its negative regulator E3 ligase Mdm4 had the opposite pattern (**Figure 1J** and **S2X**).

Perturbation in 64 genes, including 29 E3s and complex members significantly affected the relative proportion of the main cell subsets, especially the balance of DC2s *vs*. cycling macrophage-like cells (**Figure 1I,J**, **STAR Methods, Table S6**; FDR < 0.15, one-sided Fisher's exact test). Guides depleted in cycling macrophage-like cells *vs*. DC2s included the cullin E3 ligase Cul3 and its substrate adaptor Keap1, which ubiquitylates Nrf2, regulating redox response and inflammation[47,48]; Vhl and Rack1, which togerher complex with Cul2[49] and ubiquitylate Hif1a; and Mdm4, which interacts with Mdm2 to stimulate p53 ubiquitylation[50]. Vhl, Cul3, Keap1 and Mdm4 were not previously described as regulators of DC2 differentiation, but deletion of Nrf2 impacts both tissue resident and BMDC functions[51] and Rack1 depletion in myeloid cells protects against viral infection in mice without affecting the number of $F4/80^{+}CD11b^{+}$ myeloid cells[52]. Conversely, guides enriched in cycling monocyte-derived macrophages *vs*. DC2s included those targeting Cul5, March6, and Wdr26, E3s not previously recognized as regulators of DC differentiation (**Figure 1I,J**).

Other enrichments and depletions further highlight the roles for E3s and related proteins in specific cell subsets. For example, mDCs were enriched for guides targeting the E3 ligase Traf2 and the TF CEBPB and depleted for guides targeting DIDO. Traf2 plays a role in TNF-mediated NF-KB and MAP kinase signaling, and TNF injection can increase DC trafficking to lymph nodes[53], suggesting a potential role for Traf2 in regulating DC migration. Enrichment of guides

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

targeting CEBPB in mDCs is consistent with our earlier findings and CEBPB's natively low level in those cells (**Figure 1I,J**)[16]. Concomitantly, DC2.2s where CEBPB expression is higher, are enriched for guides targeting Rfwd2 (COP1), which in turn targets CEBPB for degradation[54]. Conversely, guides targeting Dido1, a gene previously characterized only in embryonic stem cell differentiation[55,56], were depleted in mDCs, suggesting a novel role in mDC differentiation (**Figure 1I,J**). In another example, guides targeting the E3 ligase Arih2, the E3 ligase substrate Nf1, and the E2 enzyme Ube2f were enriched in DC1s and macrophages and depleted in DC2s (**Figure 1I,J**). This is consistent with and expands on their established roles: Arih2 ubiquitylates substrates of Ube2f, the Nedd8 E2 that mediates neddylation of Cul5-Rbx2[57], causing degradation of the NF-KB inhibitor Ikbb in DCs[58], and inactivating mutations in Nf1 lead to uncontrolled cell growth and plasmocytoid DC (pDC) neoplasms[59].

Many guides, including those targeting members of the WD-repeat protein subfamily, were specifically enriched in different DC2 subsets, including multiple members of a single complex in the same subset, suggesting that distinct DC2 subsets are controlled by different pathways (FDR < 0.15, one-sided Fisher's exact test, **Figure 1J** and **S2Y**, **Table S6**). For example, perturbations of each of the four members of the KEAP1:NEDD8-CUL3:RBX1 complex were specifically enriched in DC2.4s. DC2.3s were enriched for guides targeting E3 ligase substrates and members of the mTOR pathway, including mTORC2 (Mtor and Rictor) and mTORC1 (Mtor and Rptor), consistent with and refining the role of mTor signaling in regulating differentiation and immune functions in DCs[60]. DC2.2s were enriched for guides targeting Fbox E3 ligase component members, including Fbxo42, a regulator of the TAK1 pathway that activates p38[61], and Fbxw7, a regulator of antiviral immunity in macrophages[62]. DC2.1s were enriched for guides

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

targeting Eif3f, Naca, Rfwd2 (Cop1), and Pparg. In addition to its role as a TF, Pparg is an E3 ligase that targets p65[63] and regulates lipid metabolism in DCs, leading to ferroptosis and impairing maturation[64], and DC-T cell interactions in type-2 immunity[65]. DC2.5s were enriched for guides targeting Trim33, an E3 that interacts with Pu.1 (SPI1) and regulates the NLRP3 inflammasome, LPS response, and macrophage activation[66,67], as well as for guides targeting Ube2i (Ubc9), an E2 SUMO-conjugating enzyme with a role in lupus, regulating sumoylation-based suppression of type I IFN and pDCs[68]. Thus, we identified regulators of six DC2 substates, including multiple complex members similarly associated with the same state(s).

**Six co-functional modules of E3 ligases regulate eleven gene programs**

To relate these broad changes to regulatory mechanisms, we next learned a regulatory model associating 329 impactful perturbed genes (affecting the level of at least 15 genes) to 1,041 significantly impacted targets (affected by at least four of the 329 perturbations), and clustered the perturbed genes and impacted targets (**STAR Methods**) into six co-functional gene modules (M1-6) and eleven co-regulated gene programs (GP1-11), respectively (**Figure 2** and **Table S7-8, STAR Methods**).

The eleven programs (**Figure 2C,D** and **S3A-K, Table S8**) consisted of genes strongly co-regulated across the perturbations and were enriched for different immune and cellular processes. We annotated the programs by their enrichment in functional categories and previously described signatures (**Table S3,8**) as response to oxidative stress (GP1), ER stress response (GP2), pyruvate metabolism (GP3), motility and cell maintenance (GP4), protein homeostasis and phagocytosis (GP5), translation (high in mDCs and DC1s) (GP6), mDCs (GP7), TNF/LPS

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

response (GP8), autophagy regulation and inflammation (GP9), MHC-I antigen presentation (GP10), and DC2 and MHC-II antigen presentation (GP11)[39]. Some programs are expressed broadly across all cell subsets (*e.g.*, GP5, **Figure S3E,M**) and others are quite specific (*e.g.*, GP6 in DC1s and mDCs; **Figure S3F,M**). The programs were not necessarily independent of each other: some were regulated in anti-correlated ways (*e.g.*, opposite regulatory effects on GP6 (DC1/mDC expressed translation) *vs*. GP9 (autophagy and inflammation) and 11 (DC2 and MHC-II presentation)), while others were regulated in similar (though not identical) ways (*e.g.*, GP4 (motility and cell maintenance), GP5 (protein homeostasis and phagocytosis), and GP6 (translation)) (**Figure 2C**).

The programs refined ones we previously defined in the same system under perturbation of 24 TFs[16] (*e.g.*, P2 genes partition into mDC (GP7) and translation (GP6)) and uncovered new functional groupings (*e.g.*, DC2 MHC-II antigen presentation (GP11), response to ER stress (GP2), and pyruvate metabolism (GP3)) (**Figure S3L**), showing that the expanded scope and nature of perturbations could reveal additional regulatory processes.

**Regulatory patterns and targets reveal the functional roles and co-associations of E3s**

The six co-functional modules of 329 regulators each impacted a different subset of programs (**Figure 2C,D**), such that modules M1 and 2 were generally positively correlated with each other and negatively correlated with modules M3, 4, and 5, reflecting opposing functions.

The co-functional modules included 97 E3 ligases, 65 E3 complex members and 3 ubiquitin-like domain proteins, 85 of which did not have prior literature evidence in DCs or inflammation, and

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

can be deemed as novel functional annotations (**Table S9**), based on their co-membership and their target programs (**Figure 2D**). Three of the E3 ligase or complex members in the model are "authorities" in the E3 circuit, as their expression is impacted (positively and negatively) by a particularly high number of other E3s (**Figure 3A** and **S4A,B**): E3s Rack1 (which targets Hif1a and BimEL[49,52,69]) and Rnf128 (Grail; regulates Tbk1 and interferon and antiviral response[70–72]), and E3 complex member Socs3 (regulates Jak2 and gp130 and response to cytokines). All three have known roles in DC biology or inflammation: Rack1 suppresses type 1 IFN responses rendering mice more susceptible to viral infection[52], Rnf128 targets the DC maturation marker Cd83 for degradation[72], and Socs3 regulates JAK/STAT signaling, inflammation and macrophage polarization[73]. Notably, only 14 of 329 regulators were previously identified in a genome-wide CRISPR screen for regulators of TNF protein expression in the same system[5], and these were further distinguished by our model into different modules, with most (9 of 14) in M6 (*e.g.*, Ube2f, Rbck1, Rnf31, Spop, Traf6, and Nedd8 (positive regulators), and Rc3h1 (negative regulator)). This shows the expanded power of functional discovery based on comprehensive expression profiles compared to one highly validated reporter/marker.

M1 regulators strongly activated ER stress (GP2), protein homeostasis and phagocytosis (GP5) and translation (GP6) through the action of regulators of ribosome biogenesis (predicted E3s Bop1, Wdr36 and Nol10, Cul4 substrate receptor Dcaf11 (VprBP); processome members E3 adaptors Dcaf13 and Wdr33), including those involved in the stress response (predicted E3s Wdr43, Wdr75, and Utp18), and cellular migration (E3 Aamp, predicted E3 Bop1) (**Table S9**). This is consistent with the enrichment of ribosome biogenesis and stress genes in the translation

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(GP6) and ER stress (GP2) programs. Additionally, M1 regulators repressed the mDC (GP7), TNF/LPS response (GP8), and MHC-I antigen presentation (GP10) programs.

M2 regulators had generally similar impacts to those of M1, repressing MHC-I antigen presentation (GP10) and the TNF/LPS response (GP8) (albeit more weakly) and activating protein homeostasis and phagocytosis (GP5) and DC1/mDC expressed translation (GP6). Consistently, they included multiple negative regulators of NFkB and interferon signaling (E3 complex member Bid, CopA and Copb2, Ikbkg, the SUMO E3 HDAC4 and E3 Mib1 (which prevent or signal Ikba degradation, respectively), predicted E3 Nsmaf, E3 Ppp1r11 and the substrate Map3k7), and known regulators of cell differentiation in general (Bptf, Dido1, Gemin5, E3 and TF Zbtb7a), and of DC differentiation and maturation in particular (Ogt, Vdr (regulated by Mdm2), and TFs and predicted E3s Zbtb14 and Zbtb49), as well as the cell cycle and cell growth E3s APC (Anapc13, Cdc27, and Fzr1), CCNF, E4f1, Ring1, and Brca1; E3 adaptor Ddb1; E3 complex members Mdm4 and Pa2g4; DUB, Wdr48, and Fbxo11. Module members included multiple components of one complex and pathway, such as Cul4b and Ddb1 from the same cullin E3 complex (below).

M3 regulators had largely opposite effects to M2 and M1: they activated the TNF/LPS response (GP8) and autophagy and inflammation (GP9), and repressed translation (GP6) and mDCs (GP7). Module members included many known regulators of inflammation (substrate Gnb1, the E3 and TF Pparg; E3s Traf3 and Wdr82 (which regulates TRAF3)), mTOR signaling (E3s Rictor and Traf2, E3 interacting partner Mlst8; substrates Mtor, Rheb and Rptor)[74–77], autophagy (E3 Rnf216 (Triad3); predicted E3s Pik3r4 and Wdfy3), and ER transport (E3 Syvn1, substrates

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sec13 and Sec31a). In particular, the module includes Traf2, Pparg and Cebpb. CEBPB is known to activate DC2s and repress mDCs[16], and guides targeting Traf2 and CEBPB were consistently enriched in mDCs (**Figure S1I,J**). CEBPB and PPARG physically interact[78] and their targeting guides were enriched in CD2.2 cells. Other regulators include PAF1, which induced Tnf expression and inflammatory signaling (**Figure 2D**), and regulators of ER transport (Sec13, Sec31a, and Syvn1), all previously identified and validated as positive regulators of TNF expression in this system[5]. Our model now further relates them to mTOR pathway and autophagy regulators and to repressing mDC-like states.

M4 regulators activated ER stress (GP2), pyruvate metabolism (GP3), protein homeostasis and phagocytosis (GP5), TNF/LPS response (GP8), and DC2 MHC-II antigen presentation (GP11) and repressed oxidative stress (GP1) and translation (GP6). They included DNA repair and splicing regulators, such as the E3s Plrg1 and Prpf1 and the predicted E3 Cdc40 (Prp17), suggesting a link between DNA repair and splicing and metabolism/translational control.

M5 regulators activated autophagy and inflammation (GP9), MHC-I antigen presentation (GP10), and DC2 MHC-II antigen presentation (GP11), and repressed pyruvate metabolism (GP3), motility and cell maintenance (GP4), translation (GP6), and LPS/TNF response (GP8), through the action of regulators of antigen presentation or inflammation (SUMO E3 Pias1 and its substrate Prdm1 (negative regulators of MHC-II) and the substrates Syk, Nfkb1, and Tnf), endocytosis/trafficking (predicted E3s Wdfy2, Wdr81 and Wdr91, E3 March6), and negative regulators of CEBPB, the key DC2 TF (Rfwd2 (Cop1) and Det1 of the Cul4-Rfwd2-Det1 complex)[54,79–81].

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Finally, M6 regulators activated mDCs (GP7) and TNF/LPS response (GP8) and repressed phagocytosis and granulation (GP1) and pyruvate metabolism (GP3), through the action of multiple cytokines and positive regulators of NF-kb (LUBAC E3s Rbck1 and Rnf31 (whose substrate Ikbkg is in "opposing module" M2), the substrate Rela; E3s Spop, Rc3h1, Traf6 and Cbl; Tlr4; and E3 complex member Bcl6), positive regulators of cytoskeleton organization and migration (RING-like Phf8[82–84], Ub and SUMO substrate Ptpn1[85]), and E3 Keap1, a negative regulator of redox and stress that targets Nrf2[86–89]. M6 members also included multiple cullin and RING-like ligases (Cul1, Cul3, Cul5, Keap1, Rnf31, Rbck1, Brap, Arih2, Traf6), and help assign putative roles to other E3s, such as Brap, a GWAS gene for psoriasis and carotid atherosclerosis, which activated inflammatory responses in human aortic smooth muscle cells[90,91] but did not have a previously known cellular role in immune cells.

The co-functional modules also impacted global shifts in cell state/type distributions, consistently with the programs they regulate. For example, cells perturbed for M5 regulators were associated with a shift from naive DCs to macrophage-like cells, consistent with the module's role as an activator of the DC2 / MHC-II presentation (GP11) program, while perturbation to M1 and M2 regulators had the opposite effect (**Figure 3B,E** and **S4C,E,F**, FDR < 0.1, Fisher's exact test). The co-functional modules also impacted distributions within one cell state/type. For example, different modules had distinct effects on the cells within DC2.1, DC2.2 and DC2.3 subsets, (**Figure 3C,D** direction 2), orthogonally to the maturation gradient (**Figure 3C** direction 1, **Figure S4D**).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Co-functional modules are enriched for physically interacting E3s**

We next asked if our genetic regulatory network can be aligned and consistent with molecular mechanisms, such as co-complex membership, physical interactions and impact of E3s on TFs that regulate gene expression directly. First, to relate the co-functionality of E3s and complex members to shared molecular mechanisms, we searched for known protein interactions between each pair of regulators (from the STRING database[92], **STAR Methods**) and compared those to their module membership (**Figure S4G**).

There was a significant enrichment of physical interactions between module members for four of the six co-functional modules (M1, 2, 4, and 6), as well as between one pair of modules (M3 and M5) ($P<0.05$, degree preserving permutation test, **STAR Methods**), suggesting that co-functional effects are congruent with joint underlying molecular mechanisms. As expected, physical interactions between members of the same (different) module were generally associated with positive (negative) correlation in functional effects (**Figure 3F** and **S4G**). Such interactions include Gbr2, Ptpn11, and Rack1 (M1), Pparg, Crebbp, Ep300, Ankfy1, and Cul2 (M3); or Cul1, Skp1a, Rnf7, Fbxw1, Rbx1, Cul3, Nedd8, Arih1, and Keap1 (M6). Within components of the NFKB signaling pathway (**Figure S4H**), multiple known TNF activators (Rela, Rbck1, Rnf31; all from M6) both physically interact with TNF and have positively correlated effects, whereas several TLR/NFKB signaling inhibitors (Nfkb1[93], Cyld[94], and Tab1[95]; all in M5) have physical interactions but negatively correlated effects, showing consistency between the genetic model and molecular mechanisms.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The analysis highlighted basic "rules" of co-regulation between E3 cullin, adaptor, and substrate recognition adaptor proteins, where multiple components of each Cullin complex grouped together in the same module (except the common interactor Rbx1), while the Cullin substrate recognition adaptor proteins were in other modules, consistent with their specializing or directing Cullin E3 complexes to different substrates and pathways (**Figure S4I**). For example, the SCF core complex members Cul1, Skp1 and Rbx1 are members of M6, as is Cul3, but Cul1-Skp1-Rbx1 substrate-specific adaptors or Cul3-Rbx1 adaptors are partitioned to multiple other modules (*e.g.*, Cul1-associated adaptors Fbxl14 and Fbxl5 in M2; Fbxl13 and Fbx03 in M3; Fbxw7 in M5; and Fbxo33 and Fbxw11 in M6; Cul3-associated adaptors Klhl3, Klhl24, Klhl6 in M2, M3, and M5, respectively, and Cul5-associated adaptor Socs3 in M5; **Figure S4I**). Similarly, Cul4b and its adaptor Ddb1 are part of M2, but Dda1, which recruits and organizes substrate receptors with both Cul3 and Cul4 complexes, is in M6. Furthermore, each of the interacting pairs of Rbx1 and Arih1 and Cul5 and Arih2 are in M6, consistent with their physical interaction and known functional roles in forming highly specific neddylated CRLs[57], but their expected adaptors were not (e.g., most F-box proteins for Cul1 and Arih1; except Fbxw7 and 11). This highlights the versatility and modularity of the ubiquitin system, whereby different adaptors/receptors target different substrates regulating different aspects of DC lifecycle, and how a systematic Perturb-Seq screen and computational analysis can decipher this organization.

**E3 perturbation effects on gene programs explained by modulation of TF activities**

Next, we examined how the E3 ligases may propagate to the transcriptional level, by combining our genetic model with one associating TFs to their physical targets, to infer TFs whose activity

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(as inferred from the expression of its known targets[96,97]) is impacted by each perturbation (**STAR Methods**).

Overall, the 329 knockout (KO) perturbations were significantly associated with inferred changes in activity of 123 TFs (**Figure S4J**), with 32 TFs with the most prominent effects from 41 E3 ligase and complex members (**Figure 3G**). Most notably, Cul3 and Keap1 perturbations decreased the activity of many TFs, including Nfkb1, Nfkb2, Rela, Relb, Jund, Irf1, Cebpb, Nfya, Klf4, Foxo3, and Foxo4 and increased the (inferred) activity of just four of the 32 highly regulated TFs: Atf6, Wt1, Srf, and Pax6 (**Figure 3G** box 1). Importantly, we show Cul3 and Keap1 perturbations increased (inferred) activity of the well known CUL3-KEAP1 substrate and master regulator of oxidative stress, Nrf2 (**Figure S4J**). The set of TFs whose activity was impacted by Cul3 and Keap1 perturbations was further partitioned to two subsets based on their opposing regulation by different adaptors and receptors. For example, Rack1, Dcaf13 and Vprbp perturbations increased the (inferred) activity of Foxo3/4, Klf4, Twist1, and Smad1 (**Figure 3G**, box 2), while perturbations of M6 E3s, including Spop, Traf6, Fbxw7, Fbxw11, and Cul1 decreased the (inferred) activity of Irf1, Nfkb2, Nfkb1, Rela, Relb, Rel, and Jun (**Figure 3G**, box 3).

This analysis detects established links between E3s and regulated TFs, supporting its validity. For example, Hif1a activity increased following KO of any member of the E3 ligase complex VHL-TCEB1-TCEB2, which binds and ubiquitylates Hif1a for degradation[98] (**Figure S4J**). Cebpb and Jun's inferred activity increased in cells with KO of the E3 ligase Rfwd2 (**Figure 3G**), a member of the CUL4-DDB1-DET1-RFWD2 complex that targets Cebp family TFs and

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Jun for ubiquitination and degradation[80]. Indeed, Rfwd2 represses programs that are activated by Cebpb (GP8) or Jun (GP1 and 8) and are enriched for their bound targets (**Figure S4K**). Because Rfwd2 knockout similarly increases the (inferred) activity of Foxl2 and JunD, these could be additional Rfwd2 substrates.

By relating joint TF targets, E3 targets and E3 regulated programs, we explain E3s impact on different programs through different mediating TFs. For example, by this analysis, Rela mediates Cul3 and Keap1 effects on GP7 and GP8, but not on GP1 (**Figure 3H**), and E3 Fbxw11's impacts on GP7 (but not GP4) (**Figure 3I**). Notably, Cul3, Keap1, Fxbw11 and Rela itself are all members of M6, and the expression of Rela's targets in GP7 is decreased when any of these regulators is perturbed (KO) in our screen (**Table S8**; **Figure 2C,D**). However, because KO of Cul1, Keap1 and Fbxw11 leads to (inferred) reduction in activity of Rela, it is unlikely that Cul3, Keap1 or Fbxw11's effects is through targeting of Rela for degradation (see **Discussion**).

**Perturbed E3 ligases impact multiple statistically-independent pathways**

The effect of perturbing one gene on another gene's expression can be due to various pathways with direct or indirect dependencies and indeed pairs of programs in our regulatory model are dependent, as reflected by their pair-wise positive and negative correlations (**Figure 2C**). To decompose the observed regulatory effects into a set of statistically-independent factors, we performed Independent Components Analysis (ICA)[99] on the regulatory matrix, recovering 15 independent latent factors whose weighted sums optimally explained the perturbation effects (**Figure S5A-E, STAR Methods**), and annotated them by enrichment in functional gene sets and known marker genes (**Table S10**). On average, the factors explained 35% of the variance of

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

observed effects of the 329 impactful regulators (27 explained well (>70%); 72 poorly (<20%), **Figure 4A**, bottom, **STAR METHODS**). The latent factors have little to no correlation (maximal Spearman $\rho = 0.14$) and little overlap in highly loading genes (**Figure S5B,G,** maximum Jaccard similarity index = 0.09), but perturbed genes can affect multiple independent factors (**Figure S5A,H**, maximum Jaccard similarity index = 0.4).

Each factor simultaneously captures both induced and repressed genes along with diametrically opposed regulators. For example, the LPS response factor (#5, **Figure 4B-D**) includes activation of LPS and TNF response genes and repression of genes more highly expressed in DC2s *vs.* mDCs and DC1s (**Figure 4C,D**). It is positively regulated by well-established activators of TNF and the LPS response (*e.g.*, Rnf31, Traf6, Paf1, Ikbkg, Rela, Cebpb) [5], and repressed by known negative regulators (*e.g.*. Rfwd2[54]), as well as by previously-uncharacterized regulators, both positively (E3 adaptor Skp1a) and negatively (E2 Ube2n; E3 substrate adaptor Smu1). The DC immune control factor (#6, **Figure 4E-G**) consists of inflammation and cytokine response genes in two opposing patterns, capturing the maturation gradient from immunostimulatory monocyte derived macrophages and DC2s to immunomodulatory mDCs genes, and its corresponding, diametrically-opposed regulation by E3 Traf2 and E3 adapter Ptpn11 *vs.* March6 and Fbxw7, respectively.

Because one regulator can be associated with multiple ICA factors, this decomposition groups together different subsets of multi-subunit E3 complex members. For example, all four components of the Cul3-Skp1a-Rbx1-Nedd8 complex are associated as negative regulators of the response to oxidative stress factor (#2) (**Figure 4H-J**). Conversely, different combinations of

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

subunits of one complex are associated with different ICA factors, predicting how interactions of one core complex with different substrate recognition adaptor proteins can drive a variety of gene regulation programs (as also described above). For example, in the Rfwd2/Cul4a/Ddb1/Rbx1/Det1 complex, Rfwd2 (Cop1), the E3 substrate recognition adaptor protein that forms an active E3 complex, has outlier loadings indicating regulation of ICA Factors 3, 5, 9, 10, and 11 (**Table S10**). Other members of this complex regulate other factors, not impacted by Rfwd2 perturbation (*e.g.*, Ddb1 regulates Factor 2, **Table S10**). Interestingly, while Rfwd2 and Det1 knockout are both similarly strongly associated with the same factors (#3, 9, and 11), other complex members knockouts (Rbx1, Cul4b) have only weak associations in those same factors (**Figure 4A**, right panel). Thus, Rfwd2 or Det1 may interact with other E3 or cullin complex members, just as the Cul4 complex interacts with other substrate recognition adaptor proteins.

**Intra-module genetic interactions and inter-module additivity in combinatorially-perturbed cells**

In the regulatory network, many of the E3s and other regulators impact the same genes and processes when perturbed individually, but, given the possible non-additivity of biological interactions, determining their effect if perturbed jointly (combinatorial perturbation) requires an additional experiment. In our large screen, 177,871 cells had more than one guide assigned (with 10,244 cells with guides targeting two or more of the 329 singly-impactful regulators; **Figure S6A**), opening the way to test for such genetic interactions. However, because of random sampling from an enormous number of possible combinations, few to no cells were profiled for any specific combination, such that we could not determine the joint effect of any individual

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

combination with confidence. Instead, we reasoned that we can leverage the co-functional modules to group all cells perturbed by a pair of guides from a given pair of modules (including the same module) to gain statistical power to decipher genetic interactions between or within modules rather than between their individual constituent genes. We thus analyzed double perturbations in a setting where cells are assigned to perturbed modules instead of perturbed genes. (The few detected triply perturbed cells were removed prior to analysis, as were module 4 perturbations given the very small number of cells.)

The proportion of genes whose expression has a significant interaction term due to a combinatorial perturbation was much greater in intra- *vs.* inter-module combinations (**Figure 5A**). Thus, when two genes within the same module (intra-module combination) were perturbed in the same cell, the combined effects on the expression of genes were often different than the sum, with a super-linear relationship in M1, M2, and M6 (**Figure 5A,B, STAR METHODS**). Conversely, when two genes from different modules were perturbed in the same cell (inter-module combinations), the impact on most genes was additive (**Figure 5A**).

At the level of expression of the individual affected genes, we found a range of patterns, with almost half (490 of 1,041 tested genes) with at least one significant interaction term (either positive (synergistic) or negative (antagonistic)) as a result of at least one of the 15 inter-module KO pair groups, and 650 with interaction terms in the five intra-module perturbation pairs (FDR < 0.1, **Figure 5C,D** and **S6B,D,E**, **STAR METHODS**). In particular, a joint perturbation of an M3 and M5 regulator yielded non-additive effects in many genes (**Figure 5D**), enriched for biosynthetic, translation, cytokine production, and inflammatory response genes. These included

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

buffering for translation (GP6) and MHC-I presentation (GP10) program genes; synergy for protein homeostasis and phagocytosis (GP5) program genes, and dominance for mDC (GP7) program genes (**Figure 5D**). Notably, the impacts of single perturbations in M3 and M5 regulators are often correlated (**Figure 2C** right and **5D**). Thus, similarly to the joint perturbation of two regulators from the same module, non-additive effects may be more prevalent for regulators from different modules but with similar effects on gene programs when individually perturbed.

**com$\beta$VAE predict combinatorial perturbations within and across modules**

We next asked how well the effects of the double knockouts can be predicted from profiles of single knockout cells. As a baseline, we first used a simple linear model (**STAR METHODS**) to assess the overall effects of each of the six co-functional modules on the 1,041 response genes, and predict the $\log_2$ fold changes of these genes in 20 pairwise module combinations by adding the individual KO group effects. As expected from our analysis above, additive effects explained most of the intra-module interactions quite poorly (**Figure 5B**), while a substantial fraction of the variance in fold changes was explained for some of the inter-module pair combinations (**Figure S6F-H**). Inter-module groups where the additive model performed worse involved pairs of perturbations where more target genes show significant interactions, such as M2-M5 and M3-M5 (**Figure 5A** and **Figure S6C,F-H**), or those with fewer double KO cells, like M1-M4 (**Figure S6A,F-H**), which may have affected the estimation of the ground truth values. For genes with significant interaction terms, the additive model lost its predictive power in most of the module pairs (**Figure S6F-H**), including a change in effect direction for many genes between prediction and observation (**Table S11**).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We hypothesized that we could gain better prediction performance by learning the interaction effects based on the latent structure of the gene expression profiles. To test this hypothesis, we developed com$\beta$VAE, a conditional variational autoencoder (CVAE)[100,101], where the latent variables of the observed data are distributed conditioned on input data labels (**Figure 5E**, **STAR METHODS**). We trained our model with 89,463 single KO cells (of 329 impactful KOs, 80,189 cells for training, 9,274 cells for validation) and a random sample of 70% of control cells, conditioning on 7 groups (controls + 6 regulator modules). With the remaining 30% of control cells, we used the trained model to generate profiles based on the counterfactual questions "*What would be the profile of this control cell if it had a single knockout from module x*?" and "*What would be the profile of this control cell if it had a double knockout, one from module x and another from module y*?" Note that this model only addresses inter-module interactions, as the conditioning is done per KO module (**STAR Methods**). Finally, we calculated the expression fold changes between these generated cells and the population of control cells.

While the explained variance in the expression fold changes for generated KO profiles of single genes from the groups observed during training (single knockout modules) was quite high ($0.77<r^2<0.95$, mean 0.85), estimates for double knockouts varied based on the module pair (**Figure S7A,B**). The profiles of cells with pairs of KOs from M3*M5, the module pair with the highest number of significant inter-module interactions (**Figure S6C**), were estimated far better by com$\beta$VAE ($r^2$=0.23) than by the additive model ($r^2$=0.02) (**Figure S6F, S7A,B**). Moreover, com$\beta$VAE had smaller mean absolute errors than the additive model when predicting the fold changes of genes with significant interaction terms (FDR< 0.1), especially in module pairs with

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the highest number of genes with significant interactions (M3*M5, M3*M6 and M6*M5; **Figure S6G, S7C**). Higher values of Beta, the hyperparameter that changes the weight of the Kullback-Leibler (KL) loss term (which acts as a regularizer on the latent space distribution of the gene expression embeddings as well as the KO embeddings, **STAR Methods**), increased the explained variance in single KOs modules and in double KO module pairs with fewer genes with interaction terms, but reduced it for pairs with more genes with interaction terms (**Figure 5F** and **S6C, S7D**). Thus, greater latent space disentanglement (*i.e.*, larger regularization parameter Beta) leads to better conditioning on the individual single KO groups, and better learning of additive effects at the cost of non-additive ones.

To evaluate how the predictions change when we include some double KO cells during training, we trained our model with the training set of the singly perturbed cells and the double KO cells of either M3M5, M5M6, or both. Interestingly, including double KO cells from one pair of modules increased the prediction performance of *other* unseen double KO groups (**Figure 5G** and **S7E**). Furthermore, when double KO cells were included in the training, higher beta values increased the prediction performance of inter-module groups with more interaction terms (**Figure S7F,G, Table S11**). Thus, including some combinatorial perturbations during training along with better latent space disentanglement helps the model to learn both the generative factors and connections between them, leading to better prediction of unseen combinations.

### *In vitro* perturbation-defined regulators and programs are associated with genetic risk in inflammatory diseases

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To determine if our regulatory network contributes to or is active in human disease, we asked whether impactful regulators in the modules or the programs they regulate are also likely to be causal in human disease, based on either heritability signals, regulation *in vivo* during disease progression, or both. We thus tested the modules and programs for enrichment in common-variant driven disease associations using sc-linker[102] and MAGMA[103] (**STAR Methods**) with GWAS summary statistics from nine immune-related diseases (average N = 79.5K, **Figure 6A,B** and **Table S12,S13**). We also compared each gene program to cell type specific disease progression programs induced in disease *vs*. healthy tissue by scRNA-Seq[102] (**Figure 6C,D**, **STAR Methods**). We found support for both relationships.

Among the co-functional modules, common and low frequency variants in genes in module M1 were enriched for heritability across all traits tested (1.62-fold on average, $P = 1.52x10^{-5}$), and especially immune-related traits (1.94 fold; $P = 2X10^{-4}$), compared to genes constituting all the modules (**Table S12**). In particular, M1 member gene and predicted E3 *WDR36* has a significant MAGMA score in allergy/eczema and blood traits and E3 adapter *PTPN11* in type 1 diabetes (T1D) and blood traits (Z-scored per-trait MAGMA scores, Bonferroni correction $\alpha = 0.1$). Notably, perturbing module M1 activates the mDC (GP7), TNF/LPS response (GP8), and MHC-I Ag presentation (GP10) programs and represses ER stress (GP2), protein homeostasis and phagocytosis (GP5) and translation (GP6) (**Figure 2C,D** and **Table S13**), consistent with the association of variants in this regulating module with inflammatory disease. Additionally, many E3s and E3 complex members in M6 are associated with risk of immune-related traits (**Figure 6A**), including Traf6 (eczema and RA), Rbck1 (CD), Bcl6 (eczema), Keap1 (eczema, IBD, and lupus), Brap (T1D), and Cul1(IBD) (Z-scored per-trait MAGMA scores, Bonferroni MTC a =

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

0.1; **Figure 6A**). Several co-regulated gene programs also showed heritability enrichment in immune diseases. These included the mDC program (GP7) (**Figure 6B**, highest score by both sc-linker and MAGMA), especially *REL* (rank 1 average MAGMA scores across immune diseases), *JAK2* (rank 3) and *STAT5A* (rank 8); and the motility and cel1 maintenance (GP4) program in IBD and related traits (**Figure 6B**), with top driving genes *CCL2* (rank 2) and *CCL7* (rank 9)[104,105].

Concomitantly, several of the perturbation-affected programs were enriched (relative to all genes expressed in this cell type) in disease progression programs from multiple cell types and diseases, especially those of three key cellular processes that are also implicated in modulating immune responses – mitochondrial metabolism[26,106], ER stress[107], and translation/antigen presentation. This is consistent with a model where processes affected by heritable variation in regulators lead to dysregulation of their target programs. For example, translation program genes (GP6) were enriched in disease progression programs of immune and non-immune cells in inflammatory diseases, including ulcerative colitis (UC) and multiple sclerosis (MS) (**Figure 6C,D**); ER stress response genes (GP2) were enriched in disease progression programs in epithelial cells in UC, lung fibrosis and asthma (**Figure 6C**); and disease progression programs in macrophages and DCs in UC are enriched for the translation program (GP6) (**Figure 6D**). These programs are all regulated by module M1, which is itself enriched for heritability of disease risk, as noted above. Thus, our analysis suggests that at least two modules (M1 and M6) may contribute to disease risk through their respective activation or repression of disease induced programs including ER stress (GP2), motility and cell maintenance (GP4), and translation (GP6), including by E3 risk genes Wdr36 (M1) and Keap1, Cul1, and Rbck1 (M6). In addition, the GP4 program is enriched for

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genes with rare-variant driven association in Crohn's disease[108], including *COX4I1*, *POLD4* and *NPY* (ranked 1, 2 and 4; **Figure 6F**). Interaction between *NPY* neurons and immune cells in the enteric nervous system has previously been implicated in IBD pathogenesis, and *NPY* expression changes occur in animal models of IBD[109,110].

Finally, we utilized our model as a "look up" resource for proposing regulators of rare risk variants for Crohn's disease[108]. Our model suggests that Il10ra expression is repressed by Egr2 and that expression of Ccr7, a chemokine receptor regulating many aspects of DC function and guiding DCs to lymph nodes[111], is repressed by Ldb2, Traf2, and Rnf165 (**Figure 6E**). While Traf2 deletion impacts inflammation in both DCs and keratinocytes[112], Traf2-based regulation of Ccr7 has not been previously reported. Increased Ccr7 expression could be an important mechanism in increased T cell infiltration and inflammation controlled by Traf2.

Thus, multiple components of our model, including regulatory modules and their impacted programs are congruent with both risk genes for human immune and inflammatory disease and the dysregulated expression programs observed in patients. This highlights the relevance of our *in vitro* screen to human disease.

## DISCUSSION

In this study, we characterized a large gene family, the E3 ligases, and their interacting partners in the cellular response of primary immune cells to an inflammatory signal. We showed the power of systematic Perturb-Seq to relate different members of one gene family as regulators in

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

distinct co-functional modules, and their impact on individual genes, co-regulated gene programs, and cell state distributions, across a mixed population of related cell types. The modular organization of the regulatory network further allowed us to study and predict the impact of genetic interactions and relate *in vitro* perturbations in a model system to the mechanisms underlying disease risk in humans.

**E3s regulate key phases of the DC life cycle**

The independent factors and programs of target genes regulated by perturbations in E3s and associated proteins span multiple stages in DC life cycle (**Figure 7**), showing the capacity of our screen to capture many phenotypes, and the breadth of roles E3s play in the immune response and the DC life cycle, many of which are novel roles. These include, in order of the lifecycle, regulation of: (**1**) differentiation towards DC2s (*e.g.*, Cul3- Keap1), DC1 state (*e.g.*, Arih2), and mDCs (*e.g.*, Traf2); (**2**) sensing, including the response to LPS (factor #5), ER stress (#8), and oxidative stress (#2), ribonucleotide synthesis (#7), and metabolism and energy (#10), regulated by *e.g.*, the Keap1-Cul3-Rbx1 complex, the Tceb1-Tceb2-Rbx1-Vhl complex, and Traf6; (**3**) DC migration, including chemotaxis (#3,9,11), cytoskeleton organization (#4), and myeloid migration (#13), regulated by *e.g.*, Pparg, Rfwd2, Brap, and the Keap1-Cul3-Rbx1 complex; (**4**) antigen presentation and associated processes (antigen presentation (#12), endopeptidases (#15), and translation (#1)), regulated by *e.g.*, Ambra1, March6, Traf2, and Traf3; and (**5**) production of chemokines that promote either a regulatory or immunostimulatory response (#6.1, #6.2, #14), regulated by *e.g.*, Traf2, Traf3, March6, Pias1, Plgr1, Ptpn11, and Prpf19.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The detailed regulatory model highlights many novel regulatory relations and helps address open questions. For example, it has been unclear whether DC maturation and migration are inextricably linked[113,114]. Our model shows that while the expression of some migration factors (#11; #13) follow a cell maturation gradient, other factors (#3, #4, and #9) express migration genes independent of DC maturation. While the same regulators are shared across migration and maturation in some factors (*e.g.*, Cul3-Keap1 and the CLR1 complex which co-regulate in Factor #4), they regulate them in opposite ways in others (*e.g.*, Cul3-Keap1 and the CLR1 complex in Factor # 9 or #13). This suggests that some, but not all, of the DC migration program is controlled independently of maturation. Several E3s are regulators of multiple programs along the DC lifecycle (*e.g.*, Keap1-Cul3-Rbx1, Fbxw11-Cul1-Skp1a, Fbxw7, March6), while others play specific roles in key stages (*e.g.*, Rfwd2 (Cop1) in migration; Wdr70 in immunostimulatory *vs*. immunoregulatory response).

**Congruent genetic effects and physical interactions relate complex members and characterize E3 partners and substrates**

Screening both E3s and their interacting partners and substrates allowed us to relate functional (genetic) effects to physical interactions and molecular mechanisms. Co-functional modules of regulators are enriched for physical protein-protein interactions and members from the same E3 complex. While multiple components of each Cullin complex grouped together in the same module (except the common interactor Rbx1), the Cullin substrate recognition adaptor proteins were in other modules, highlighting their specializing or directing Cullin E3 complexes to different substrates and pathways. Because programs may not be independent, and regulators are

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

strictly partitioned to separate modules, when regulators have multiple roles, this partitioning can mask their full set of relationships. For example, in the CUL4-DDB1-RBX1-DET1-RFWD2 E3 cullin complex, Cul4b and its adaptor Ddb1 are both members of M2, and Rbx1, Cul1/3/5, and Det1 are in M5. This challenge is addressed when considering independent factors associated with overlapping regulators: Rfwd2 (Cop1) co-regulated the response to LPS with Rbx1 (factor #5), the response to oxidative stress with Ddb1 (#2), and chemotaxis with Det1 (#9). Thus, the ICA factors allow us to relate different adaptors with partly overlapping effects (*e.g.*, Rbx1, Det1, and Ddb1) and the way in which they combine with different substrate recognition adaptor proteins.

Combining co-functional genetic profiles with physical interactions showed congruence between genetic relations and molecular mechanisms, and helped suggest new interactions between E3s and putative adaptors. For example, surprisingly, Rfwd2, but not other members of the CUL4-DDB1-RBX1-DET1-RFWD2 complex, is a regulator of factors 3, 10, and 11, suggesting that it may interact with other complexes to ubiquitylate targets. Other regulators of these factors that have strong physical interactions with Rfwd2 and could be such candidates include: Ptpn11 (Shp2) (factor 3 and 10 regulator) that acts as an adaptor with p38-pRfwd2 to bind and catalyze Ub-mediated degradation of FASN[115]; Wdr82 and Ep300 (factor 10 and 11); Anapc13[116], Cul2, Cul5, and Huwe1 (factor 10); and Crebbp, Skp1a, Nedd8, Cul1, and Wdr5 (factor 11).

Because knockouts of E3s and their substrates should have opposite effects when the substrate is targeted for degradation, and similar effects when Ub modification is activating, we can predict the directionality of protein level regulation by the correlation between expression profiles of E3-

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

substrate pairs in our screen. For example, the regulatory profile of Cebpb is negatively correlated with that of Rfwd2 in all ICA factors where both are regulators (#3,5,11), consistent with the targeting of CEBPB for degradation by the CUL4-DDB1-RBX1-DET1-RFWD2 complex. Conversely, Fbxw11 KO leads to repression of (inferred) activity of all of NFKB1 (p105/p50 precursor), NFKB2 (p100/p52 precursor), Rela (p65) and Relb. Because the targets of these transcriptional activators are repressed both by the TFs own KO and by Cul3, Keap1 and Fxbw11 KO, it is unlikely that this effect is mediated by the TF's degradation. For Cul3 and Keap1, the effect is likely through direct CUL3-KEAP1 ubiquitin modification and subsequent degradation of IKBKB[117]. As for Fxbw11, it is reported to directly bind with NFKB1 and NFKB2[118], and this binding is enhanced in the presence of a proteasome inhibitor[119]. While the current model suggests that full length Nfkb1 is processed constitutively and Fbxw11 (also known as BTrCP2) targets Nfkb2 for complete degradation upon stimulation such as LPS activation[118,120], our data may not be fully consistent with such a model. We hypothesize that Fbxw11 could be part of the system that interacts with the proteasome to processes p105 (Nfkb1) and p100 (Nfkb2) to generate active p50 and p52, respectively. These analyses may be particularly helpful for multi-subunit E3 Ligase complexes reusing core scaffolds and adaptors with different substrate recognition adaptor proteins.

**Scaled genetic perturbation and combinatorial screens**

We leveraged several efficiencies to enable Perturb-seq at scale, including hashing and overloading (40,000 cells per droplet channel; 5-fold cost reduction) and shallow sequencing (15,900 reads per cell on average; 2-fold reduction). Future efficiencies could include pre-barcoding (for even higher overloading[121]), cheaper sequencing[122], and further guide-compressed

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

screens[123],(Yao et al, bioRxiv 2023). For guide compression we note that we initially aimed for a larger number of perturbations per cell, but these have been challenging to achieve in our primary cells, and may require cells from CRISPRi[124] engineered mice.

The large scale of our screen and the modular organization of the regulatory circuit opened a path to systematically tackle genetic interactions. First, our large-scale screen encompassed a relatively large number of cells with multiple perturbations per cell, but as this was a random sample, any particular combination was present in too few cells to directly estimate their effects. However, because of the organization of the regulators in co-functional modules, we could assess genetic interactions at the level of modules, testing for the prevalence of significant inter- or intra-module interactions globally, as well as their impacts on individual genes. This analysis showed that intra-module interactions are far more prevalent than inter-module interactions and impact specific target processes. Consistently, the impact of most inter-module combinations of perturbations can be quite well predicted by a naïve additive (linear) model. Moreover, learning the interaction effects was improved by comβVAE, a conditional variational autoencoder we developed that relies on the latent structure of the expression profiles. This shows the power of combining rich profiles and modular structures to allow prediction of unobserved experiments. Using a higher number of perturbations per cell, implementing 'compressed screens'[123],(Yao et al, bioRxiv 2023) and dedicated gene editing tools, such as Cas12[125], should help further facilitate the dissection of genetic interactions.

**Leveraging cell screens to decipher human genetics**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The reverse genetic approach of Perturb-Seq screens can complement and help interpret the results of forward genetic studies in humans, such as GWAS, especially in systems like primary DCs, where human cell models are lacking. By considering both regulators and regulated programs from our model in the context of associations from human GWAS and single cell profiles from relevant human disease tissues, we found that both common and low frequency variants in regulators in module M1 are enriched for heritability in immune-related traits, including the predicted E3 *WDR36* (in allergy/eczema and blood traits) and E3 adapter *PTPN11* in T1D and blood traits. Moreover, two of the programs affected by perturbations in module M1 regulators – ER stress (GP2) and translation (GP6) are differentially expressed in relevant cell types in immune disease, including macrophages and DCs in UC (GP2) and fibrosis and asthma (GP6).

**Limitations of our study**

Although our study explores genetic circuits regulated by E3 family genes whose direct action is post-transcriptional, we rely on expression profiles as phenotypes, limiting our ability to draw direct mechanistic conclusions. We partly address this by *post hoc* analysis of protein-protein interaction and inference of TF activity, but further mechanistic studies and regulatory models that explicitly include physical interactions will be needed to generate a comprehensive model that is both causal (genetic) and mechanistic. For genetic interactions analysis, we were limited in our ability to transduce BMDCs at high MOI, resulting in fewer cells with multiple perturbations than desired. While module-level analysis allowed us to characterize genetic interactions and test our ability to predict them, this is a simplification. Moreover, our comβVAE currently focuses on inter-module interactions and further developments are needed to predict

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

intra-module interactions, which are more prevalent. Finally, although our goal was to relate E3 family members and complexes in inflammatory circuits to human disease, no human DC line exists and patient-derived material is limited in scale and accessibility for genetic perturbations. We therefore screened mouse primary cells, and then related this signal to human genetics signal to prioritize regulators that may also play a large role in human health and disease.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## STAR Methods

### Mice

Six- to eleven-week-old female, constitutive Cas9-expressing mice were obtained from the Jackson labs (strain # 026179). All animal protocols were reviewed and approved by the MIT Committee on Animal Care (CAC protocol 0618-034-21) and all experiments conformed to the relevant regulatory standards.

### Bone marrow derived dendritic cells

BMDCs were differentiated and perturbed as previously described[16]. Cells were grown in RPMI media (ThermoFisher 21870-076) supplemented with 10% heat inactivated FBS (Invitrogen), 100 U/mL penicillin/streptomycin (GIBCO 15140122), 2 mM L-glutamine (ThermoFisher 25030081), 10 mM HEPES (GIBCO 15630080), 1 mM Na pyruvate (ThermoFisher 11360070), 1X MEM nonessential amino acids (VWR 45000-700), 55 μM β-mercaptoethanol (GIBCO 21985023), and 20 ng/mL recombinant murine GM-CSF (PeproTech 315-03). On day 0, bone marrow was extracted from mouse femur and tibia by cleaning surrounding tissue and crushing the bones gently via mortar and pestle. Bone marrow was filtered with a 70μm cell strainer, and red blood cells were lysed in 2 mL RBC lysis buffer (Sigma R7757) for 10 minutes at room temperature. RBC lysis was quenched with 18 mL media, cells were spun at 1,500 RPM and resuspended in 25 mL media. Following a final 70μm filtration, white blood cells were plated in 1,000mm non tissue culture-treated plastic dishes with 10 mL media at 200,000 cells/mL. On day 2, cells were fed with 10 mL media and infected with lentivirus (described below). On day 5, 12 mL media were removed carefully avoiding nonadherent cells, and 10 mL fresh media were

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

added. On day 7, 5 mL media were added to cultures. On day 8, cells were collected, spun at 1,500 RPM and resuspended in 10 mL fresh media at $1x10^6$ cells/mL. On day 9, cells were stimulated for 3 hours with 100 ng/mL LPS (Invivogen, tlrl-peklps) and harvested by scraping. Cells then underwent antibody staining for cell hashing (described below) and $mKate2^+$ (perturbation vector) and $GFP^+$ (Cas9) cells were enriched by Fluorescence Activated Cell Sorting (FACS) (~8% population) prior to single cell library generation (**Figure S1**).

**Selection of genes for perturbation**

All 898 genes annotated as 'E3 family' were identified from the *Mus musculus* species in the iUUCD 2.0 database[7] on April 2019. The 'E3 family' gene search included members with 'E3 activity', 'E3 adaptor' and 'ULD/UBD' designations in iUUCD. This list was supplemented with 1,054 *Mus musculus* genes identified by an NCBI Gene search of the term 'E3 activity', to a final non-redundant list of 1,137 E3s and interaction partner genes.

**Design and construction of feature barcoding lentiviral vector**

To generate a lentiviral perturbation vector compatible with Perturb-seq Feature Barcoding technology (10x Genomics), a new FB-LentiGuide-Puro-mKate2 vector was designed (pRDA_122: Supplemental Sequence File; AddGene #TBD), where sgRNAs contain a 3' terminal binding sequence (gctcacctattagcggctaagg) complementary to Feature Capture oligos present on 10X v3 and v3.1 beads. To enable FACS or puromycin selection of perturbed cells, expression of mKate2-2A-PuroR was driven by the Ef1α promoter. The designed vector was generated by GenScript.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Cloning of guide pools**

A 3,390 Perturb-Seq guide library was designed with three guides targeting each of the 1,130 genes using the Broad Institute Genetic Perturbation Platform Web sgRNA Designer (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design)[126], along with 330 control guides (165 nontargeting guides and 165 targeting intergenic regions). The pooled CRISPR library was cloned into the FB-LentiGuide-Puro-mKate2 vector backbone as previously described[17].

**Cell hashing and overloading**

BMDCs were washed with PBS and resuspended in 50μL PBS+ 2% BSA and 0.1% Tween20 and stained with BioLegend hashing antibodies (BioLegend 155801, 155803, 155805, 155807, 155809, 155811, 155813, 155815) as previously described[127] and mKate2$^{+}$GFP$^{+}$ cells were sorted by FACS (**Figure S1A**). 40,000 cells were loaded per channel on 46 3' v3 channels according to the manufacturer's instructions[128].

**Single cell RNA-seq library generation**

ScRNA-seq libraries were generated following the manufacturer's instructions (v3 User Guide, 10x User Guide, with Feature Barcoding)[128].

**Feature barcoding (gRNA) and hashtag libraries generation**

The cDNA amplification reaction was mixed by adding 5 μL of 2 μM HTO additive (**Table S14**), 15 μL Feature cDNA primer 2000096, 50 μL Amp Mix (10x Genomics, 2000047 or 2000103) and 30 μL cDNA followed by PCR (98°C for 3 min; 98°C for 15 sec, 63°C for 20 sec,

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

72°C for 60 sec x 11 cycles; 72°C for 1 min; hold at 4°C). cDNA supernatant was selected by collecting the eluant from the 0.6X SPRI cleanup of the cDNA amplification reaction. Eluant was taken from the 0.6X SPRI cleanup of the cDNA amplification reaction, and another 60µL of SPRI beads were added to the 150µL cDNA supernatant. After performing two 80% ethanol washes, elution was performed in 50µL EB buffer (https://www.qiagen.com/us/products/discovery-and-translational-research/lab-essentials/buffers-reagents/buffer-eb/). An additional 1.0X SPRI elution was performed in 30µL EB buffer. This SPRI-purified cDNA supernatant was used as template for both hashtag and feature barcoding library generation.

To generate feature barcode (single cell gRNA) libraries, 50µL 10X Amp mix was next mixed with 45µL Feature SI Primers 1 and 5µL SPRI-purified cDNA supernatant followed by PCR (98°C for 45 sec; 98°C for 20 sec, 60°C for 30 sec, 72°C for 20 sec x 15 cycles; 72°C for 1 min; hold at 4°C). Product was purified with 1.0X SPRI beads, eluting in 30µL EB. Next, the product was run on a 2% TBE agarose gel at 140 volts for 40 minutes and the 250-300 bp gel fragment was purified by adding 800µL agarose dissolving buffer (https://www.zymoresearch.com/products/adb-agarose-dissolving-buffer) to each sample and incubating at 55°C at 1,250 RPM for 10 minutes. The dissolved gel was added to a Zymo DNA Clean & Concentrator-5 kit column and columns were spun for 30 seconds at maximum speed, followed by two 200 µL washes with the included wash buffer (https://www.zymoresearch.com/products/dna-clean-concentrator-5). Columns were spun for four minutes at max speed to dry, then transferred to a new tube followed by elution in 15µL water (D4014). 5µL gel purified product was mixed with 50µL 10X Amp mix, 35µL Feature SI

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Primers 2, and 10µL Chromium i7 sample indices (PN-120262/PN-220103: Chromium i7 Multiplex Kit) followed by PCR (98°C for 45 sec; 98°C for 20 sec, 54°C for 30 sec, 72°C for 20 sec x 6 cycles; 72°C for 1 min; hold at 4°C) and the product was purified with 0.7X SPRI and eluted in 30µL EB.

To generate hashtag libraries, 5µL SPRI-purified cDNA supernatant was next combined with 25µL NEBNext 2X MasterMix (https://www.neb.com/products/m0541-nebnext-high-fidelity-2x-pcr-master-mix#Product%20Information) and mixed with 19µL water, 0.5µL 10µM SI-PCR primer, 0.5µL 10 µM K_HTOX primer (**Table S14**) followed by PCR (98°C for 10 sec; 98°C for 2 sec, 72°C for 15 sec x 21 cycles; 72°C for 1 min; hold at 4°C). Product was purified with 2.0X SPRI and eluted in 15µL TE buffer (https://www.thermofisher.com/order/catalog/product/AM9849).

**Library sequencing**

Gene expression, hashtag, and feature barcoding libraries were pooled and sequenced together across 46 Illumina HiSeq 2500 lanes (R1 28 I1 8 R2 90), yielding on average 15,900 scRNA-Seq reads per cell (R3: 350M reads per lane total of 12 lanes, 78% spike-in), 680 hashing reads per cell (R3: 15M reads, 4% spike-in), and 3,600 feature barcoding reads/cell (R3: 80M reads, 18% spike-in).

**Read alignment and demultiplexing**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ScRNA-seq reads were mapped to the reference mouse genome mm10_v3.0.0 with Cumulus[129] executed through Terra (https://app.terra.bio/), using CellRanger 3.0.2. Demultiplexing of cell-hashing and feature barcoding data was performed with DemuxEM[34] as implemented in Cumulus.

**scRNA-seq preprocessing**

Empty droplets were removed at FDR > 0.01 by EmptyDrops [36] with parameters: lower= 200, niters= 10,000, ignore=10 and retain= 1000. Droplets with either <300 detected genes, <1,000 total UMIs or >15% mitochondrial reads were removed, retaining 1,071,671 droplet profiles. 13,811 genes expressed in at least 400 of these droplets were retained. 838,201 droplets were categorized as singlets based on hashtags by DemuxEM[34]. Cells were assigned a feature (guide) if they had at least 2 feature barcode UMIs, and feature barcode-UMI pairs with <20% of the reads per cell were removed[130], yielding 341,664 cells assigned one barcode and 177,871 cells assigned at least two. 186 targeting guides detected in <20 single perturbed cells were removed from further analysis, retaining 3,204 targeting guides.

**scRNA-seq expression matrix and dimensionality reduction**

Single cell expression matrix and feature barcodes were processed in an anndata object format in Scanpy[131]. Raw counts were saved in the 'counts' layer for downstream analysis. Expression counts per cell were normalized, to a total of $10^4$ counts per cell, and normalized values were log transformed (natural log), after adding a pseudocount of 1.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A *k*-nearest neighbor (*k*=15) cell neighborhood graph was constructed with the first 50 principle components (PCs) of the log normalized expression matrix and clustered with the Leiden algorithm[132] (resolution=0.5). Gene signatures were computed as the average expression of the gene set in the cells minus the average expression of a reference set of genes that is randomly sampled from the same expression bins.

Outlier control guides were identified by PCA of the log-normalized expression matrix of the 44,074 control cells with one of 330 control guides, followed by fitting a linear regression model to each of the top 100 PCs with Python statsmodels package[133], where in each model one PC was the response and the binary feature barcode matrix of the control guides were the covariates. To identify outliers, the 330 X 100 coefficients matrix was fitted with each of four algorithms in scikit-learn Python package[134]: isolation forest[135], elliptic envelope[136], local outlier factor[137], and one-class SVM, and the 9 non-targeting and 22 intergenic guides that were predicted as outliers by at least three methods were removed.

**Prediction of corresponding gene expression clusters in unperturbed cells with and without LPS stimulation**

Cluster assignment of LPS unstimulated unperturbed and LPS stimulated unperturbed cells were predicted by a logistic regression model trained on the LPS stimulated perturbed dataset, with the 10 cluster scores of the top 100 marker genes of the clusters as covariates.

**Guide and knockout enrichment analysis**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To test guide depletion in the screen *vs*. the initial guide pool, distributions of the ratio between the number of cells assigned a (single) guide in the screen *vs*. number of guides in the initial pool were generated for each of the 3,390 targeting and 330 control guides. The ratio distribution of the control guides was taken as background, to calculate an empirical P-value of the depletion of each targeting guide. Targeting guides with ratio of at most 0.08587 were identified as depleted ($CDF_{null}$ (0.08587) = 0.0498).

One-sided Fischer's exact tests were used to test the enrichment (separately, depletion) of cells with a particular guide (or guides targeting the same gene) in each cell subset, where the test schema was the tested group versus rest, and a Benjamini-Hochberg FDR was calculated.

**Identification of congruent guides targeting the same gene**

The effects of each of 3,204 targeting guides on 6,685 genes expressed in at least 5% of the 341,664 singly-perturbed cells were learned using a negative binomial regression model, with control cells as reference, and correcting for the total number of detected genes in a cell, % mitochondrial reads and cell states identified by the initial Leiden clustering. The pairwise Pearson correlation coefficient between the effect size profiles of each pair of guides targeting the same gene were calculated. If no pairs had a positive correlation, all guides were discarded. If all three pairs had $r > 0.015$, all three guides were retained. Otherwise, only the guide pair with the highest positive correlation was retained.

**A linear regulatory model of knockout effect**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A mixed effects negative binomial linear regression model was fit for each of the 6,685 affected (response) genes, where the gene expression values were the response variable, the cell states identified by Leiden clustering were the random effect covariate, and the knockout target gene confounders (number of detected gene/cell, %mitochondrial reads/cell) were the fixed effect covariates, and control cells were the reference. Benjamini-Hochberg FDR was used to correct for multiple hypotheses (6,658 tested genes) with a stringent threshold, such that most regulatory coefficients close to zero were not significant. To generate a background distribution for the number of genes significantly affected due to lentivirus infection or off-target effects, the same model was fit for each of the 299 control guides, testing one control guide against the rest of the control guides. Based on this background distribution, 329 perturbed genes were retained as ones with significant (FDR < 0.1) effect on at least 15 of the 6,685 tested genes.

**Identification of co-functional modules and co-regulated gene programs**

The 329 retained knockouts (perturbed genes) were grouped based on their effects on the 1,041 genes that were affected by at least by 4 perturbations. To this end, the top 50 PCs of the scaled and centered effect size matrix B ($B_{ij}$ = effect size of knockout of gene *i* on gene *j*) were used to calculate a *k*-nearest neighbors graph (*k*=10) of the knockout (perturbed) genes, and the Leiden algorithm (resolution=0.64) was used to identify the 6 clusters as co-functional modules.

Similarly, to identify co-regulated gene programs, the top 50 PCs of the scaled and centered $B^T$ were used to construct a *k*-NN graph of the 1,041 response genes (*k*=10), and the Leiden algorithm (resolution=0.8) was used to identify 8 gene programs. Three of these programs were selected for further subclustering upon manual inspection, resulting in 11 gene programs.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Embedding cells jointly on KO module information and their gene expression profiles**

To assess the change in cell distributions across DC2.1, DC2.1 and DC2.3 subsets upon perturbations, supervised-UMAP [138] was used (target_weight = 0.5, kNN n_neighbors=18) to embed cells based on their normalized expression profiles and module assignment (M1-M6).

**Calculating Wasserstein distances within and across modules**

The average population distances between cell subsets perturbed for members of each module (or controls) or between randomly sampled cell subsets perturbed for members of the same modules, were calculated by sampling without replacement 300 cells (100 times), computing Wasserstein distances between pairs of cell populations using the Python Optimal Transport Library [139], and averaging across 100 iterations.

**Protein-protein interaction analysis**

*Mus musculus* protein-protein interaction network data was downloaded from STRING DB (version 11.5) and interactions with experimental evidence score > 0 were selected. Interactions between the 329 knockout targets were used to generate a protein-protein interaction graph. To test for enrichment of intra- and inter-module interactions, 400 random degree-preserving graphs were generated using the BiRewire R package[140] and the distribution of number of intra- and inter-module interactions in these graphs was used as the null distribution to calculate empirical P-values for the corresponding observed number of interactions.

**Inference of KO effects on TF factor activity**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Expression scores of high confidence targets (Levels A and B) activated by 123 mouse TFs in DoRothEA[97,141,142], were calculated using the R package decoupleR[96]. A linear regression model was used to infer the effects of each 329 KOs on the expression of targets of each of the 123 TFs, where in each model the response variable was the expression score of each TF's target genes, and the covariates were the 1-hot encoded feature barcode matrix (with control cells as the base level) and possible confounders (cell clusters, %mitochondrial reads, number UMIs).

**ICA module factorization**

Independent components analysis[99,143] was used to identify statistically independent factors from the 1041 X 329 effect size matrix $\boldsymbol{\beta}$ from the mixed effect linear model ($b_{ij}$ = estimated effect of knockout of gene *j* on expression of gene *i*). A source of variation $s_k = \{s_{1k}, s_{2k}, s_{3k}, \ldots, s_{1041k}\}$ is defined as the set of relative weights (*i.e.*, relative expression states) of genes 1, … , 1,041, such that the effects of perturbation *j* on expression of gene *i* is a weighted sum of the effects over *P* different sources, written as:

$$b_{ij} = a_{1j}s_{i1} + a_{2j}s_{i2} + \cdots + a_{pj}s_{ip}$$

where $a_{1j}, a_{2j}, \ldots, a_{pj}$ are the mixing weights and $a_{pj}$ is the overall effect of perturbation *j* on source *p*. Thus, in matrix $\boldsymbol{\beta}$, each row is an observation of a gene's expression changes due to the varying effects of the knockouts on various pathways (sources of variation) to which the gene belongs.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Identifying the $P$ underlying pathways $s_1,\ldots,s_P$ affected by 329 knockouts is formulated as finding a source matrix $S$ (1,041 by $P$) and mixing matrix $M$ ($P$ by 329) which are both unknown:

$$B = SM$$

ICA relaxes this factorization problem by assuming that the source signals are independent and requiring that they be non-Gaussian [99]. Although modeling total perturbation effects as a linear combination of factors may miss nonlinear relationships, the nonlinear separation problem is not identifiable.

ICA decomposition was computed using Information-Maximization (Infomax)[144], as implemented in the ICA package in R[145]. The optimal number of latent sources, $P$ was determined considering (**1**) the number of non-Gaussian components estimated by the Ladle estimator[146,147]; (**2**) reconstruction error of the original matrix from the obtained statistically independent components; and (**3**) prediction power of the identified factors for the effects of unseen perturbations during fitting. For (**2**), the ICtest R package[146] was used to compute the ladle estimates (gn) for different number of factors, where 'gn' is the sum of the vectors giving the measures of variation of the eigenvectors and the normalized eigenvalues of the fourth order blind identification (FOBI) matrix and the estimated number of Gon-gaussian components is the value where gn takes its minimum. For (**3**), for $P$ between 2 and 30, 80% (263) of the 329 perturbations were randomly sampled 10 times, ICA was fitted each time to the 1,041 by 263 matrix and the effects of each of the remaining 66 perturbations was predicted with a simple linear regression model, where the 1,041 by P matrix $S$ was the covariate matrix.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

After selecting *P*=15, the full 1,041 X 329 effect size matrix $\boldsymbol{\beta}$ was decomposed and for each factor $s_k$ in *S*, gene *i* was defined as a prominent gene defining this factor if it had outlier weights $s_{ik} < Q_1(s_k) - 1.5\big(Q_3(s_k) - Q_1(s_k)\big) \quad | \quad s_{ik} > Q_3(s_k) + 1.5\big(Q_3(s_k) - Q_1(s_k)\big)$. Likewise, a KO of gene *j* was defined as highly affecting factor *k* if it had outlier weights in component $m_k$ of the mixing matrix M, $m_{jk} < Q_1(m_k) - \big(Q_3(m_k) - Q_1(m_k)\big) \quad | \quad m_{jk} > Q_3(m_k) + \big(Q_3(m_k) - Q_1(m_k)\big)$.

**Genetic interaction analysis**

For inter-module interactions, for each of the 1,041 response genes, linear regression models were fit as follows:

$$Y_i = \beta_0 + \sum_{j=1}^{6} \beta_j M_j \; + \sum_{j=1}^{6} \sum_{t=(j+1)}^{6} \beta_{jt} M_j M_t + \; \epsilon$$

where $Y_i$ is the normalized expression level of gene *i* (corrected for cell cluster, % mitochondrial reads and number UMIs), $M_i$ is a binary covariate denoting if the cell had a perturbation in a gene from module *i*. The model was fit with single KO cells and inter-module double KO cells. P-values of the $\beta$ estimates were corrected with Benjamini-Hochberg FDR for the 1,041 tested genes.

For intra-module interactions, for each KO module we randomly partitoned (for 50 times) the KO module $M_j$ into two equal bins in terms of the number of KO genes, $M_{j_1}$ and $M_{j_2}$, and for each response gene *i* fit the model:

$$Y_i = \beta_0 + \beta_1 M_{j_1} + \beta_2 M_{j_2} + \beta_3 M_{j_1} M_{j_2} + \; \epsilon$$

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where $Y_i$ is the normalized expression level of gene $i$ corrected for confounders as above. Benjamini-Hochberg FDR was first calculated for the p-values of the $\beta$ estimates for each of the 50 iterations, $\beta$ estimates for which FDR >= 0.1 were set to zero, and then the mean of the 50 parameter estimates as taken as the inferred intra-module interaction effect.

**Deep learning model to predict interactions**

To learn models that predict the effect of combinatorial perturbations, conditional VAEs (CVAEs) [101] were used, which model the distribution of a high-dimensional output as a generative model conditioned on the auxiliary covariates. In general, CVAEs aim to learn the marginal likelihood of the data in such a generative process:

$$\max_{\phi,\,\theta} \mathbb{E}_{q_\phi(z|x,a)}[\log p_\theta(x|a,z)]$$

where $x \in \mathbb{R}^D$ is a dataset of samples with labels (conditioned variable) $a$ and generated by ground truth factors $z$, while $\phi, \theta$ parametrize the distributions of the CVAE encoder and the decoder respectively. This can be rewritten as:

$$\log p_\theta(x|a,z) = D_{KL}(q(z|x,a) \parallel p(z|a)) + \mathcal{L}(\phi,\theta;x,a,z)$$

where $D_{KL}(\parallel)$ is the non-negative Kullback-Leibler (KL) divergence between the true and the approximate posterior and $\mathcal{L}(\phi,\theta;x,a,z)$ is the evidence lower bound (ELBO) on the log-likelihood of the data:

$$\log p_\theta(x|a,z) \geq \mathcal{L}(\phi,\theta;x,a,z) = \mathbb{E}_{q_\phi(z|x,a)}[\log p_\theta(x|a,z)] - D_{KL}(q(z|x,a) \parallel p(z|a))$$

To make optimization tractable in practice, $p(z|a)$ is typically set to the isotropic unit Gaussian $\mathcal{N}(0,I)$[148]. The ELBO for the VAE and CVAE factorizes across the samples[101,148,149]. Therefore, it is straightforward to apply computationally efficient minibatch based stochastic gradient

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

descent (SGD) and learn the parameters $\phi$ of the encoder ($q_\phi(z|x,a)$) and $\theta$ of the decoder ($p_\theta(x|a,z)$) by deep neural networks[148].

In our com$\beta$VAE model, ($x_i$,$a_i$) denotes a single data point *i*, where $x \in \mathbb{R}^G$ is the G-dimensional observed log-normalized gene expression profiles of *G* genes in a single cell *i* and $a_i \in A = A_{i1} \times A_{i2} \times \ldots A_{iP}$ representing the *P*-dimensional auxiliary (independent) discrete covariates of the same cell, such as the knockout(s) perturbing the cell, or confounders (cell subtype or cell cycle phase). The perturbation covariates in $A$ are assumed to be independent binary covariates and a cell can have multiple perturbations, while other covariates are one-hot encoded. The latent variables $z_i$ are genenrated conditionally on a *D* dimensional vector $v \in \mathbb{R}^D$ which are the embeddings of $a_i$ learnt with a single hidden layer neural network of D units which is jointly trained with the encoder-decoder framework. An adjustable hyperparameter β is introduced to the original CVAE objective, which was previously shown to result in more disentangled latent representations z in standard VAE models [150,151] :

$$\mathcal{L}(\phi, \theta; x, v, z) = \mathbb{E}_{q_\phi(z|x,v)}[\log p_\theta(x|v,z)] - \beta D_{KL}(q(z|x,v) \parallel p(z|v))$$

Assuming the data generating process described above, our objective is to train a model such that, a target counterfactual distribution $x_i'$ of gene expression $x_i$ can be generated if cell *i* had the covariates $a_i'$ instead of $a_i$. For the network architecture after benchmarking the number of hidden units [2,3,4,5,6], and number of units per layer [32, 64, 128, 512, 1024], the encoder and decoder networks were defined with 2 hidden layers, with 512 units in each layer, and ReLU (rectified linear unit) activation function used between the hidden layers. The dimensions

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[16,32,64,128] were benchmarked for the dimensions of $z$ and $v$, and both were set to 64. Models were trained and benchmarked with log-transformed normalized gene expression values of 1,041 genes corrected for cell clusters, % mitochondrial reads and total UMI counts to minimize the effect of confounders in evaluating generated *vs*. observed effects. The model was implemented in Pytorch, and trained with hyperparameters batch_size =1000, max_epochs=8000, optimizer=adam, learning_rate= 0.001, weight_decay=0.

**Disease progression gene programs**

Disease progression programs were defined as previously described[102] using publicly available scRNA-seq datasets, processed, annotated, and analyzed as previously described[102]. A gene-level non-parametric Wilcoxon rank sum differential expression test was performed between cells from healthy and disease tissues of the same cell type as previously described[102].

**Identification of heritability signal**

Gene programs and co-functional modules were tested for enrichment in heritability signal using both scLinker[102] and MAGMA[103] over a set of 60 relatively independent diseases and traits (average N = 297K)[102].

In sclinker, each program or module were combined with enhancer-gene linking strategies defined by either SNPs in enhancers linked to genes based on the Roadmap[152] and Activity-By-Contact (ABC) SNP-To-Gene (S2G) strategies either aggregated across all biosamples related to blood (RoadmapUABC-Blood). For each gene score X and S2G strategy Y, a combined annotation X × Y was defined by assigning to each SNP the maximum gene score among genes

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

linked to that SNP (or 0 for SNPs with no linked genes); this generalizes the standard approach of constructing annotations from gene scores using window-based strategies[153,154]. Heritability analysis of these sclinker annotations was performed using stratified LD score regression[155,156] conditional on a set of 86 baseline coding, conserved and LD-related annotations (baseline-LDv2.1[157]). The Enrichment Score (Escore) metric[102] reported was derived from heritability enrichment analysis and its corresponding p-values.

For MAGMA analysis, the MAGMA z-score was computed for each gene module or program using a 0kb window based strategy for linking SNPs to genes[103] and then a gene set enrichment analysis of the MAGMA z-scores was performed for each with respect to a set of 1,000 sets of same size of randomly selected genes from across all perturbation programs (using the fgsea software[158]).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## References and Notes


1. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
2. Bock, C. *et al.* High-content CRISPR screening. *Nat. Rev. Methods Primer* **2**, 1–23 (2022).
3. Uffelmann, E. *et al.* Genome-wide association studies. *Nat. Rev. Methods Primer* **1**, 1–21 (2021).
4. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).
5. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675–686 (2015).
6. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* **175**, 1958-1971.e15 (2018).
7. Zhou, J. *et al.* iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations. *Nucleic Acids Res.* **46**, D447–D453 (2018).
8. Park, Y., Jin, H., Aki, D., Lee, J. & Liu, Y.-C. The ubiquitin system in immune regulation. *Adv. Immunol.* **124**, 17–66 (2014).
9. Kattah, M. G., Malynn, B. A. & Ma, A. Ubiquitin-Modifying Enzymes and Regulation of the Inflammasome. *J. Mol. Biol.* **429**, 3471–3485 (2017).
10. Senft, D., Qi, J. & Ronai, Z. A. Ubiquitin ligases in oncogenic transformation and cancer therapy. *Nat. Rev. Cancer* **18**, 69–88 (2018).
11. Greb, J. E. *et al.* Psoriasis. *Nat. Rev. Dis. Primer* **2**, 1–17 (2016).
12. Lee, M. N. *et al.* Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* **343**, 1246980 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

13. Oikawa, D. *et al.* Molecular bases for HOIPINs-mediated inhibition of LUBAC and innate immune responses. *Commun. Biol.* **3**, 1–17 (2020).

14. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882.e21 (2016).

15. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).

16. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).

17. Frangieh, C. J. *et al.* Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. *Nat. Genet.* **53**, 332–341 (2021).

18. Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896.e15 (2016).

19. Jin, X. *et al.* In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* **370**, eaaz6063 (2020).

20. McFaline-Figueroa, J. L. *et al.* A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat. Genet.* **51**, 1389–1398 (2019).

21. Norman, T. M. *et al.* Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science* **365**, 786–793 (2019).

22. Papalexi, E. *et al.* Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat. Genet.* **53**, 322–331 (2021).

23. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* (2022) doi:10.1016/j.cell.2022.05.013.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

24. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).

25. Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257–263 (2009).

26. Jovanovic, M. *et al.* Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038 (2015).

27. Rabani, M. *et al.* High resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698–1710 (2014).

28. Rabani, M. *et al.* Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* **29**, 436–442 (2011).

29. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).

30. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).

31. Platt, R. J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).

32. Chevrier, N. *et al.* Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell* **147**, 853–867 (2011).

33. Garber, M. *et al.* A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* **47**, 810–822 (2012).

34. Gaublomme, J. T. *et al.* Nuclei multiplexing with barcoded antibodies for single-nucleus genomics. *Nat. Commun.* **10**, 2907 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

35. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

36. Lun, A. T. L. *et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 63 (2019).

37. El-Brolosy, M. A. *et al.* Genetic compensation triggered by mutant mRNA degradation. *Nature* **568**, 193–197 (2019).

38. Helft, J. *et al.* GM-CSF Mouse Bone Marrow Cultures Comprise a Heterogeneous Population of CD11c(+)MHCII(+) Macrophages and Dendritic Cells. *Immunity* **42**, 1197–1211 (2015).

39. Maier, B. *et al.* A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature* **580**, 257–262 (2020).

40. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).

41. Jiang, A. *et al.* Disruption of E-cadherin-mediated adhesion induces a functionally distinct pathway of dendritic cell maturation. *Immunity* **27**, 610–624 (2007).

42. Erlich, Z. *et al.* Macrophages, rather than DCs, are responsible for inflammasome activity in the GM-CSF BMDC model. *Nat. Immunol.* **20**, 397–406 (2019).

43. Chinnam, M. *et al.* MDM2 E3 ligase activity is essential for p53 regulation and cell cycle integrity. *PLoS Genet.* **18**, e1010171 (2022).

44. Bollag, G. *et al.* Loss of NF1 results in activation of the Ras signaling pathway and leads to aberrant growth in haematopoietic cells. *Nat. Genet.* **12**, 144–148 (1996).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

45. Dasgupta, B. & Gutmann, D. H. Neurofibromin regulates neural stem cell proliferation, survival, and astroglial differentiation in vitro and in vivo. *J. Neurosci. Off. J. Soc. Neurosci.* **25**, 5584–5594 (2005).

46. Sharma, M. D. *et al.* Activation of p53 in Immature Myeloid Precursor Cells Controls Differentiation into Ly6c+CD103+ Monocytic Antigen-Presenting Cells in Tumors. *Immunity* **48**, 91-106.e6 (2018).

47. Saha, S., Buttari, B., Panieri, E., Profumo, E. & Saso, L. An Overview of Nrf2 Signaling Pathway and Its Role in Inflammation. *Mol. Basel Switz.* **25**, E5474 (2020).

48. Zhang, D. D., Lo, S.-C., Cross, J. V., Templeton, D. J. & Hannink, M. Keap1 Is a Redox-Regulated Substrate Adaptor Protein for a Cul3-Dependent Ubiquitin Ligase Complex. *Mol. Cell. Biol.* **24**, 10941–10953 (2004).

49. Cai, W. & Yang, H. The structure and regulation of Cullin 2 based E3 ubiquitin ligases and their biological functions. *Cell Div.* **11**, 7 (2016).

50. Pant, V., Xiong, S., Iwakuma, T., Quintás-Cardama, A. & Lozano, G. Heterodimerization of Mdm2 and Mdm4 is critical for regulating p53 activity during embryogenesis but dispensable for p53 and Mdm2 stability. *Proc. Natl. Acad. Sci.* **108**, 11995–12000 (2011).

51. Williams, M. A. *et al.* Disruption of the transcription factor Nrf2 promotes pro-oxidative dendritic cells that stimulate Th2-like immunoresponsiveness upon activation by ambient particulate matter. *J. Immunol. Baltim. Md 1950* **181**, 4545–4559 (2008).

52. Qin, C. *et al.* RACK1 T50 Phosphorylation by AMPK Potentiates Its Binding with IRF3/7 and Inhibition of Type 1 IFN Production. *J. Immunol. Baltim. Md 1950* **207**, 1411–1418 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

53. Martín-Fontecha, A. *et al.* Regulation of Dendritic Cell Migration to the Draining Lymph Node□: Impact on T Lymphocyte Traffic and Priming. *J. Exp. Med.* **198**, 615–621 (2003).

54. Ndoja, A. *et al.* Ubiquitin Ligase COP1 Suppresses Neuroinflammation by Degrading c/EBPβ in Microglia. *Cell* **182**, 1156-1169.e12 (2020).

55. Fütterer, A. *et al.* DIDO as a Switchboard that Regulates Self-Renewal and Differentiation in Embryonic Stem Cells. *Stem Cell Rep.* **8**, 1062–1075 (2017).

56. Liu, Y. *et al.* The death-inducer obliterator 1 (Dido1) gene regulates embryonic stem cell self-renewal. *J. Biol. Chem.* **289**, 4778–4786 (2014).

57. Kostrhon, S. *et al.* CUL5-ARIH2 E3-E3 ubiquitin ligase structure reveals cullin-specific NEDD8 activation. *Nat. Chem. Biol.* **17**, 1075–1083 (2021).

58. Lin, A. E. *et al.* ARIH2 is essential for embryogenesis, and its hematopoietic deficiency causes lethal activation of the immune system. *Nat. Immunol.* **14**, 27–33 (2013).

59. Szczepaniak, A. *et al.* Germline missense NF1 mutation in an elderly patient with a blastic plasmacytoid dendritic cell neoplasm. *Int. J. Hematol.* **110**, 102–106 (2019).

60. Sukhbaatar, N., Hengstschläger, M. & Weichhart, T. mTOR-Mediated Regulation of Dendritic Cell Differentiation and Function. *Trends Immunol.* **37**, 778–789 (2016).

61. Nagler, A. *et al.* A genome-wide CRISPR screen identifies FBXO42 involvement in resistance toward MEK inhibition in NRAS-mutant melanoma. *Pigment Cell Melanoma Res.* **33**, 334–344 (2020).

62. Song, Y. *et al.* E3 ligase FBXW7 is critical for RIG-I stabilization during antiviral responses. *Nat. Commun.* **8**, 14654 (2017).

63. Hou, Y., Moreau, F. & Chadee, K. PPARγ is an E3 ligase that induces the degradation of NFκB/p65. *Nat. Commun.* **3**, 1300 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

64. Han, L. *et al.* PPARG-mediated ferroptosis in dendritic cells limits antitumor immunity. *Biochem. Biophys. Res. Commun.* **576**, 33–39 (2021).

65. Nobs, S. P. *et al.* PPARγ in dendritic cells and T cells drives pathogenic type-2 effector responses in lung inflammation. *J. Exp. Med.* **214**, 3015–3035 (2017).

66. Gallouet, A.-S. *et al.* Macrophage production and activation are dependent on TRIM33. *Oncotarget* **8**, 5111–5122 (2017).

67. Xue, J. *et al.* Tumour suppressor TRIM33 targets nuclear β-catenin degradation. *Nat. Commun.* **6**, 6156 (2015).

68. Lu, A. *et al.* Aim2 Couples With Ube2i for Sumoylation-Mediated Repression of Interferon Signatures in Systemic Lupus Erythematosus. *Arthritis Rheumatol. Hoboken NJ* **73**, 1467–1477 (2021).

69. Corsini, E. *et al.* Immunostimulatory effects of RACK1 pseudosubstrate in human leukocytes obtained from young and old donors. *Oncotarget* **6**, 6524–6534 (2015).

70. Nurieva, R. I. *et al.* The E3 ubiquitin ligase GRAIL regulates T cell tolerance and regulatory T cell function by mediating T cell receptor-CD3 degradation. *Immunity* **32**, 670–680 (2010).

71. Song, G. *et al.* E3 ubiquitin ligase RNF128 promotes innate antiviral immunity through K63-linked ubiquitination of TBK1. *Nat. Immunol.* **17**, 1342–1351 (2016).

72. Su, L. L., Iwai, H., Lin, J. T. & Fathman, C. G. The transmembrane E3 ligase GRAIL ubiquitinates and degrades CD83 on CD4 T cells. *J. Immunol. Baltim. Md 1950* **183**, 438–444 (2009).

73. McCormick, S. M. & Heller, N. M. Regulation of Macrophage, Dendritic Cell, and Microglial Phenotype and Function by the SOCS Proteins. *Front. Immunol.* **6**, 549 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

74. Lalani, A. I. *et al.* Myeloid cell TRAF3 regulates immune responses and inhibits inflammation and tumor development in mice. *J. Immunol. Baltim. Md 1950* **194**, 334–348 (2015).

75. Murakami, T. *et al.* Cutting Edge: G Protein Subunit β 1 Negatively Regulates NLRP3 Inflammasome Activation. *J. Immunol. Baltim. Md 1950* **202**, 1942–1947 (2019).

76. Ricote, M. & Glass, C. K. PPARs and molecular mechanisms of transrepression. *Biochim. Biophys. Acta* **1771**, 926–935 (2007).

77. Zhu, K. *et al.* WDR82 Negatively Regulates Cellular Antiviral Response by Mediating TRAF3 Polyubiquitination in Multiple Cell Lines. *J. Immunol. Baltim. Md 1950* **195**, 5358–5366 (2015).

78. Lefterova, M. I. *et al.* PPARgamma and C/EBP factors orchestrate adipocyte biology via adjacent binding on a genome-wide scale. *Genes Dev.* **22**, 2941–2952 (2008).

79. Lau, O. S. & Deng, X. W. The photomorphogenic repressors COP1 and DET1: 20 years later. *Trends Plant Sci.* **17**, 584–593 (2012).

80. Wertz, I. E. *et al.* Human De-etiolated-1 regulates c-Jun by assembling a CUL4A ubiquitin ligase. *Science* **303**, 1371–1374 (2004).

81. Yoshida, A., Kato, J.-Y., Nakamae, I. & Yoneda-Kato, N. COP1 targets C/EBPα for degradation and induces acute myeloid leukemia via Trib1. *Blood* **122**, 1750–1760 (2013).

82. Asensio-Juan, E., Gallego, C. & Martínez-Balbás, M. A. The histone demethylase PHF8 is essential for cytoskeleton dynamics. *Nucleic Acids Res.* **40**, 9429–9440 (2012).

83. Gu, L. *et al.* The Histone Demethylase PHF8 Is Essential for Endothelial Cell Migration. *PloS One* **11**, e0146645 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

84. Zhou, W. *et al.* PHF8 upregulation contributes to autophagic degradation of E-cadherin, epithelial-mesenchymal transition and metastasis in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res. CR* **37**, 215 (2018).

85. Martin-Granados, C. *et al.* A key role for PTP1B in dendritic cell maturation, migration, and T cell activation. *J. Mol. Cell Biol.* **7**, 517–528 (2015).

86. Bellezza, I., Giambanco, I., Minelli, A. & Donato, R. Nrf2-Keap1 signaling in oxidative and reductive stress. *Biochim. Biophys. Acta Mol. Cell Res.* **1865**, 721–733 (2018).

87. Hammer, A. *et al.* Role of Nuclear Factor (Erythroid-Derived 2)-Like 2 Signaling for Effects of Fumaric Acid Esters on Dendritic Cells. *Front. Immunol.* **8**, 1922 (2017).

88. Rojo de la Vega, M., Chapman, E. & Zhang, D. D. NRF2 and the Hallmarks of Cancer. *Cancer Cell* **34**, 21–43 (2018).

89. Sihvola, V. & Levonen, A.-L. Keap1 as the redox sensor of the antioxidant response. *Arch. Biochem. Biophys.* **617**, 94–100 (2017).

90. Lc, T. *et al.* Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nat. Commun.* **8**, (2017).

91. Liao, Y.-C. *et al.* BRAP Activates Inflammatory Cascades and Increases the Risk for Carotid Atherosclerosis. *Mol. Med. Camb. Mass* **17**, 1065–1074 (2011).

92. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

93. Yang, H.-T. *et al.* NF-κB1 inhibits TLR-induced IFN-β production in macrophages through TPL-2-dependent ERK activation. *J. Immunol. Baltim. Md 1950* **186**, 1989–1996 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

94. Yoshida, H., Jono, H., Kai, H. & Li, J.-D. The tumor suppressor cylindromatosis (CYLD) acts as a negative regulator for toll-like receptor 2 signaling via negative cross-talk with TRAF6 AND TRAF7. *J. Biol. Chem.* **280**, 41111–41121 (2005).

95. Yang, Y. *et al.* The E3 ubiquitin ligase RNF114 and TAB1 degradation are required for maternal-to-zygotic transition. *EMBO Rep.* **18**, 205–216 (2017).

96. Badia-i-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinforma. Adv.* **2**, vbac016 (2022).

97. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).

98. Haase, V. H. The VHL tumor suppressor: master regulator of HIF. *Curr. Pharm. Des.* **15**, 3895–3903 (2009).

99. Hyvärinen, A. & Oja, E. Independent component analysis: algorithms and applications. *Neural Netw. Off. J. Int. Neural Netw. Soc.* **13**, 411–430 (2000).

100. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).

101. Sohn, K., Lee, H. & Yan, X. Learning Structured Output Representation using Deep Conditional Generative Models. in *Advances in Neural Information Processing Systems* vol. 28 (Curran Associates, Inc., 2015).

102. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat. Genet.* **54**, 1479–1492 (2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

103. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

104. Carson, W. F. *et al.* Enhancement of macrophage inflammatory responses by CCL2 is correlated with increased miR-9 expression and downregulation of the ERK1/2 phosphatase Dusp6. *Cell. Immunol.* **314**, 63–72 (2017).

105. Morris, A. B., Adams, L. E. & Ford, M. L. Influence of T Cell Coinhibitory Molecules on CD8+ Recall Responses. *Front. Immunol.* **9**, (2018).

106. Pearce, E. J. & Everts, B. Dendritic cell metabolism. *Nat. Rev. Immunol.* **15**, 18–29 (2015).

107. Cubillos-Ruiz, J. R. *et al.* ER Stress Sensor XBP1 Controls Anti-tumor Immunity by Disrupting Dendritic Cell Homeostasis. *Cell* **161**, 1527–1538 (2015).

108. Sazonovs, A. *et al.* Large-scale sequencing identifies multiple genes and rare variants associated with Crohn's disease susceptibility. *Nat. Genet.* **54**, 1275–1283 (2022).

109. El-Salhy, M. & Hausken, T. The role of the neuropeptide Y (NPY) family in the pathophysiology of inflammatory bowel disease (IBD). *Neuropeptides* **55**, 137–144 (2016).

110. Chandrasekharan, B., Nezami, B. G. & Srinivasan, S. Emerging neuropeptide targets in inflammation: NPY and VIP. *Am. J. Physiol. Gastrointest. Liver Physiol.* **304**, G949-957 (2013).

111. Rodríguez-Fernández, J. L. & Criado-García, O. The Chemokine Receptor CCR7 Uses Distinct Signaling Modules With Biased Functionality to Regulate Dendritic Cells. *Front. Immunol.* **11**, 528 (2020).

112. Etemadi, N. *et al.* TRAF2 regulates TNF and NF-κB signalling to suppress apoptosis and skin inflammation independently of Sphingosine kinase 1. *eLife* **4**, e10592 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

113. Flores-Romo, L. In vivo maturation and migration of dendritic cells. *Immunology* **102**, 255–262 (2001).

114. Liu, J., Zhang, X., Cheng, Y. & Cao, X. Dendritic cell migration in inflammation and immunity. *Cell. Mol. Immunol.* **18**, 2461–2471 (2021).

115. Muramatsu, H. *et al.* Mutations of an E3 ubiquitin ligase c-Cbl but not TET2 mutations are pathogenic in juvenile myelomonocytic leukemia. *Blood* **115**, 1969–1975 (2010).

116. Schwickart, M. *et al.* Swm1/Apc13 is an evolutionarily conserved subunit of the anaphase-promoting complex stabilizing the association of Cdc16 and Cdc27. *Mol. Cell. Biol.* **24**, 3562–3576 (2004).

117. Lee, D.-F. *et al.* KEAP1 E3 Ligase-Mediated Downregulation of NF-κB Signaling by Targeting IKKβ. *Mol. Cell* **36**, 131–140 (2009).

118. Heissmeyer, V., Krappmann, D., Hatada, E. N. & Scheidereit, C. Shared pathways of IkappaB kinase-induced SCF(betaTrCP)-mediated ubiquitination and degradation for the NF-kappaB precursor p105 and IkappaBalpha. *Mol. Cell. Biol.* **21**, 1024–1035 (2001).

119. Kim, T. Y. *et al.* Substrate Trapping Proteomics Reveals Targets of the βTrCP2/FBXW11 Ubiquitin Ligase. *Mol. Cell. Biol.* **35**, 167–181 (2015).

120. Zhang, Q., Lenardo, M. J. & Baltimore, D. 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. *Cell* **168**, 37–57 (2017).

121. Datlinger, P. *et al.* Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. *Nat. Methods* **18**, 635–642 (2021).

122. Simmons, S. K. *et al.* Mostly natural sequencing-by-synthesis for scRNA-seq using Ultima sequencing. *Nat. Biotechnol.* 1–8 (2022) doi:10.1038/s41587-022-01452-6.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

123. Cleary, B. & Regev, A. The necessity and power of random, under-sampled experiments in biology. (2020) doi:10.48550/arXiv.2012.12961.

124. Gemberling, M. *et al.* Transgenic mice for in vivo epigenome editing with CRISPR-based systems. 2021.03.08.434430 Preprint at https://doi.org/10.1101/2021.03.08.434430 (2021).

125. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

126. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

127. McFarland, J. M. *et al.* Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nat. Commun.* **11**, 4296 (2020).

128. 10x Genomics. User Guide: Chromium Single Cell 3' Reagent Kits v3 with Feature Barcoding technology for CRISPR screening. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://assets.ctfassets.net/an68im79xiti/7oWTi4259uwu06kmeCQG4g/caacae48b97b58660c7547cea9c067b6/CG000184_ChromiumSingleCellSingleCell3v3_FeatureBarcodingtechnology_CRISPR_RevA.pdf.

129. Li, B. *et al.* Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. *Nat. Methods* **17**, 793–798 (2020).

130. Dixit, A. Correcting Chimeric Crosstalk in Single Cell RNA-seq Experiments. 093237 Preprint at https://doi.org/10.1101/093237 (2021).

131. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

132. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

133. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proc. 9th Python Sci. Conf.* 92–96 (2010) doi:10.25080/Majora-92bf1922-011.

134. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. (2012) doi:10.48550/arXiv.1201.0490.

135. Liu, F. T., Ting, K. M. & Zhou, Z.-H. Isolation-Based Anomaly Detection. *ACM Trans. Knowl. Discov. Data* **6**, 3:1-3:39 (2012).

136. Rousseeuw, P. J. & Driessen, K. V. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics* **41**, 212–223 (1999).

137. Breunig, M. M., Kriegel, H.-P., Ng, R. T. & Sander, J. LOF: identifying density-based local outliers. *ACM SIGMOD Rec.* **29**, 93–104 (2000).

138. Sainburg, T., McInnes, L. & Gentner, T. Q. Parametric UMAP Embeddings for Representation and Semisupervised Learning. *Neural Comput.* **33**, 2881–2907 (2021).

139. Flamary, R. *et al.* POT: Python Optimal Transport. *J. Mach. Learn. Res.* **22**, 1–8 (2021).

140. Gobbi, A., Iorio, F., Jurman, G., Albanese, D. & Saez-Rodriguez, J. BiRewire: High-performing routines for the randomization of a bipartite graph (or a binary event matrix), undirected and directed signed graph preserving degree distribution (or marginal totals). (2022) doi:10.18129/B9.bioc.BiRewire.

141. Holland, C. H., Szalai, B. & Saez-Rodriguez, J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1863**, 194431 (2020).

142. Holland, C. H. *et al.* Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, 36 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

143. Herault, J. & Jutten, C. Space or time adaptive signal processing by neural network models. *AIP Conf. Proc.* **151**, 206–211 (1986).

144. Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **7**, 1129–1159 (1995).

145. Helwig, N. E. ica: Independent Component Analysis. (2022).

146. Luo, W. & Li, B. Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika* **103**, 875–887 (2016).

147. Nordhausen, K., Oja, H., Tyler, D. E. & Virta, J. ICtest: Estimating and Testing the Number of Interesting Components in Linear Dimension Reduction. (2022).

148. Kingma, D. P. & Welling, M. An Introduction to Variational Autoencoders. *Found. Trends® Mach. Learn.* **12**, 307–392 (2019).

149. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. Preprint at https://doi.org/10.48550/arXiv.1312.6114 (2022).

150. Higgins, I. *et al.* beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. in (2022).

151. Burgess, C. P. *et al.* Understanding disentangling in $\beta$-VAE. Preprint at https://doi.org/10.48550/arXiv.1804.03599 (2018).

152. Liu, Y., Sarkar, A., Kheradpour, P., Ernst, J. & Kellis, M. Evidence of reduced recombination rate in human regulatory domains. *Genome Biol.* **18**, 193 (2017).

153. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

154. Zhu, X. & Stephens, M. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat. Commun.* **9**, 4361 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

155. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

156. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

157. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).

158. Korotkevich, G. *et al.* Fast gene set enrichment analysis. 060012 Preprint at https://doi.org/10.1101/060012 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Acknowledgements**

We thank L. Gaffney and A Hupalowska for assistance with figure design and graphic, and Brian Cleary, Romain Lopez and members of the Regev and Theis labs for fruitful discussions. Work was supported by the Klarman Cell Observatory, National Human Genome Research Institute (NHGRI) Center for Excellent in Genome Science (5RM1HGP706193-09), and HHMI (AR), and NIH F32 Ruth L. Kirschstein Postdoctoral Fellowship (5F32AI138458) (PT).

**Author Contributions:**

Conceptualization: KGS, AR

Perturb-seq vector design: JD, KGS

Expanded Perturb-seq design and experimentation: KGS, OK, PIT, LZG, TD, DP, RR, NS

Sequencing: KGS, TD

Computational analysis design and execution: BE, KGS, KD, KJ, KDY

Supervision: AR, ORR

Writing – original draft: KGS, BE, AR, KD, KJ

Writing – review and editing: all authors

**Competing interests**: A.R. is a co-founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas, and until July 31, 2020 was an S.A.B. member of Thermo Fisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics and Asimov. From August 1, 2020, A.R. is an employee of Genentech, and has equity in Roche. From November 16, 2020, K.G.S. is an employee of Genentech. From March 22, 2021, P.I.T. is an employee of Genentech. From April 19, 2021, A.M.M. is an employee of Genentech. From August 22, 2022, A.R.Y. is an employee

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of Genentech. From October 19, 2020, O.R.R. is an employee of Genentech. A.R., O.R.R., K.G.S., B.E., and P.I.T. are inventors on multiple patents filed by or issued to the Broad Institute related to single cell genomics and Perturb-Seq. JGD consults for Microsoft Research, Abata Therapeutics, Servier, Maze Therapeutics, BioNTech, Sangamo, and Pfizer. JGD consults for and has equity in Tango Therapeutics. JGD serves as a paid scientific advisor to the Laboratory for Genomics Research, funded in part by GlaxoSmithKline. JGD receives funding support from the Functional Genomics Consortium: Abbvie, Bristol Myers Squibb, Janssen, Merck, and Vir Biotechnology. JGD's interests were reviewed and are managed by the Broad Institute in accordance with its conflict of interest policies. The other authors declare that they have no competing interest.

**Data and materials availability:** scRNA-seq data generated in this work are available from the Gene Expression Omnibus with accession number TBD and the Single Cell Portal with accession number TBD. Code used in analysis of sequencing data is available from GitHub link https://github.com/EraslanBas/PerturbDecode.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Supplementary Tables

**Table S1. Curated E3 gene family list**

**Table S2. Perturb-seq guides and oligo pools**

**Table S3. Gene lists for cell type signatures**

**Table S4. Top 1000 differentially expressed genes in cell clusters**

**Table S5. Enrichment or depletion of perturbed genes compared to the input library**

**Table S6. Enrichment or depletion of perturbed genes in cell subsets and clusters**

**Table S7. Regulatory coefficients in the model**

**Table S8. Gene programs and regulatory modules**

**Table S9. E3 and E3 complex member regulator metadata**

**Table S10. ICA factor genes, regulators and explained variance**

**Table S11. Regulatory coefficients in a model with genetic interactions**

**Table S12. Immunologic trait heritability in modules**

**Table S13. Extended sc-linker program heritability**

**Table S14. Primers for sequencing library generation**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Figure legends

**Figure. 1**. **Large scale Perturb-seq of the function of E3 ligases in the LPS response in BMDCs**

**A.** Study overview. Top: Experimental flow. Middle: Perturb-Seq vector design. Bottom: Example E3 ligases and complex members known to regulate different processes. **B.** PerturbDecode. Workflow diagram of key features (full diagram in **Figure S1E**). **C-E.** Screened cell populations captures multiple cell states and subtypes. Uniform Manifold Approximation and Projection (UMAP) of 519,535 cell profiles colored by cluster membership (**C**), cell type signature scores (**D-F**), cell cycle phase (**G**), or the difference of macrophage vs. DC signature scores (**STAR Methods**) (**H**). **I-J**. E3 perturbations affect DC subset distributions. **I.** Odd-ratio (color bar) of the significant (FDR < 0.15, one-sided Fisher's exact test) enrichment (pink) or depletion (blue) of guides targeting perturbed genes (rows) in each major cell subset (columns) in the screen. **J.** Summary of enrichment (pink) and depletion (blue) of guides targeting key proteins in major subsets (left) and in DC2 subtypes (right), colored by E3 family type and grouped by complex. ULD: ubiquitin-like domain; DUB: de-ubiquitylating enzyme; NS: non-significant.

See also related **Figures S1** and **S2**

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure 2. Six co-functional modules of E3 and related genes regulate 11 gene programs in the response of BMDCs to LPS**

**A.** Co-functional modules. Left: Distribution of E3 family member types across the 329 significant regulators. Right: UMAP embeddings of the regulatory profiles of the 329 regulators (KO genes), colored by their module membership (**STAR Methods**). **B.** Co-regulated programs. UMAP embeddings of the regulated profiles of 1,041 affected genes, colored by their gene program membership (**STAR Methods**). **C,D.** Regulatory model. **C.** Top left: Regulatory matrix (beta). Regulatory effect size (red/blue) of perturbing (KO) each of 329 genes (rows) on the expression of each of 1,041 affected genes (columns). Red/blue: Induction/repression in response to perturbation (KO) compared to control cells. Black horizonal and vertical line delineate co-functional modules and co-regulated programs, respectively. Top right: Co-functional modules. Covariance (green/purple) between the regulatory profiles in beta of perturbing each of 329 genes. Genes are clustered by module (as in **A**; color code on top and right). Bottom left: Co-regulated programs. Covariance (green/purple) between the regulatory profiles in beta of the effect on expression of each of 1,041 genes. Genes are clustered by program (as in **B**; color code on bottom and right). **D.** Regulatory network. Bipartite graph from co-functional modules to the co-regulated programs. Red point arrow/blue blunt arrow: module genes activate/inhibit program (*i.e.*, KO inhibits/activates program) (arrow color determined by significant mean difference). Key gene names are noted.

See also related **Figure S3**.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure 3. The genetic regulatory network is congruent with physical interactions and highlights modularity in E3-based regulation**

**A.** Three E3 ligases are highly regulated 'authorities' in the E3 network. Regulatory relations from the model (**Figure 2C**) based on a perturbation (KO) in an E3 ligase to another E3 ligase whose expression is affected. Red/blue arrows: perturbed E3 activates/inhibits target's expression (*i.e.*, KO inhibits/activates expression). **B-E.** Co-functional modules impact distribution of cell states. **B.** UMAP embedding of single cell profiles colored by Gaussian kernel density estimations of cells control (top left), M1 (top right) or M5 (bottom) guides. **C.** Supervised UMAP embedding of DC2.1, DC2.2, and DC2.3 cell profiles using a cells' co-functional module assignment as the response label (**STAR Methods**), colored by the module assignment of their guides (color code). **D.** Average Wasserstein distances (color bar) between cells with guides from different co-functional modules (rows, columns). **E.** Odd-ratio (colorbar) of significantly enrichment (pink) or depletion (green) of DC subsets (rows) in cells with guides from each co-functional module (columns) (FDR < 0.15, one-sided Fisher's exact test). **F.** Co-functional modules are enriched for protein-protein interaction partners. Physical interactions (blue, red, grey; experimental score > 0, STRING DB) or lack thereof (white) between each pair of 78 E3 ligases and adaptors with at least 24 interactions (with any of the 165 E3 ligases among the 329 regulators). Red/blue: the regulatory profiles of the physically interacting genes have significant (P<0.05) positive/negative correlation. Grey: the regulatory profiles of the physically interacting genes are not significantly correlated. Bars: co-functional modules. Rows and columns are hierarchically clustered. Full matrix in **Figure S4G**. **G,H,I.** TF explaining expression impact of E3 perturbations. **G.** Inferred activity scores (color bar) of 32 TFs (columns) whose target genes are significantly (FDR < 0.1) induced (yellow) or repressed (blue)

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

when perturbing each of 41 E3 and related genes (rows) (**STAR Methods**). (Full matrix in **Figure S4J**). **H,I.** Intersection between TF targets (DoRothEA), E3 expression targets, and gene programs.

See also related **Figure S4**.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure 4. Independent programs affected by distinct combinations of E3-ligase family members.**

**A.** E3 regulators association with independent factors from ICA. Left: Explained variance of effects on 1,041 genes of each of 203 perturbed genes (left matrix columns, sum of explained variance > 25%) or components of the CUL4-RBX1-DET1-RFWD2 complex (right matrix columns) by each of 15 latent factors (rows, main panel) and across all 15 factors (bottom). **B-J.** Member genes and regulators of example ICA factors. **B,E,H.** Effect sizes (yellow: positive; blue: negative; color bar) on significantly affected genes (columns; outlier loadings, separated by direction of effect) upon perturbation of each regulator gene associated with the factor (rows; outliers based on weights in the mixing matrix) (**STAR Methods**). Left bar: Co-functional modules; top bar: gene programs from the regulatory model. **C,D,F,G,I,J.** UMAP embedding of cell profiles (as in **Figure 1C**) colored by expression scores for each sub-factor (label on top).

See also related **Figure S5**.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure 5. Intra-module genetic interactions are more prevalent than inter-module interactions and modular structure can be leveraged to predict combinatorial perturbations**

**A.** Inter-module interactions are more prominent. Number of genes (y-axis) significantly affected by intra- (left bar) or inter- (right bars) module combinatorial perturbation (x-axis), additively (grey) or non-additively (blue). **B.** Intra-module interactions. Observed (y axis) fold-changes in expression (*vs*. control) in cells with perturbations in two genes from the same module and the expected fold change (x axis) from an additive model based on the two individual perturbations for each of 1,041 genes (dots). Slope of the first PC (red line) and variance in observed double knockouts explained by the single knockouts ($R^2$) are labeled. Module M4 is not shown to insufficient number of double-knockout cells. **C.** Inter- and intra-module interactions vary across programs. Significant effect sizes (red/blue color bar; FDR<0.1) of perturbations at the level of individual modules ($M_i$), inter-module pairs ($M_i$:$M_j$; i≠j), and intra-module pairs ($M_i$:$M_i$) (rows) on each of 1,041 genes (columns, labeled by gene program). Bottom row: Row centered mean expression in control cells. **D.** Substantial interactions between M3 and M5 regulators. Binarized significant effects (FDR<0.1, red/blue: positive/negative) on gene expression (rows, only genes with significant interaction terms) by single perturbations in regulators from M3 or M5, their additive effect (M3+M5), their interaction term (M3:M5), and observed combined effect (columns). **E-G**. comβVAE predicts combinatorial perturbations. **E.** Method overview. **F.** Distribution of explained variance ($R^2$, y axis) in fold changes of the 1,041 genes from 7 runs with the same hyperparameters at different KL loss weights (x-axis) for individual modules ($M_i$) and inter-module combinations ($M_i$:$M_j$; i≠j). **G.** Distribution of the explained variance ($R^2$, y axis) in fold changes of the 1,041 genes from 7 runs with the same hyperparameters in the

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

indicated inter-module combinations (labels on top) when the model (Beta=6.0) is trained only with data from single KOs from all modules (M) or single KOs from all modules and double KOs from one or two pairs of modules ($M_i$:$M_j$; $i \neq j$) (x axis). Boxes display the first ($Q_1$), second ($Q_2$, median) and third ($Q_3$) quartiles while the bottom and top whiskers show the intervals [$Q_1$ - 1.5 IQR, $Q_1$] and [$Q_3$, $Q_3$ +1.5 IQR], respectively.

See also related **Figures S6** and **S7**.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure 6. Co-functional modules and gene programs are enriched for risk heritability and cell programs in human immunological disease.**

**A.** Module M6 associated with genetic disease risk for human immunological disease. Significant MAGMA Z-scores (color bar; per trait; Bonferroni $\alpha < 0.1$) for immunological disease traits (columns) of module M6 E3 family genes (rows) with at least one significant score. **B.** Gene programs associated with genetic disease risk for human immunological disease. Significance ($-\log_{10}$(p-value), dot color) and effect size (dot size) of heritability enrichment in each gene program (rows) for different immune traits (columns) by sc-linker analysis with SNP annotations combined with intersection of Roadmap and ABC gene-enhancer linking strategy (left) or by MAGMA (right). **C,D.** Genes programs expressed during immunological disease progression in immune and non-immune cells. Enrichment (color bar) of gene programs (columns) for cell type specific disease progression programs in humans (rows), across diverse cell types and diseases (C) or only in DCs and macrophages in UC, fibrosis, asthma, and COVID-19 (D). **E.** Perturb-Seq regulatory model highlights E3 regulators of rare IBD disease genes. Regulatory coefficient (color bar, from the model of **Figure 2C**) of the impact of perturbing regulators (rows) on the expression of genes (columns) with rare variants associated with IBD that also have at least one significantly regulating E3 family member. **F.** Gene program enriched for rare variant in immunological disease. Significance ($-\log_{10}$(p-value), dot color) and effect size (dot size) of heritability enrichment of gene programs (rows) for CD or IBD based on rare (CD:SAIGE-GENE) or common (IBD:MAGMA and CD: MAGMA) variants.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure 7. The DC life cycle regulated by E3 ligases**

ICA factors and their key regulators grouped in each DC lifecycle stages (boxed). Bottom: UMAP embedding of cell profiles (as in **Figure 1C**) colored by expression scores (color bar) for migration-related factors.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Supplementary Materials

### Figure S1, related to Figure 1. Design and performance of large-scale Perturb-Seq in primary BMDCs

**A-D.** Sorting of perturbed cells for profiling. **A,B.** Forward scatter (x axis) versus side scatter (y axis) for selected live (A) single (B) perturbed BMDCs. **C.** GFP fluorescence (x axis; Cas9 mice cells) versus mKate2 fluorescence (y axis; Perturb-Seq vector) to select $mKate2^{+}GFP^{+}$ cells. **D.** Distribution of mKate2 expression (x axis) in sorted live, single cells. **E.** PerturbDecode. Detailed workflow. **F-J.** Quality control for impact of guide-based perturbations. **F.** Cumulative distribution function (CDF) (y-axis) of Pearson's *r* (x-axis) between the effect sizes of guides targeting the same gene (purple), different genes (red), one gene and one no-target control (green), or one gene and one intergenic control (blue). **G.** Distribution of number of genes (y-axis) (of 6,685 tested genes) significantly affected (FDR < 0.1) by non-targeting controls, intergenic controls or targeting guides (with guides targeting the same gene combined) (x-axis) (**STAR Methods**). **** $P < 2.2*10^{-16}$, one-sided Wilcoxon rank-sum test. **H.** Significant effect sizes (color bar; blue/red negative/positive fold-change; FDR<0.1) of perturbing each of 544 targets (rows) that were also among the 6,685 genes with tested expression on itself and the other 544 targets (columns). Rows and columns are ordered alphabetically. 137 of 539 genes significantly negatively affected their own expression (diagonal). **I.** Number of genes (y axis) whose expression is significantly (FDR<0.1) affected by perturbation of each of 849 perturbed genes (from 13,811 detected genes) and the mean expression of these perturbed genes (x axis, normalized log1p). Pearson's *r* and significance in the upper right corner. **J.** Distribution of number of genes (y axis) significantly affected (FDR < 0.1) by the perturbation of genes that are

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

("expressed") or are not ("not expressed") in the 13,811 detected genes. **** P-value < $10^{-4}$, one-sided Wilcoxon rank-sum test.

**Figure S2, related to Figure 1. Proportions of cells in different states and subtypes in screen are affected by perturbations.**

**A-G**. Markers of specific cell subsets in the screened BMDC population. Mean expression (dot color, mean normalized log1p expression) and fraction of expressing cells (dot size) for genes differentially expressed (columns) in each of the 10 cell clusters (A, rows) or in the major cell sub-types (B-G, rows). **H,I.** DC- vs. macrophage-like gradient. UMAP embedding of 519,535 cell profiles (as in **Figure 1C**) colored by DC (DC1+DC2+mDC genes) or macrophage signature score. **J-V.** Cell subsets and stated in unperturbed resting or LPS-stimulated BMDCs. UMAP embedding of 3,655 unperturbed unstimulated and 4,027 unperturbed and LPS stimulated (3h) BMDC profiles colored by treatment (J), inferred cell cycle phase (K), signature scores for the top 100 upregulated genes of each of the 10 clusters of **Figure 1C** (**L-U**), or their predicted major cell subtype (**V**) (**STAR Methods**). **W.** Perturbation and stimulation affect cell subtype proportions. Percentage of cells (y axis) of each of four major subtypes in the screen (color legend) in unperturbed unstimulated, unperturbed LPS stimulated and perturbed, LPS stimulated data (x-axis). * P < $2.2*10^{-16}$, one-sided Fisher's exact test. **X,Y.** Specific perturbations are enriched or depleted in specific cell subsets and clusters. Odd-ratio (color bar) of enrichment (pink) or depletion (blue) (FDR < 0.15, one-sided Fisher's exact test) of cells with a perturbed gene (rows) in cell cycle phases in major subtypes (X, columns) or in the 10 cell clusters (Y, clusters as in **Figure 1C**).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure S3, related to Figure 2. Gene programs in the regulatory model**

**A-K.** Gene programs expressed in different cell subsets. UMAP embedding of cell profiles (as in **Figure 1C**) colored by expression scores of each program genes. **L,M.** Relation of gene programs to other programs and cell subset signatures of BMDCs. **L.** Jaccard index (left) and fractional overlap (right) between each gene program (rows, "A") and programs in an earlier small Perturb-Seq screen of 24 TFs in the LPS-stimulated BMDCs [16] (left, columns, "B") or DC subset signatures [39] (right, columns, "B"). **M.** Distribution of program scores (y axis) for DC1-, DC2-, mDC-, and macrophage-like cell subsets (x axis). * $P < 0.05$ one-*vs*.-rest one-sided Wilcoxon rank sum test.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure S4, related to Figure 3. Relation between regulatory (genetic) and physical (protein-protein and TF-target) interactions.**

**A,B.** Three E3 ligases are 'authorities' highly regulated by other E3s. Binarized regulatory effects (blue/red: negative/positive) of perturbing each of 60 E3 ligases in the model that significantly affected the expression of at least one other of the 60 E3 ligases, for all 60 E3 ligases (A) or for the 15 that both impact and are impacted by another E3 (B). Module membership is labeled by color on left and top. Negative effects of the KOs on its own RNA level are not shown in A and shown in B. **C-F.** Co-functional modules affect distribution of cell states. **C.** UMAP embedding of cell profiles (**as in Figure 1C**) colored by Gaussian kernel density estimations of cells with control guides (top left) or guides targeting genes with each module (label on top). **D.** Supervised UMAP embedding of DC2.1, DC2.2, and DC2.3 cell profiles using a cells' co-functional module assignment as the response label (as in **Figure 3C**, **STAR Methods**), colored by the difference of macrophage and DC signature Z scores. **E.** Percentage of cells (y axis) with guides targeting genes in each module (x axis) belonging to each of the cell clusters of **Figure 1C** (colors). **F.** Percentage of cells (y axis) in each of the cell clusters of **Figure 1C** (x axis) with guides targeting genes in each module (colors). **G-I.** Congruence between protein interactions and genetic effects. **G.** Physical interactions (blue, red, grey; experimental score > 0, STRING DB) or lack thereof (white) between each pair of 165 E3 ligases among the 329 regulators (rows, columns). Red/blue: the regulatory profiles of the physically interacting genes have significant ($P<0.05$) positive/negative correlation. Grey: the regulatory profiles of the physically interacting genes are not significantly correlated. Genes are sorted by module membership (colors on top and left). **H.** Physical interactions (edges, STRING DB experimental score > 0) between NFkB signaling pathway components included in the regulatory model (nodes), colored by significant

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(P<0.05) positive (red) or negative (blue) correlation of their perturbation effects. **I.** Top: physical interactions (blue, red, grey; experimental score > 0, STRING DB; color code as in G) or lack thereof (white) between each perturbed CLR E3 ligases (rows) and their CLR complex members, including adaptor domain proteins (columns). Columns are ordered by CLR physical complexes / interactions (boxes and dashed lines). Bottom: As on top, except all significant covariances are displayed regardless of interaction evidence. **J,K.** TFs impacted by and mediating the effects of E3 perturbation. **J.** Inferred activity scores (colorbar) of 109 TFs (columns) whose target genes are significantly (FDR < 0.1) induced (yellow) or repressed (blue) (by at least 10 perturbations) when perturbing each of 156 E3 and related genes (rows) (that affected at least 10 of the 109 TFs) (**STAR Methods**). **K.** Intersection between TF targets (DoRothEA), E3 expression targets, and gene programs.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure S5, related to Figure 4. Independent factors in the regulatory model**

**A,B.** Genes and regulators associated with each factor. Regulators (A, rows) and member genes (B, columns) associated (green) with each factor (columns) by their outlier weights in the mixing matrix or outlier loadings in the matrix of source signal estimates, respectively. **C-E.** Choice of number of factors. 'gn'-main criterion for the ladle estimate [146] (C, y-axis, **STAR Methods**), explained variance (**D**, $R^2$, y-axis) after matrix reconstruction with estimated components, and distribution of explained variance (**E**, y axis) in randomly sampled unseen perturbation responses (**STAR Methods**) for ICA decomposition with different numbers of components (x-axis). **F.** Factor expression across cells. UMAP embedding of cell profiles (as in **Figure 1C**) colored by expression scores for each sub-factor (label on top). **G,H.** Jaccard index (color bar) for each pairs of factors (columns and rows) based on genes with outlier loadings in the matrix of source signal estimates (**G**) or regulators with outlier weights in the mixing matrix per factor (**H**).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure S6, related to Figure 6. Inter- and intra-module genetic interactions**

**A.** Combinatorial perturbations in the E3 screen. Number of cells (color bar, number) with pairs of guides targeting genes in the same or pair of modules (rows, columns). **B.** Genes vary in the extent of genetic interactions on their expression. Number of significant interaction terms (FDR < 0.1, y-axis) on the expression of each gene (x axis) from combinatorial perturbations across all intra-modules (red) and inter- (orange) module interactions. **C.** Module pairs vary in the extent of interactions. Number of target genes (y axis, of 1,041 in the regulatory model) with a significant effect on expression (blue) due to single perturbations in genes in one module ($M_i$, x axis) or with a significant interaction term due to perturbation in two genes from the same ($M_i$:$M_i$, x axis) or different ($M_i$:$M_j$, $i \neq j$, x axis). **D,E.** Specific inter-module interactions across genes. Binarized significant effect (FDR < 0.1) (red/blue: positive/negative) on the expression of genes (row, only genes with significant interaction terms) by single perturbations in regulators from two different modules ($M_i$, $M_j$, $i \neq j$), their additive effect ($M_i$+$M_j$), their interaction term ($M_i$:$M_j$), and the observed effect (columns). Gene program membership is labeled on left. **F-H.** Observed vs. additive effects across genes. Fold changes in gene expression observed (*y* axis) following inter-module combinatorial perturbation or predicted by an additive model (*x* axis) for each of the 1,041 genes (F, dots) or only genes with either significant (G, FDR < 0.1) or non-significant (H, FDR >= 0.1) inter-module interaction terms. $R^2$: explained variance in observed fold changes; MAE: mean absolute error of the predictions.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure S7, related to Figure 6. ComβVAE for prediction of combinatorial perturbations**

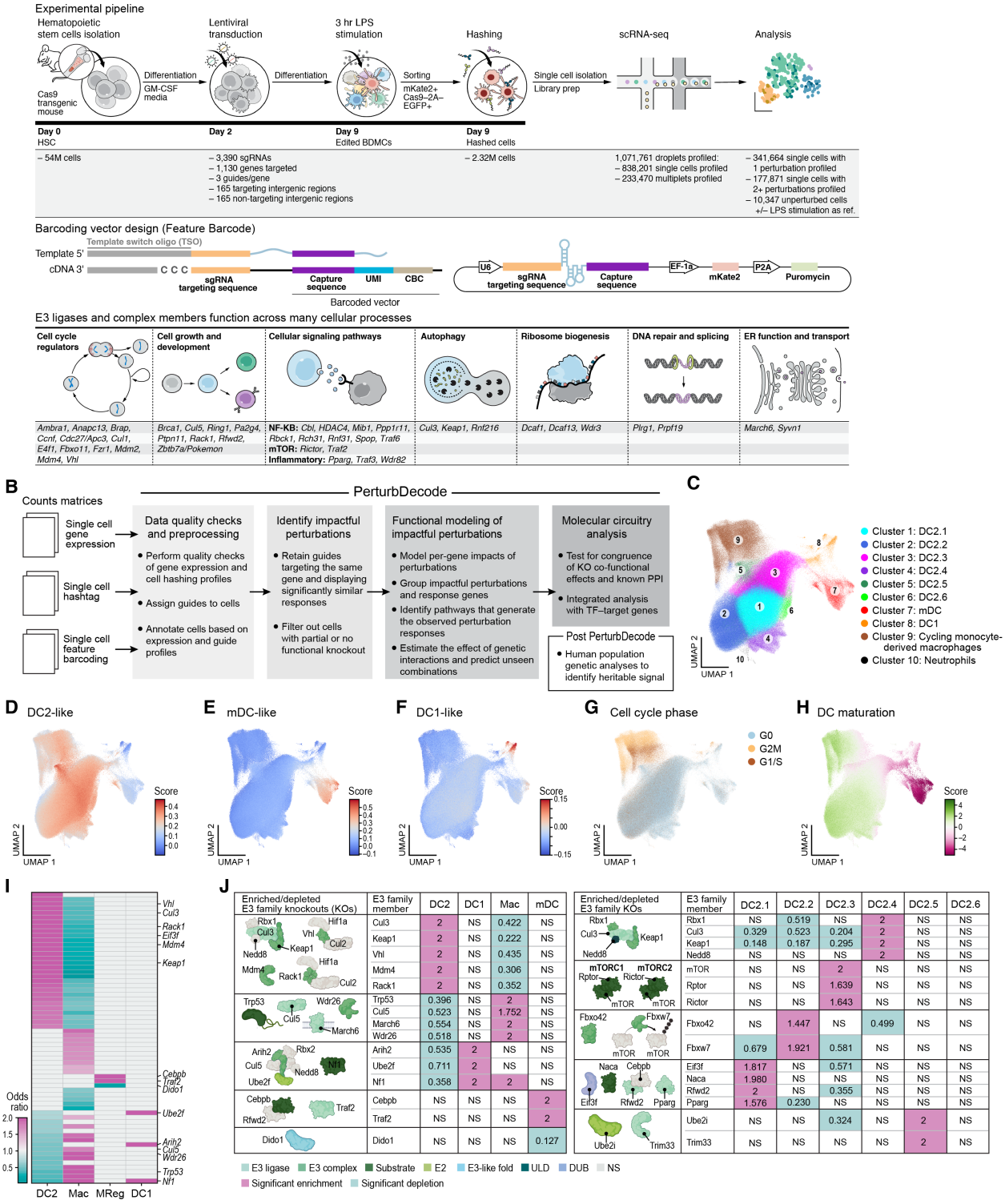
**A,B.** Fold change predictions from comβVAE. Fold changes in gene expression observed (*y* axis) following inter-module combinatorial perturbation (y-axis) or predicted by comβVAE (*x* axis) for each of the 1,041 genes (A, dots) or only genes with significant (B, FDR < 0.1) inter-module interaction terms. The diagonal entries in (A) reflect the prediction in single knockouts. **C.** comβVAE performance for different modules and module combinations. Distribution of explained variance (top, y axis, $R^2$) and mean absolute error (bottom, y axis, MAE) of the predictions of the comβVAE model for each module ($M_i$) or inter-module combination ($M_i$:$M_j$, i≠j) (x axis) across 7 runs with the same hyperparameters. **D.** Impact of changes in KL loss weight. Distribution of explained variance in fold changes (D, y-axis) of the comβVAE at different KL loss weight values (*x* axis) for each module ($M_i$) or inter-module combination ($M_i$:$M_j$, i≠j) across 7 runs with the same hyperparameters. **E-G.** Impact of training with doubly perturbed cells. **E.** Distribution of the explained variance in fold changes (y axis, $R^2$) in the indicated inter-module combinations (labels on top of panel) when the model (Beta=6.0) is trained only with data from singly perturbed cells from all modules (M) or singly perturbed cells from all modules and doubly perturbed cells from one or two pairs of modules ($M_i$:$M_j$; i≠j) (x axis) across 7 runs with the same hyperparameters. **F.** Explained variance in fold changes (y-axis) in each module pair (panels) when the comβVAE is learned with different KL loss weight values (*x* axis) and trained either only with singly perturbed cells (red) or with both singly perturbed and doubly perturbed cells of specific module pairs (green: M3M5, blue: M5M6, purple: M3M5 and M5M6). **G.** Explained variance in fold changes (y-axis) in select module pairs with relatively high number of genes with significant inter-module interaction terms (column headers) when the comβVAE is learned at different KL loss weight values (*x* axis) and

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

trained either only with singly perturbed cells (red) or with both singly perturbed and doubly perturbed cells of specific module pairs (row labels). All boxes in Box plots display the first ($Q_1$), second ($Q_2$, median) and third ($Q_3$) quartiles, and bottom and top whiskers show the intervals [$Q_1$ -1.5 IQR, $Q_1$] and [$Q_3$, $Q_3$ +1.5 IQR], respectively.

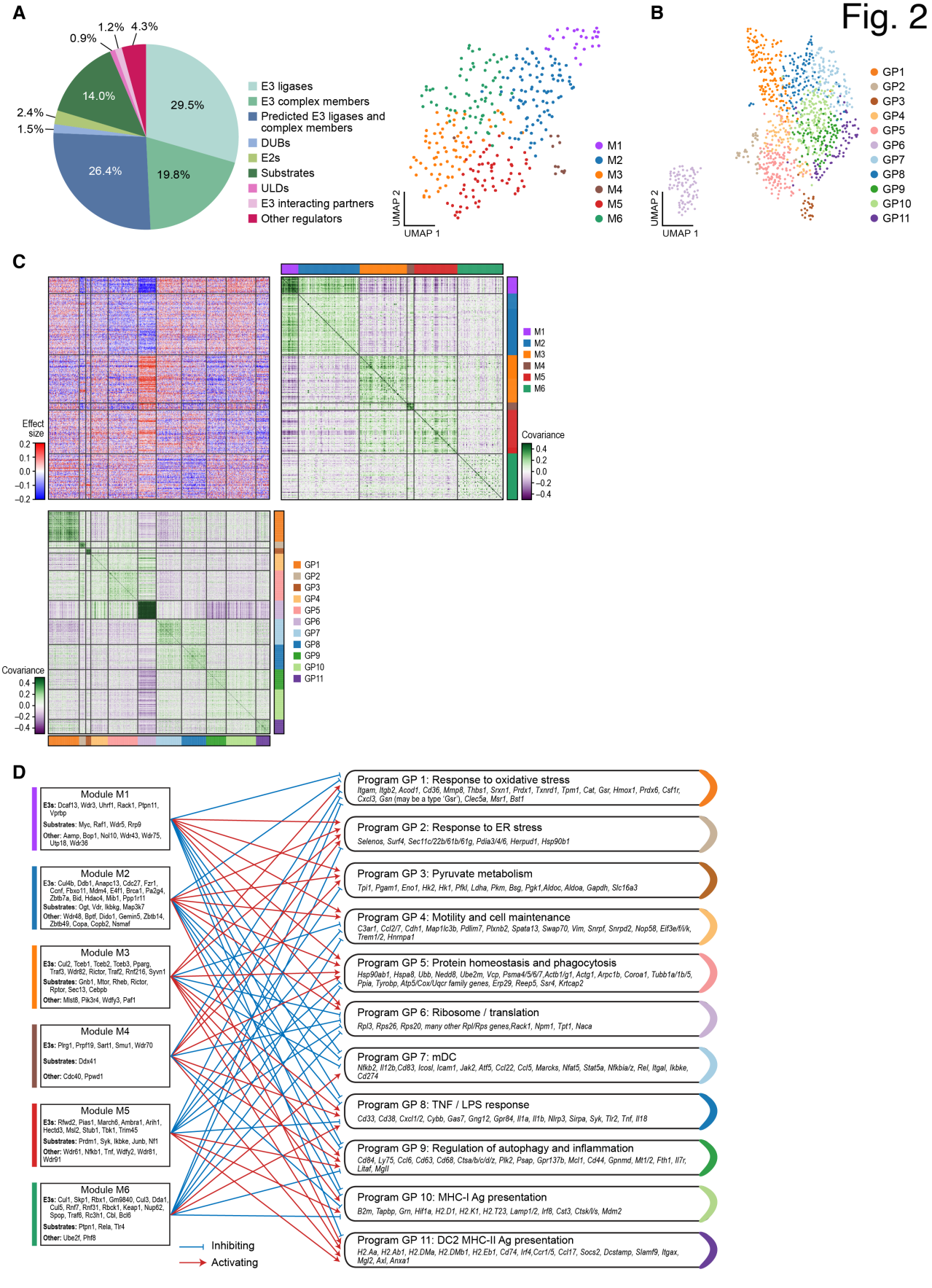
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Fig. 1



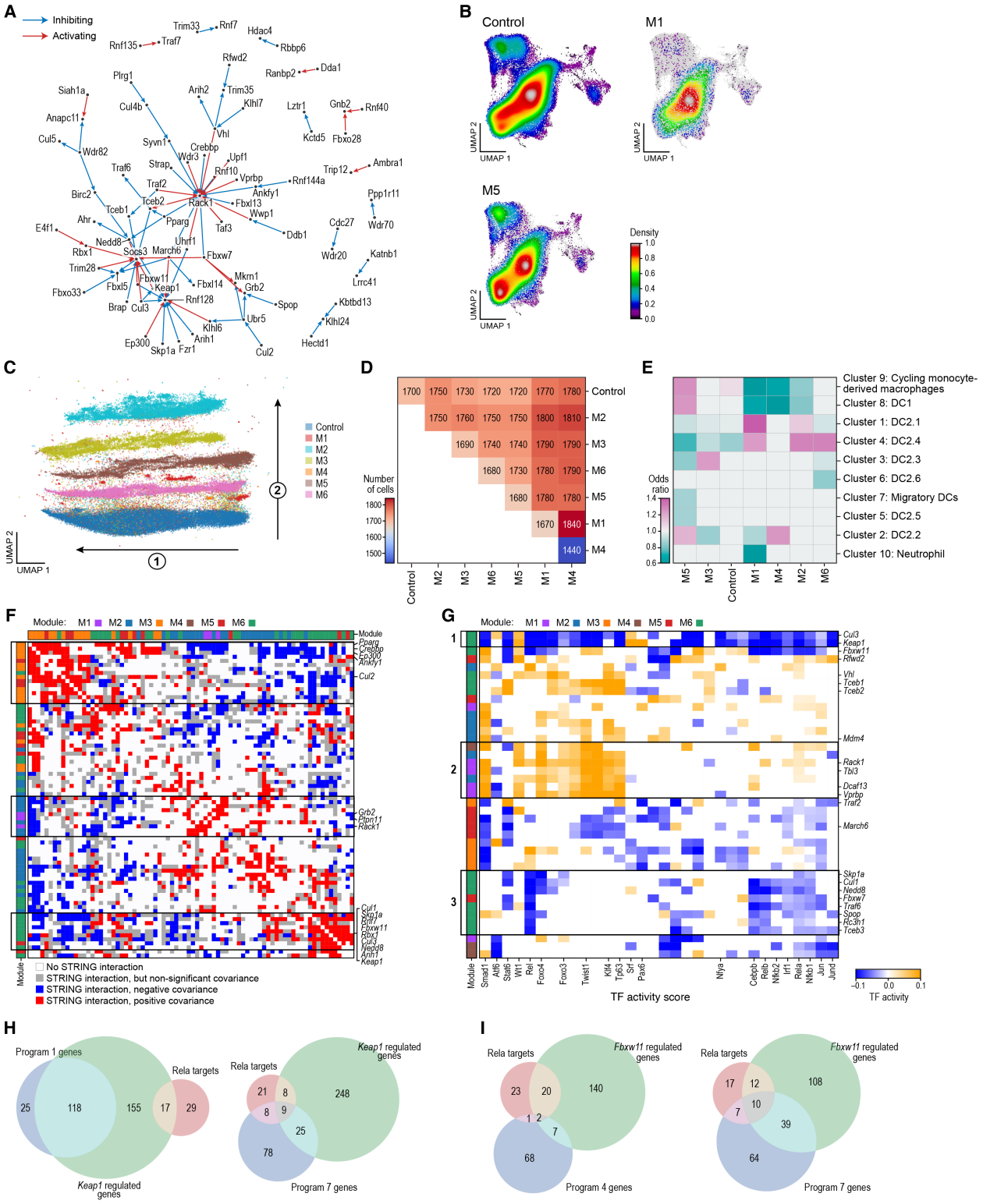
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 2



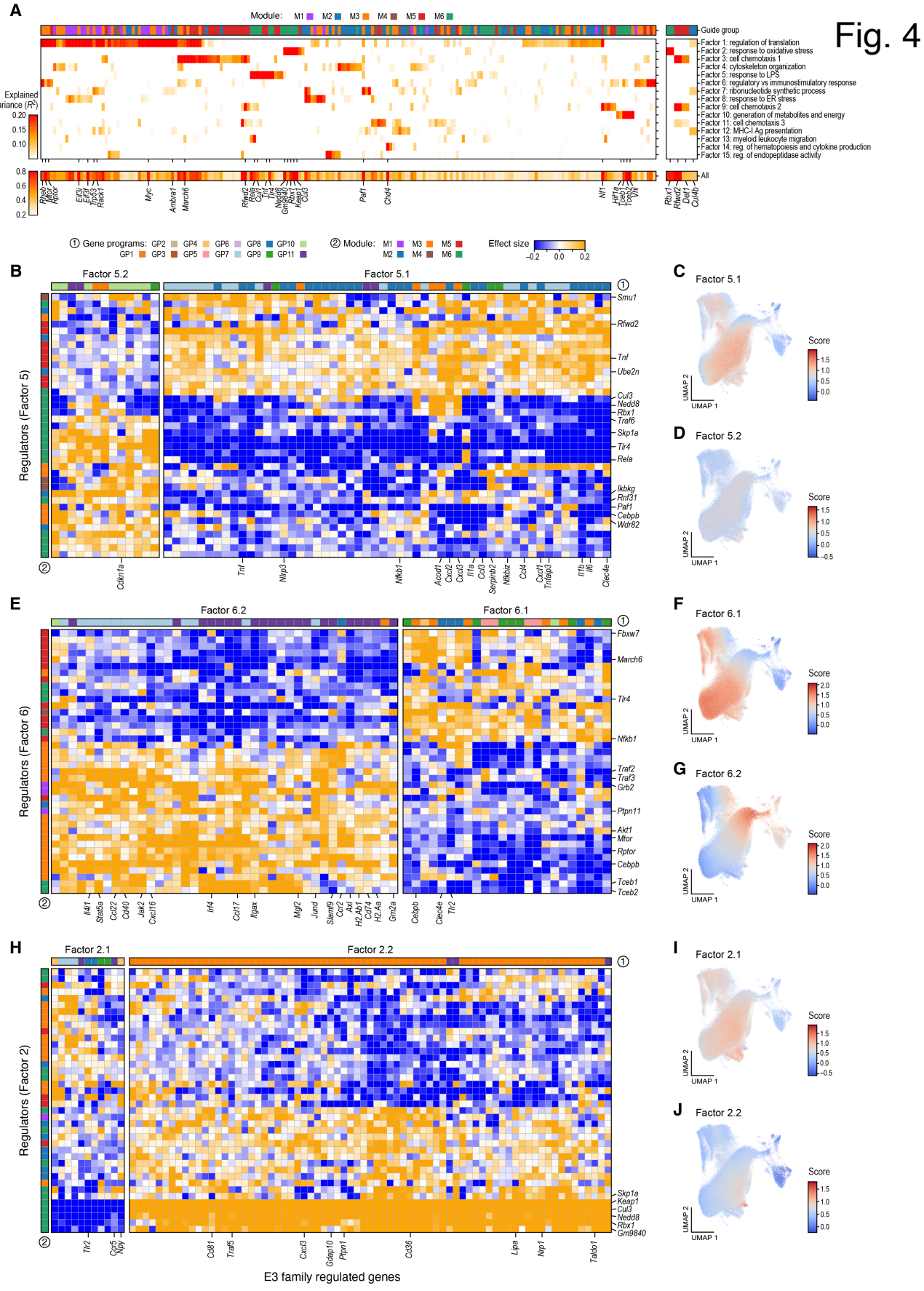
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Fig. 3



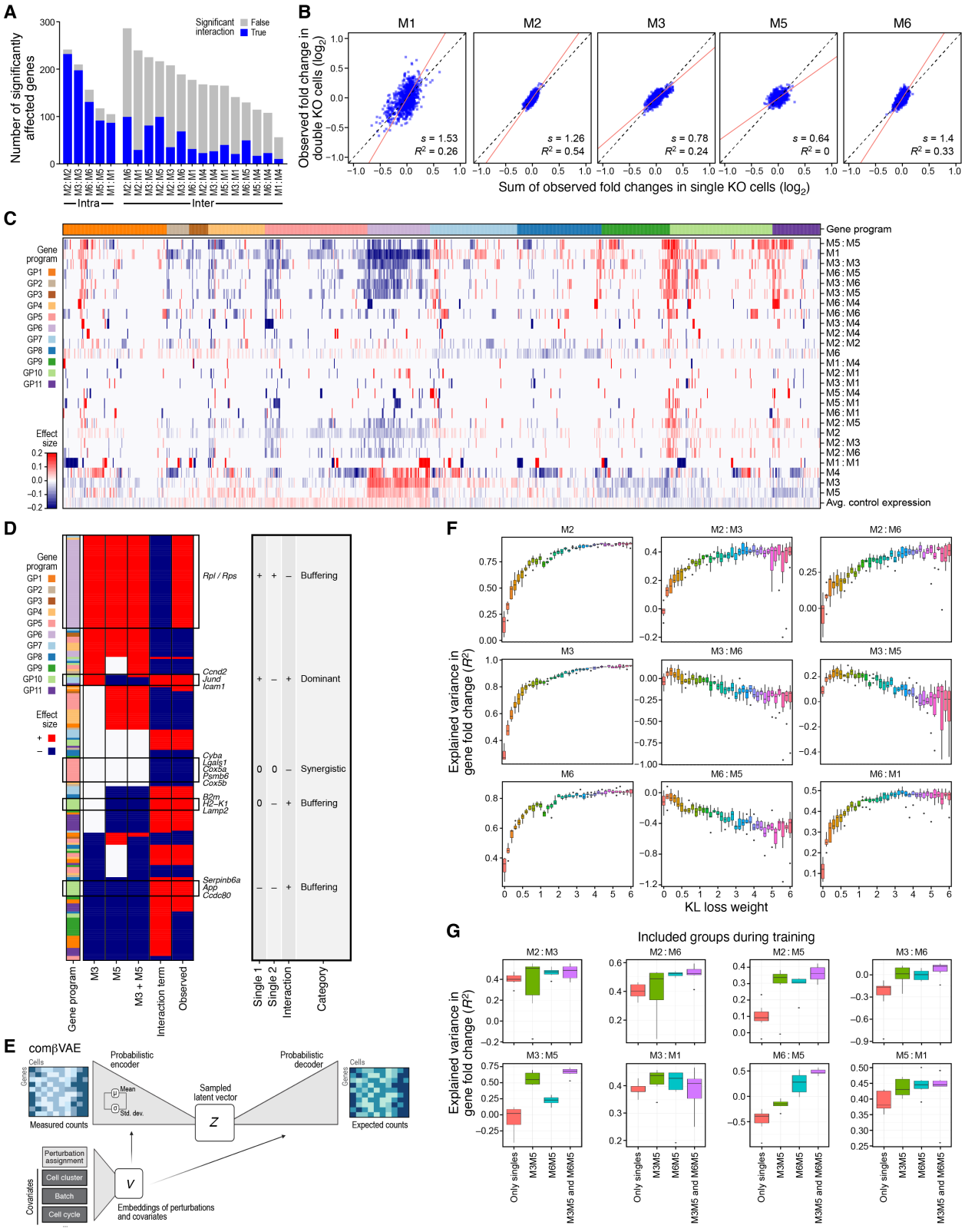
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 4



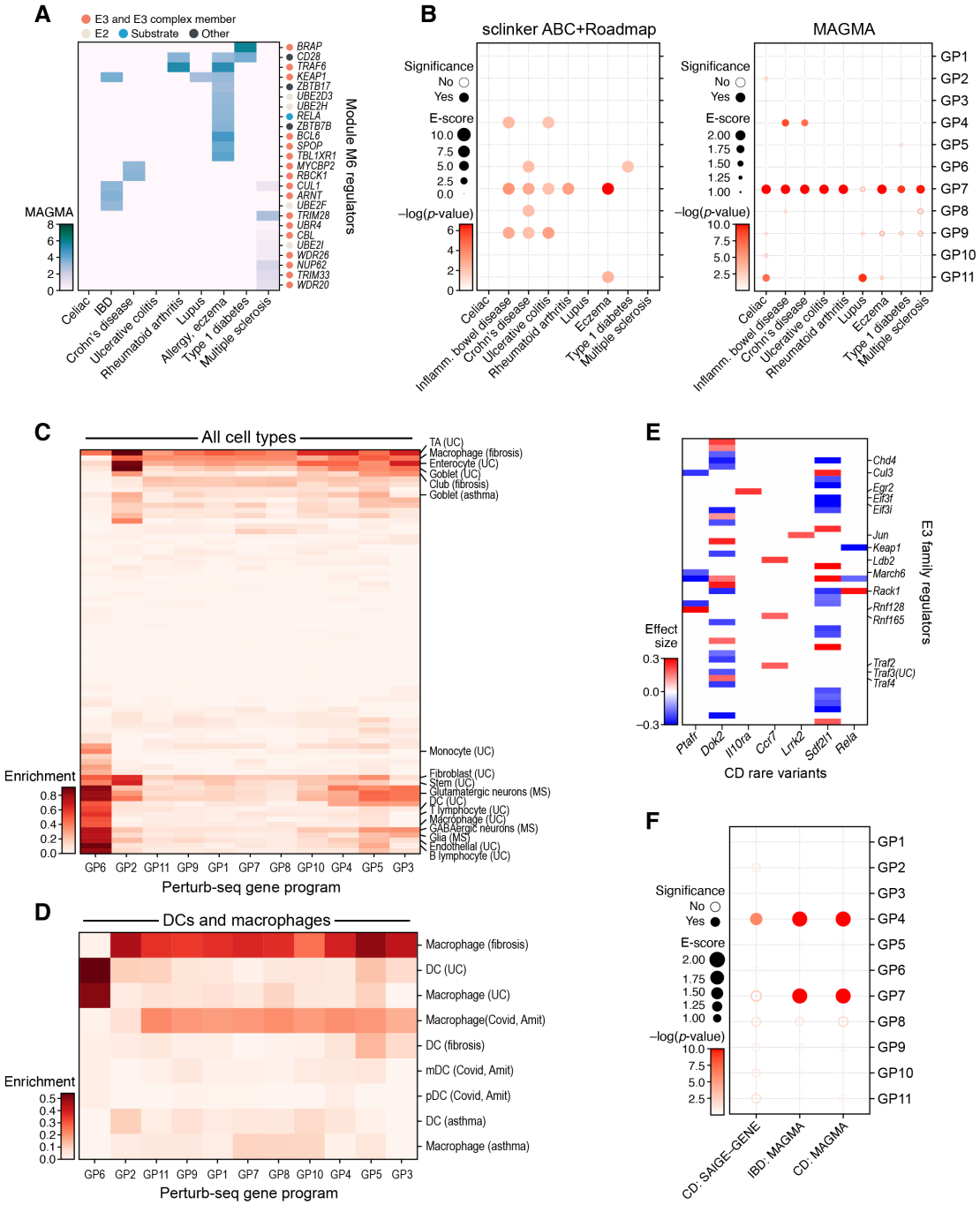
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Fig. 5



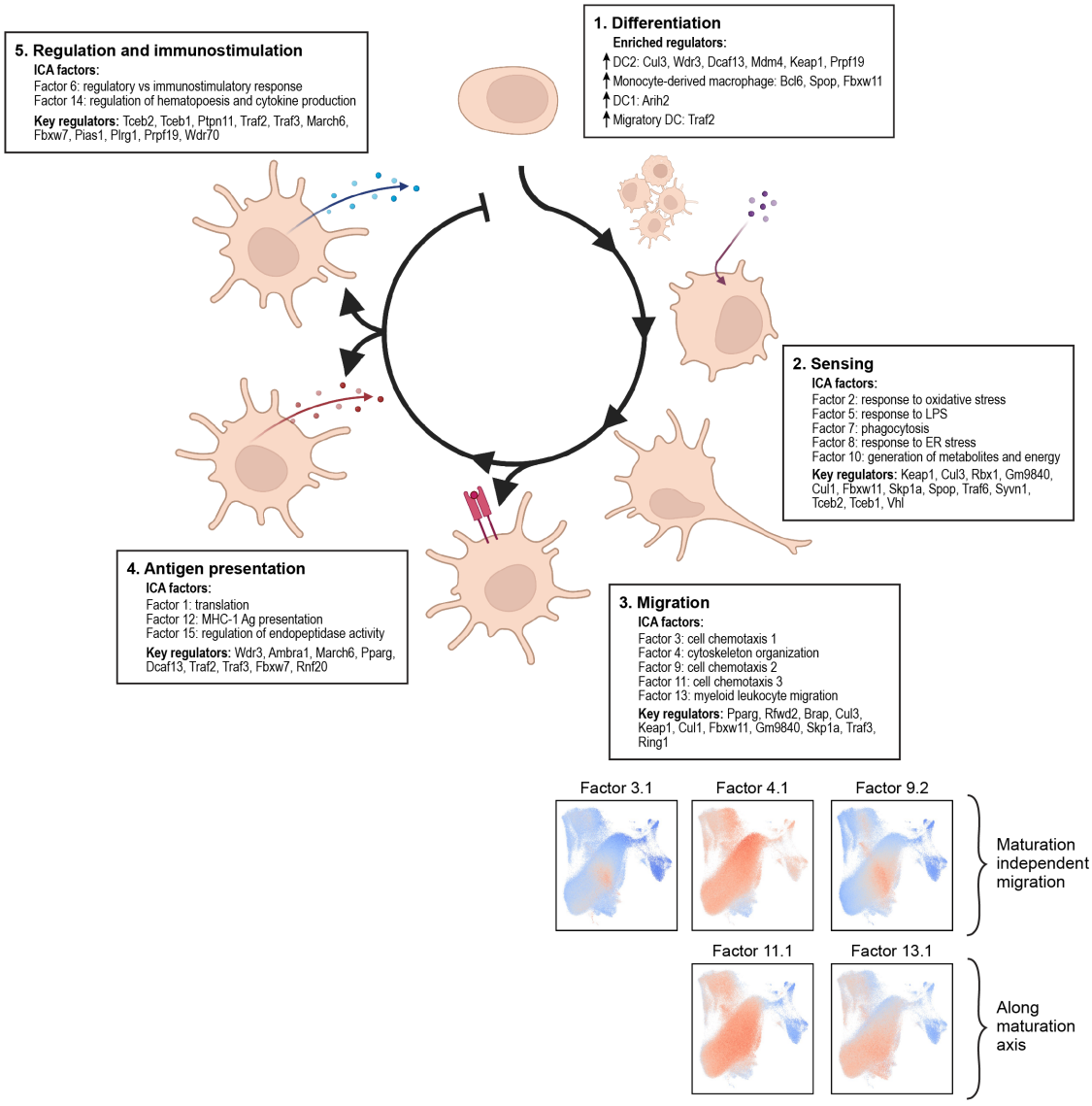
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 6



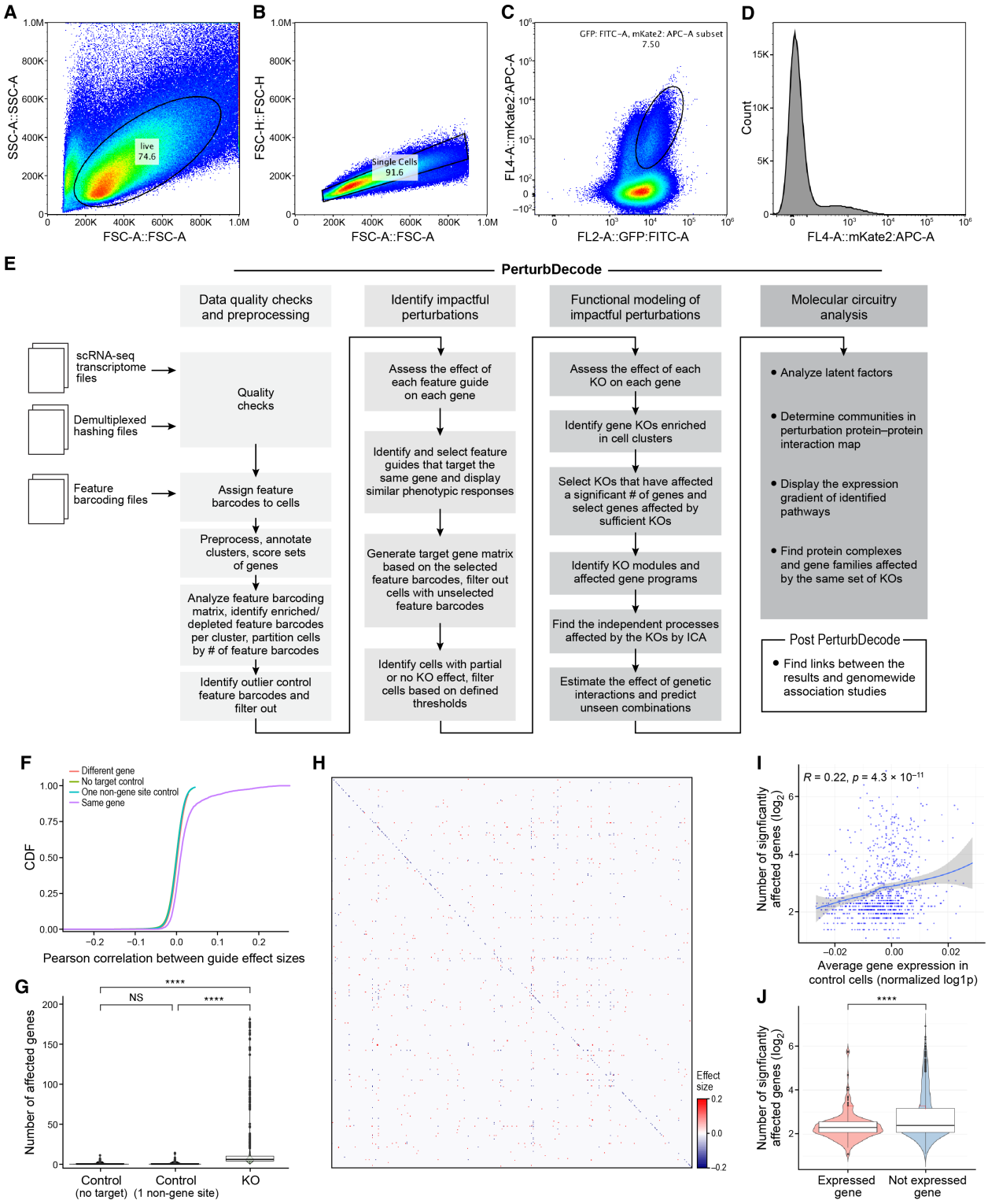
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 7



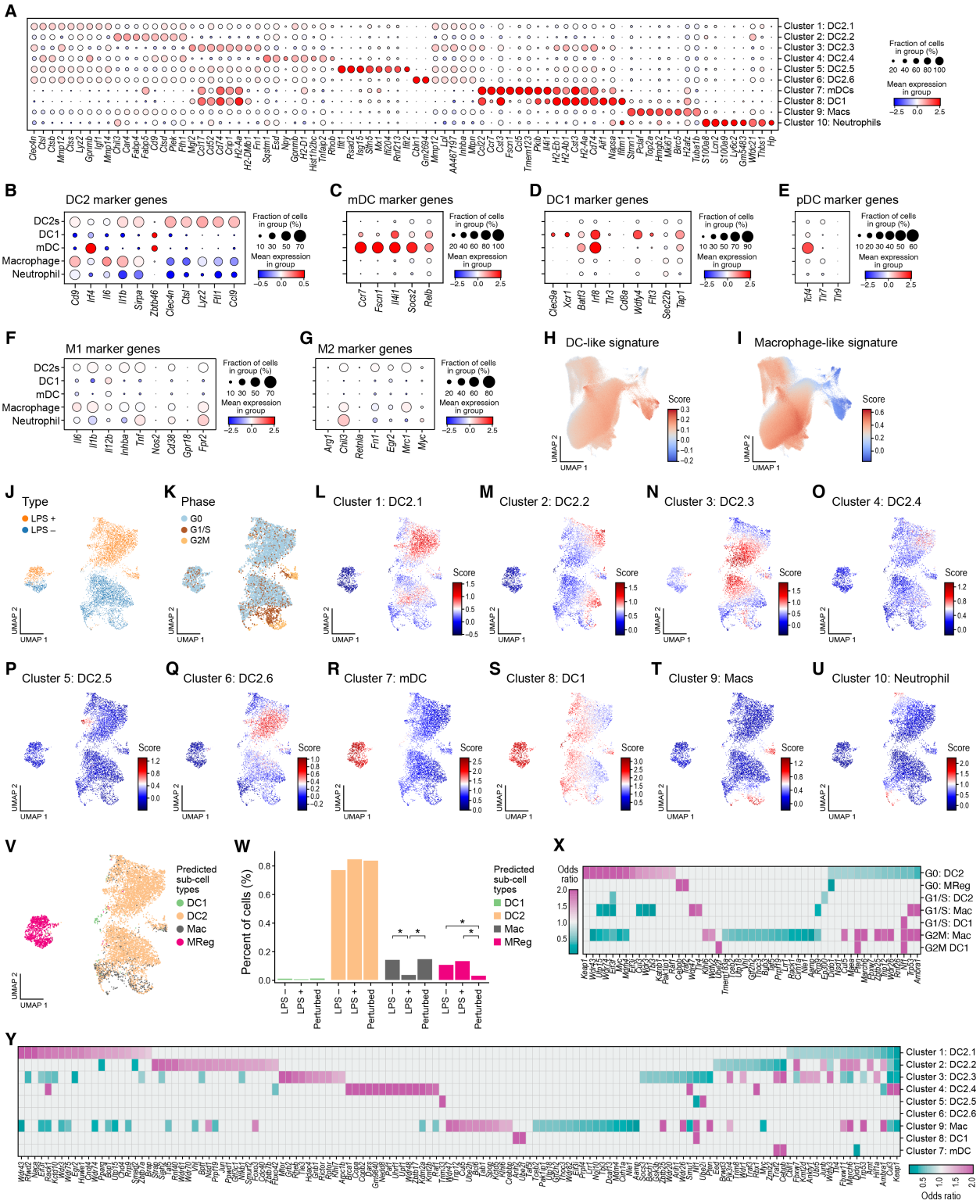
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Supp. Fig. 1



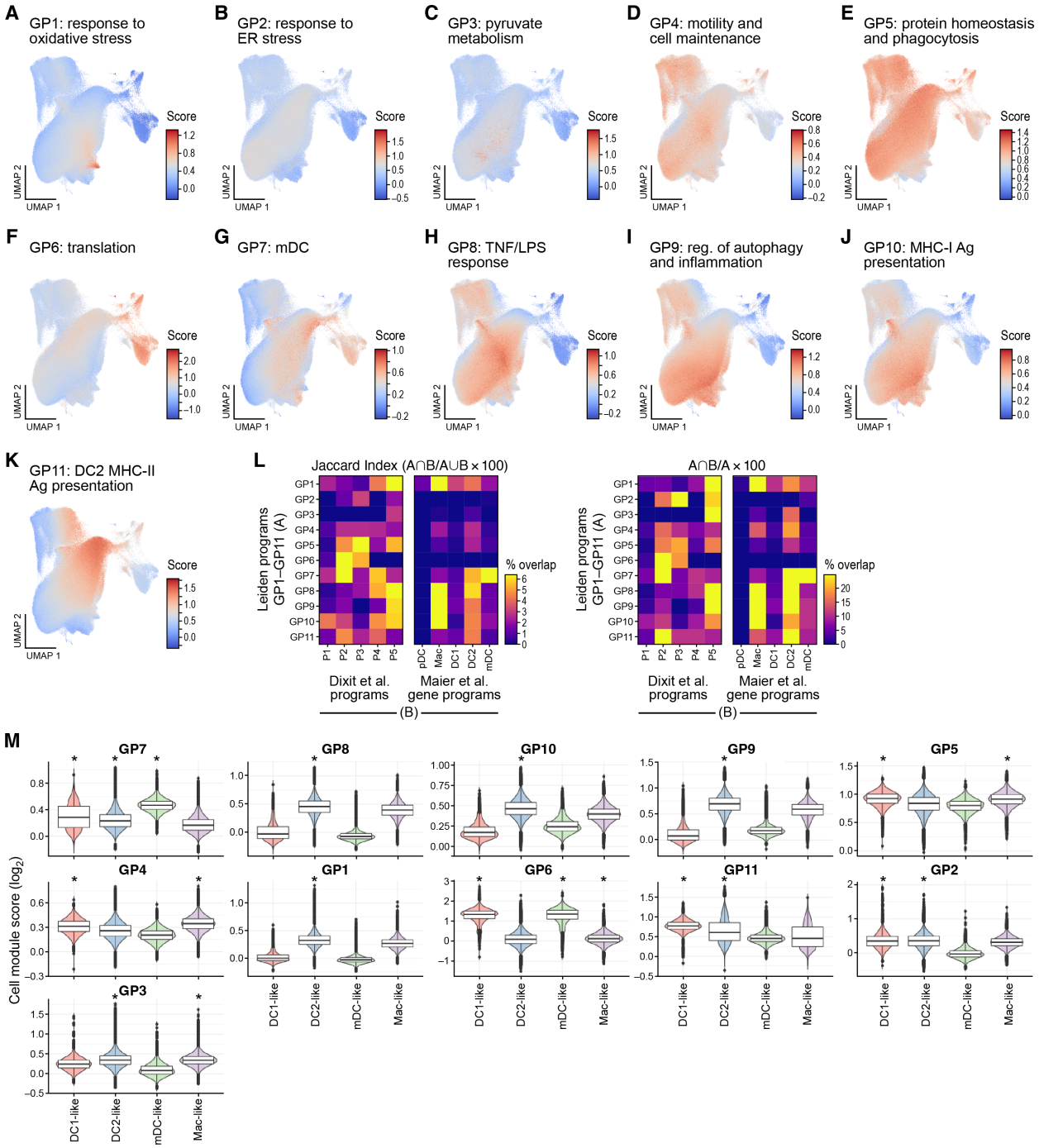
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Supp. Fig. 2



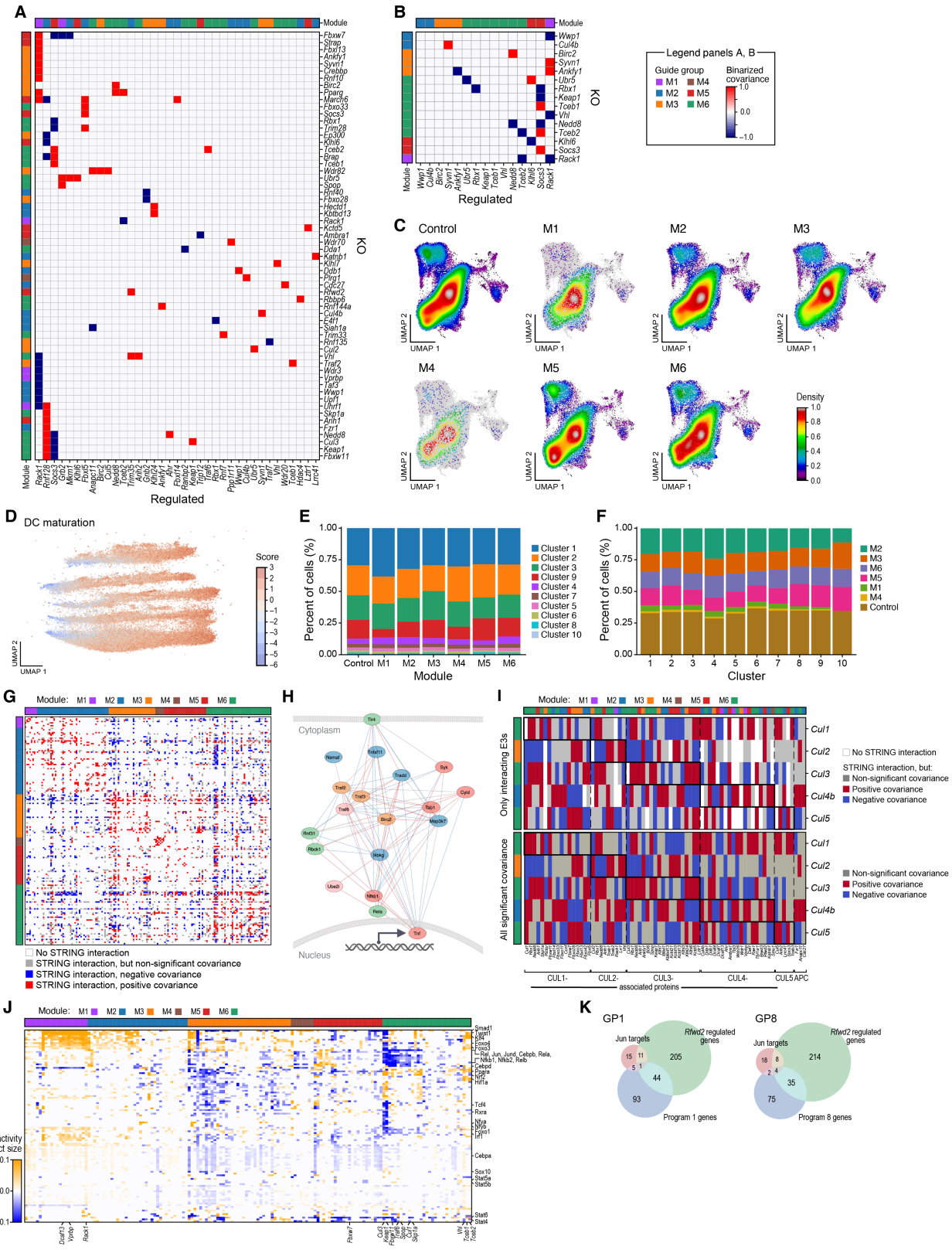
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Supp. Fig. 3



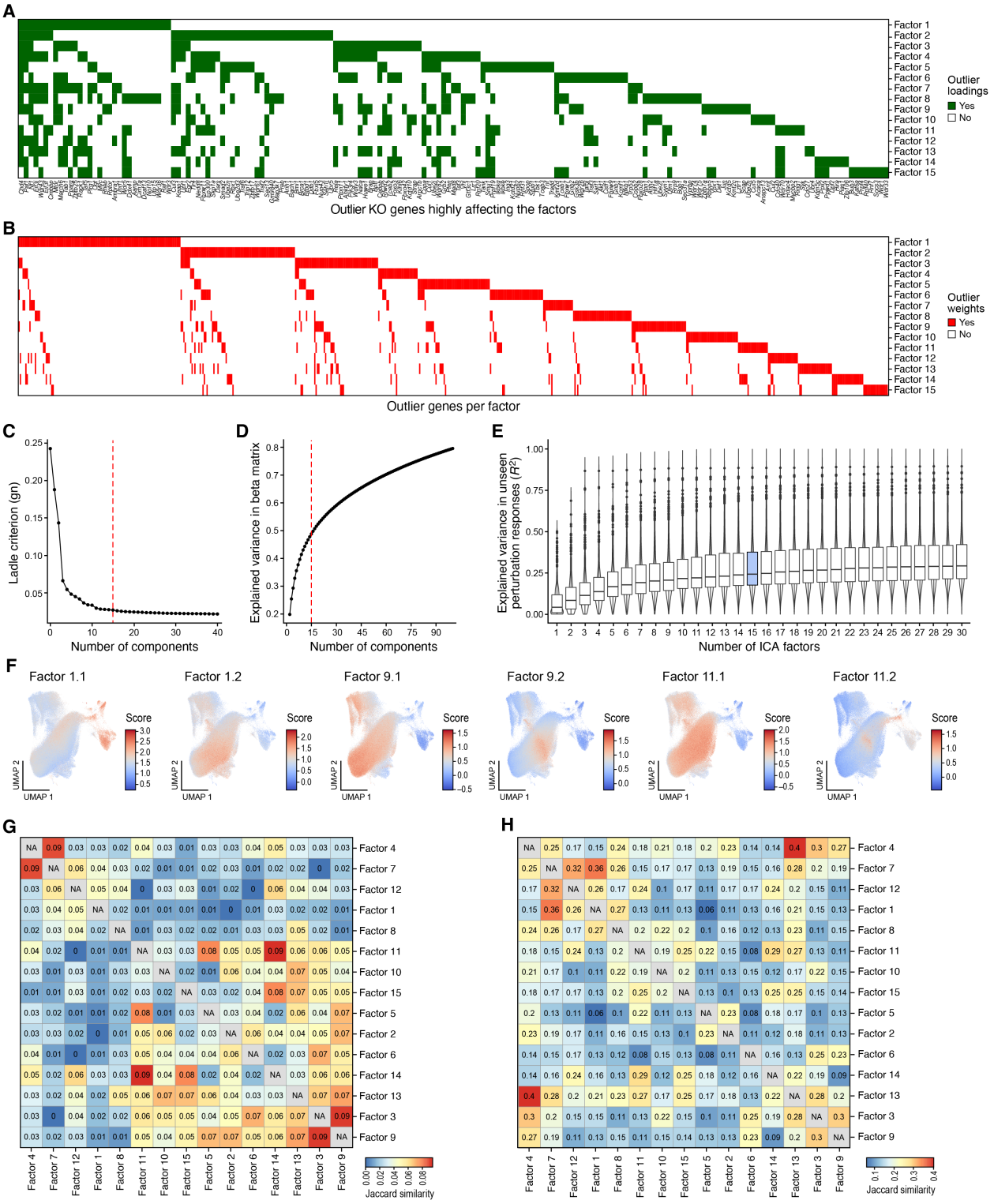
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Supp. Fig. 4



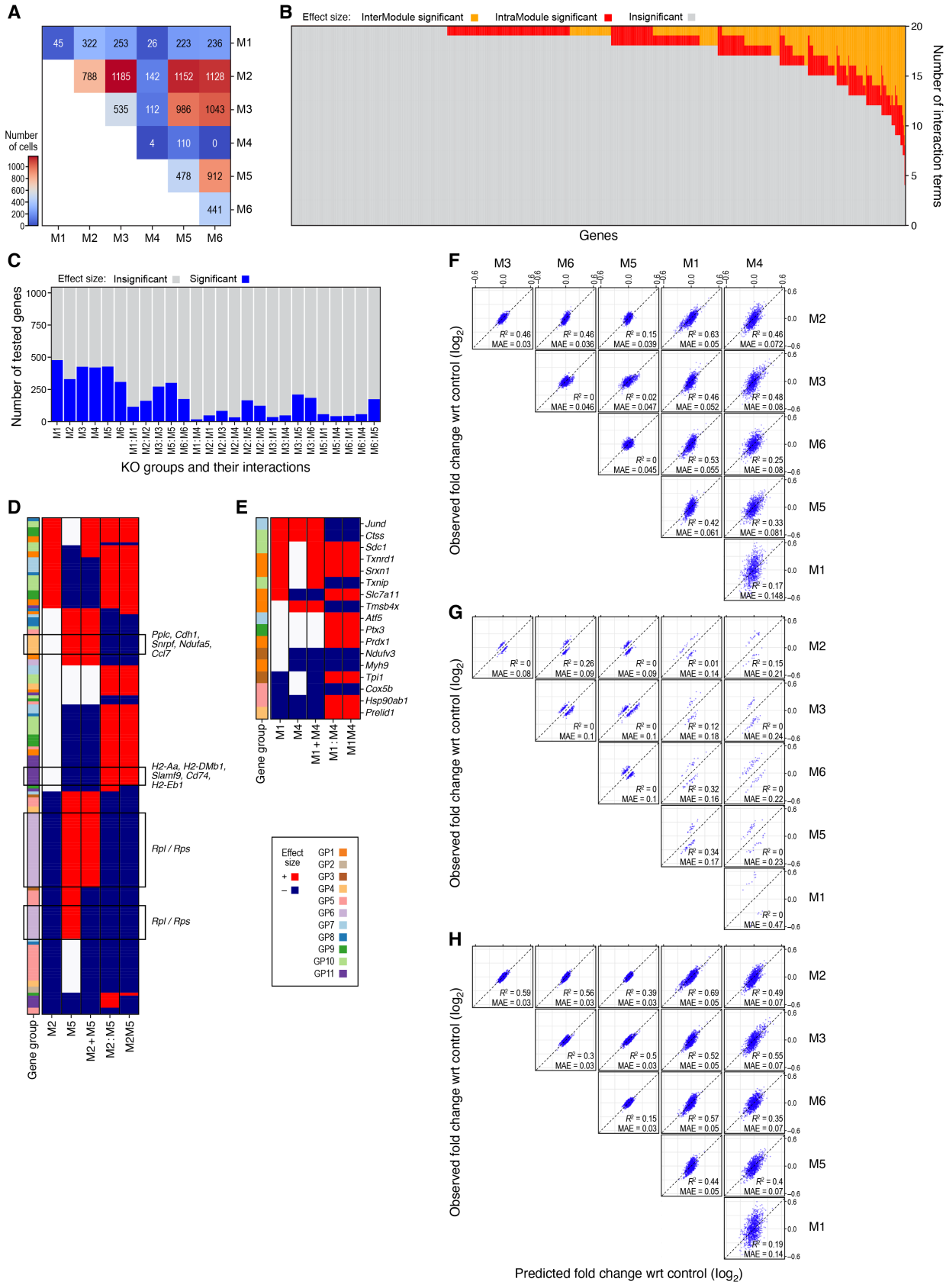
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

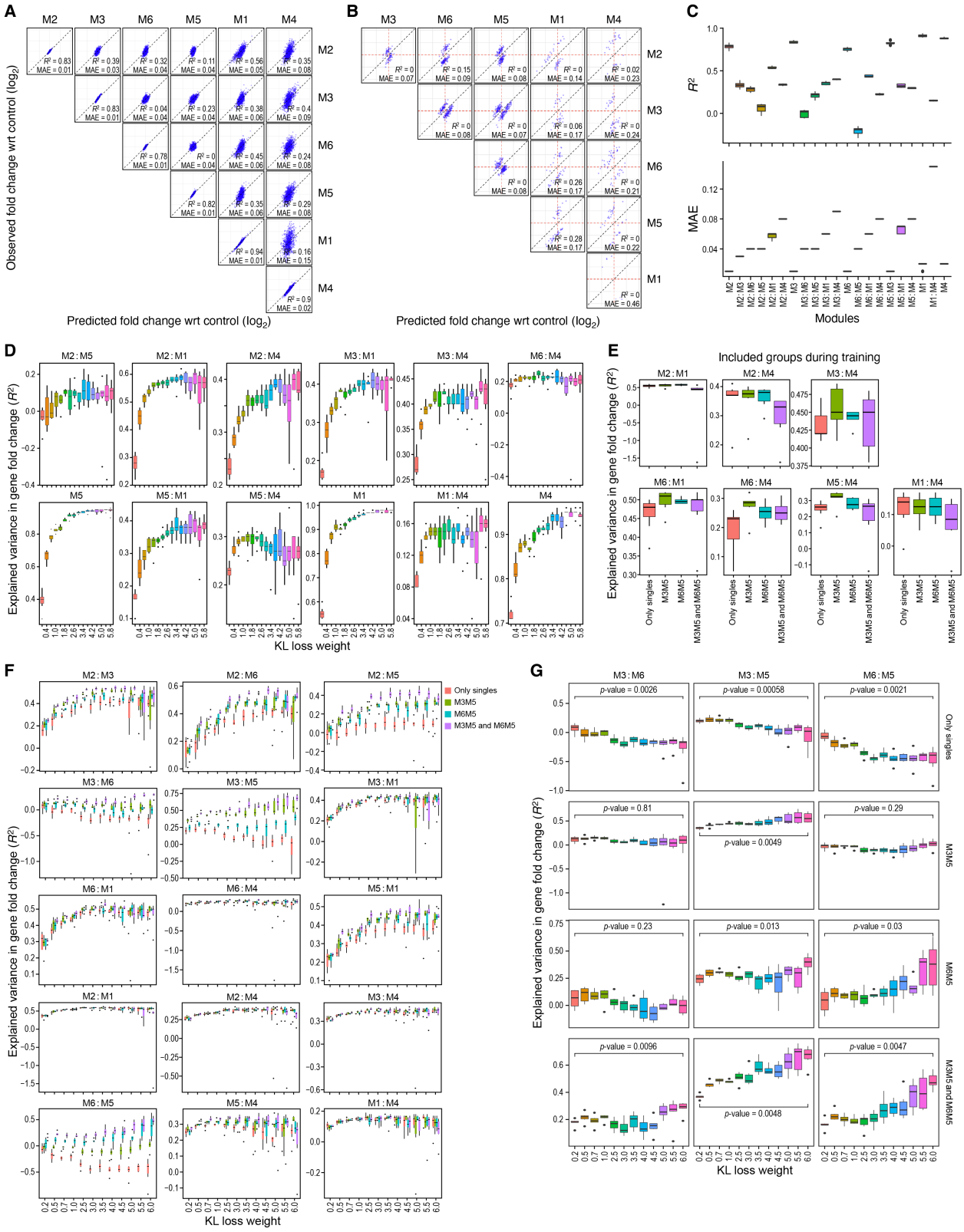
# Supp. Fig. 5



bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Supp. Fig. 6



bioRxiv preprint doi: https://doi.org/10.1101/2023.01.23.525198; this version posted January 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Supp. Fig. 7