Joint profiling of chromatin accessibility and gene expression in thousands of single cells

Junyue Cao[1,2], Darren A. Cusanovich[1]*†, Vijay Ramani[1]*, Delasa Aghamirzaie[1], Hannah A. Pliner[1], Andrew J. Hill[1], Riza M. Daza[1], Jose L. McFaline-Figueroa[1], Jonathan S. Packer[1], Lena Christiansen[3], Frank J. Steemers[3], Andrew C. Adey[4,5], Cole Trapnell[1,6,7]‡, Jay Shendure[1,6,7,8]‡

[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [2]Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA. [3]Illumina Inc., San Diego, CA, USA. [4]Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA. [5]Knight Cardiovascular Institute, Portland, OR, USA. [6]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [7]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [8]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.

*These authors contributed equally to this work. †Present address: Asthma & Airway Disease Research Center and Department of Cellular and Molecular Medicine, The University of Arizona, Tucson, AZ, USA. ‡Corresponding author. Email: coletrap@uw.edu (C.T.); shendure@uw.edu (J.S.)

Although we can increasingly measure transcription, chromatin, methylation, etc. at single cell resolution, most assays survey only one aspect of cellular biology. Here we describe sci-CAR, a combinatorial indexing-based co-assay that jointly profiles chromatin accessibility and mRNA in each of thousands of single cells. As a proof-of-concept, we apply sci-CAR to 4,825 cells comprising a time-series of dexamethasone treatment, as well as to 11,296 cells from the adult mouse kidney. With the resulting data, we compare the pseudotemporal dynamics of chromatin accessibility and gene expression, reconstruct the chromatin accessibility profiles of cell types defined by RNA profiles, and link cis-regulatory sites to their target genes on the basis of the covariance of chromatin accessibility and transcription across large numbers of single cells.

The concurrent profiling of multiple classes of molecules, e.g. RNA and DNA, within single cells has the potential to reveal causal regulatory relationships and to enrich the utility of organism-scale single cell atlases. However, to date, nucleic acid 'co-assays' rely on physically isolating each cell, limiting their throughput to a few cells per study (fig. S1A and table S1) (*1–6*).

Single-cell combinatorial indexing ("sci") methods use split-pool barcoding to uniquely label the nucleic acid contents of single cells or nuclei (*7–13*). Here we describe sci-CAR, which jointly profiles single cell chromatin accessibility and mRNA in a scalable fashion. Sci-CAR effectively combines sci-ATAC-seq and sci-RNA-seq into a single protocol (Fig. 1): (i) Nuclei are extracted, with or without fixation, and distributed to wells. (ii) A first RNA-seq 'index' is introduced by in situ reverse transcription (RT) with a poly(T) primer bearing a well-specific barcode and a unique molecular identifier (UMI). (iii) A first ATAC-seq index is introduced by in situ tagmentation with Tn5 transposase bearing a well-specific barcode. (iv) All nuclei are pooled and redistributed by FACS to multiple plates. (v) After second-strand synthesis of cDNA, nuclei in each well are lysed, and the lysate split to RNA and ATAC-dedicated portions. (vi) To provide a second priming site for amplification of 3′ cDNA tags, the RNA-dedicated lysate is subjected to transposition with unindexed Tn5 transposase. 3′ cDNA tags are amplified with primers corresponding to the Tn5 adaptor and RT primer. These primers also bear a well-specific barcode that is the second RNA-seq index. (vii) The ATAC-seq-dedicated lysate is amplified with primers specific to the barcoded Tn5 adaptors from step iii. These primers also bear a well-specific barcode that is the second ATAC-seq index. (viii) Amplicons from RNA-seq and ATAC-seq-dedicated lysates are respectively pooled and sequenced. Each sequence read is associated with two barcodes corresponding to each round of indexing. As with other sci- protocols, most nuclei pass through a unique combination of wells, receiving a unique combination of barcodes that can be used to group reads derived from the same cell. Because the barcodes introduced to RNA-seq and ATAC-seq libraries correspond to specific wells, we can link the mRNA and chromatin accessibility profiles of individual cells.

We applied sci-CAR to a cell culture model of cortisol response, wherein dexamethasone (DEX), a synthetic mimic of cortisol, activates glucocorticoid receptor (GR), which binds to thousands of locations across the genome, altering the expression of hundreds of genes (*14–17*). We collected lung adenocarcinoma-derived A549 cells after 0, 1 or 3 hrs of 100 nM DEX treatment, and performed a 96 x 576 well sci-CAR experiment. The three timepoints were each represented in 24 wells during the first round of indexing, while the remaining 24 wells contained a mixture of HEK293T (human) and NIH3T3 (mouse) cells (fig. S1B).

We obtained sci-RNA-seq profiles for 6,093 cells (median 3,809 UMIs) and sci-ATAC-seq profiles for 6,085 cells

(median 1,456 unique reads) (fig. S1, C to E). For both data types, reads assigned to the same cell overwhelmingly mapped to one species (fig. S1, F and G). We obtained roughly equivalent UMIs per cell from 'RNA-only' plates processed in parallel, albeit at a lower sequencing depth per cell. Aggregated transcriptomes of co-assayed vs. RNA-only plates were well-correlated (r = 0.97-0.98; fig. S2). In contrast, although co-assayed vs. 'ATAC-only plates' were comparable in quality and well-correlated in aggregate (fig. S3), ATAC-only plates had ~10-fold higher complexity. The lower efficiency of the co-assay for ATAC is likely explained by factors including buffer modifications and our use of only half the lysate.

There were 4,825 cells (70% of either set) for which we recovered both transcriptome and chromatin accessibility data. To confirm that paired profiles truly derived from the same cells, we asked whether cells from mixed human-mouse wells were consistently assigned as human or mouse. Indeed, 1,423/1,425 (99%) of co-assayed cells from those wells were assigned the same species label from both sci-RNA-seq and sci-ATAC-seq profiles (Fig. 2A).

We next examined the time course of GR activation. DEX treatment of A549 cells increased both transcription and promoter accessibility of markers of GR activation, including *NFKBIA*, *SCNN1A*, *CKB*, *PER1* and *CDH16 (14, 16)* (fig. S4, A and B). Unsupervised clustering or t-SNE visualization of either sci-RNA-seq or sci-ATAC-seq profiles readily separated clusters corresponding to untreated and DEX-treated cells (Fig. 2, B and C). Reassuringly, cells from co-assay plates and single-assay plates of either type were intermixed (fig. S4C).

88% and 93% of co-assayed cells in clusters 1 and 2 of sci-ATAC-seq data were found in corresponding sci-RNA-seq clusters (fig. S4, D and E). Cells with concordant vs. discordant assignments did not significantly differ in read depth (*P*-value > 0.1, Welch two-sample t-test), but notably fell on the border between clusters 1 and 2 in either t-SNE (Fig. 2D and fig. S4F). While most discordant cells (70%) were from 0 hrs, the remainder tended to derive from 1 hrs rather than 3 hrs (5% of 1 hr vs. 1% of 3 hr cells, *P*-value = 2.2e-16, Fisher's Exact Test). Although we cannot rule out that this is due to imperfect clustering, these discordantly assigned cells potentially reflect transitional states in GR activation.

Differential expression (DE) analysis of sci-RNA-seq data revealed significant changes in 2,613 genes (5% FDR) (table S2). For comparison, a similar analysis with bulk RNA-seq data of DEX treatment in A549 cells at 0 vs. 3 hrs (*18*) identified 870 DE genes, 536 of which were also DE here. Log$_2$ fold changes were well-correlated between the datasets for DE genes (r = 0.86, fig. S4G).

Differential accessibility (DA) analysis of sci-ATAC-seq profiles identified significant changes at 4,763 sites (5% FDR) (table S3). For comparison, a similar analysis of bulk DNase-seq data from DEX-treated A549 cells at 0 vs. 3 hrs (*18*)

identified 672 DA sites, 544 of were also DA here. Log$_2$ fold changes were well-correlated between the datasets for DA sites (rho = 0.68, fig. S4H).

Of our DA sites, 701 (15%) were promoters, of which 175 overlapped with DE transcripts. Transcripts for genes with DA promoters that were not DE were detected in significantly fewer cells than genes with DA promoters that were DE (median 10% vs. 25%, *P*-value < 5e-5, unpaired two sample permutation test based on 20,000 simulations), suggesting we may be insufficiently powered to detect DE at many genes with DA promoters. For the 175 genes that are both DA and DE, the log$_2$ fold changes were modestly correlated (rho = 0.63, fig. S4I), with 130/175 (74%) exhibiting directional concordance (exact two-sided binomial test, *P*-value = 9e-11).

We ordered cells along a pseudotime trajectory with Monocle (*19*) based on the top 1,000 DE genes (fig. S5A). Cells were ordered consistently with the time course (Fig. 2E). Of note, the aforementioned cells from 1 hour whose cluster assignments were discordant (Fig. 2D and fig. S4F) occurred significantly earlier in pseudotime than cells with concordant assignments (*P*-value = 3e-5, Wilcoxon rank sum test, fig. S5B). Of the 2,613 DE genes, 979 (37%) increased and 1,111 (43%) decreased in expression along pseudotime, while 523 (20%) exhibited transient changes (fig. S5, C and D, and tables S2 and S4). We exploited the co-assay to examine the dynamics of chromatin accessibility across RNA-defined pseudotime, identifying opening (47%), closing (32%) and transient (21%) DA sites (fig. S5E and tables S3 and S5). There were eleven genes that showed significant changes in *both* gene expression and promoter accessibility along pseudotime (5% FDR for both), with well-correlated dynamics (Fig. 2F and fig. S5, F to H).

We converted the (cell x site) matrix to a (cell x transcription factor (TF) motif) matrix, simply by counting occurrences of each motif in all accessible sites for each cell (*20*). The motifs of 91/399 (23%) of expressed TFs were DA across the treatment conditions (5% FDR) (tables S6 and S7). Where ChIP-seq data was available for the same time course (*18*), we observed consistent dynamics of increasing motif-associated accessibility (fig. S6A) and TF binding to accessible sites (fig. S6B). Motif accessibility dynamics across expression-defined pseudotime are summarized in fig. S6C. The motif of the canonical glucocorticoid receptor *NR3C1* was the most activated, even though its expression decreased (Fig. 2G), consistent with its activation by recruitment from the cytosol rather than by increased expression. In contrast, *KLF9* is a direct target of GR activation via a feed forward loop (*21*). Consistent with this, we observe that both its expression and its motif accessibility increase along pseudotime (Fig. 2G and fig. S6, D and E).

Single-cell RNA sequencing studies have recently characterized the transcriptomes of diverse cell types represented

in the mammalian kidney (*22–24*). However, little is known about the epigenetic landscapes that underlie these cell type-specific gene expression programs. To investigate this, we isolated and fixed nuclei from whole kidneys of two 8-week male mice (fig. S7A). From one sci-CAR experiment, we obtained sci-RNA-seq profiles for 13,893 nuclei (median 1,011 UMIs; fig. S7B) and sci-ATAC-seq profiles for 13,395 nuclei (median 7,987 unique reads; fig. S7C). There were 11,296 cells for which we recovered both transcriptome and chromatin accessibility profiles.

We compared sci-CAR transcriptomes with a recently published single cell RNA-seq dataset of the same tissue generated by Drop-seq (*24*). After correcting for gene length biases (Drop-seq is biased toward shorter transcripts, and sci-RNA-seq toward longer transcripts) aggregated transcriptomes were reasonably well correlated (r = 0.73, fig. S7D). Semi-supervised clustering of 10,727 sci-CAR transcriptomes (>500 UMIs) identified 14 groups, ranging in size from 74 (0.7%) to 2,358 (22.0%) cells (Fig. 3A and fig. S7, E and F). Established markers identified nearly all cell types (fig. S8, A and B). The expression profiles of proximal tubule cells separate them into three subtypes including S1/S2 cells (*Slc5a12+*, *Gatm+*, *Alpl+*, *Slc34a1+*), S3 type 1 cells (*Slc34a1+*, *Atp11a+*), and S3 type 2 cells (*Atp11a+*, *Rnf24+*) (fig. S8C) (*25*, *26*). The smallest cluster is positive for cell cycle progression markers (*Mki67* and *Cenpp*), and may represent an actively proliferating subpopulation (fig. S8D) (*25*, *26*). Cell type proportions were well-correlated between replicate kidneys, with the exception of paranephric body adipocytes (1.2% vs. 0.4%), likely due to technical variation in kidney dissection as these reside superficial to the renal fascia (fig. S7E).

We identified 8,774 genes that were DE across the 14 cell types (5% FDR), including 1,771 with >2-fold greater expression in the highest vs. second highest cell type (fig. S9, A and B, and tables S8 and S9). New marker genes were identified, such as *Daam2* for renal pericytes and *Calcr* for collecting duct intercalated cell B (fig. S9, C and D) (*25*, *26*). We examined expression of solute carrier transporters (SLCs), as these correspond to a principal function of the kidney. 208/345 (60%) of these were DE in subsets of renal tubule cell types, many corresponding to known and potentially novel reabsorption specificities (Fig. 3B, fig. S9E, and table S10).

We compared aggregated sci-CAR chromatin accessibility profiles with published bulk ATAC-seq data on adult mouse kidney (*18*), and found them to be reasonably well correlated (r = 0.75; fig. S10, A and B). Across all genes, aggregate promoter accessibility correlated with aggregate gene expression (rho = 0.26; fig. S10C). Nonetheless, a significant challenge for single cell ATAC-seq data, relative to single cell RNA-seq data, is the sparsity of the resulting matrices (*8*). Thus, our initial efforts to cluster co-assayed cells based solely on their ATAC-seq profiles failed to discover the expected diversity of

cell types. We therefore sought to leverage the co-assay aspect of these data to recover the chromatin landscapes of individual cell types.

As a first approach, we simply annotated cell types from transcriptional profiles for ~96% of the 11,296 cells that were successfully co-assayed. We then aggregated ATAC-seq signal for each cell type separately, followed by peak calling (*27*). As a second approach, we also developed an algorithm to combine the ATAC-seq profiles of cells with highly similar RNA-seq profiles prior to clustering (fig. S7A). For cells from each RNA-seq-defined cell type, we identified subsets of cells with highly similar expression profiles (a mean of 50 cells assigned to each of 222 'pseudo-cells'). We then aggregated the ATAC-seq profiles of each pseudo-cell, and performed t-SNE on these. In contrast with single-cell ATAC-seq data, pseudo-cell chromatin accessibility profiles corresponding to the same cell types clustered together (Fig. 3C). Overall, these analyses illustrate how co-assay data can be leveraged to overcome the relative sparsity of single cell ATAC-seq data and define chromatin accessibility profiles even for closely related cell types.

We identified 22,026 DA sites across the 14 mouse kidney cell types, including 2,096 promoters and 19,930 distal sites (5% FDR; Fig. 3D; fig. S10, D and E; and tables S11 and S12). In some cases, DA at a gene's promoter was concordant with DE (fig. S11, A and B), but this was the exception rather than the rule. Out of 2,096 genes with a DA promoter in at least one cell type, 132 genes were also DE (1% FDR) with a >2-fold difference between the first and second ranked cell type. Although promoter accessibility and expression of these genes across cell types are positively correlated (median rho = 0.17), the majority (112/132 or 85%) exhibited maximal promoter accessibility and gene expression in different cell types (fig. S11C). The relatively weaker correlation compared with what we observed in the A549 dexamethasone time series (rho = 0.63; fig. S4I) is potentially a consequence of the fact that in the A549 cells, we were comparing changes in promoter accessibility vs. expression, whereas here we are comparing absolute enrichment of accessibility at promoters vs. expression.

We sought to link distal *cis*-regulatory elements to their target genes based on the covariance of chromatin accessibility and gene expression across large numbers of co-assayed cells. As the sparsity of our single cell profiles makes this challenging, we worked with the aforedescribed 222 pseudo-cells (fig. S12A). For each gene, we computed correlations between its expression and the adjusted accessibility of all sites within 100 kilobases (kb) of its transcriptional start site (TSS) using LASSO (least absolute shrinkage and selection operator).

Within the top 2,000 DE genes (ranked by q-value), we linked 1,260 distal sites to 321 genes (median 3 sites per gene, out of median 19 sites within 100 kb of TSS tested; fig. S12, B and C, and table S13). 44% of sites were linked to the nearest

TSS, and 21% to the second nearest TSS (fig. S12D). Distal site-gene linkages were significantly closer than all possible pairs tested (mean 41 kb for links vs. 48 kb for all pairs tested; *P*-value < 5e-5, unpaired permutation test based on 20,000 simulations; fig. S12E).

To evaluate the possibility that the links were artifacts of regularized regression, we permuted the sample IDs of the chromatin accessibility matrix and performed the same analysis. After this permutation, only 4 links were identified (fig. S12B). To control for correlations between closely located accessible sites in the genome, we separately permuted the peak IDs. This yielded 216 links, or just 17% as many links as without permutation (fig. S12B).

The 321 genes with linked distal sites were specifically expressed in a variety of cell types (fig. S12F). For example, the link with the highest correlation is between distal convoluted tubule cell marker gene Slc12a3 and a site 36 kb downstream of its TSS and overlapping its last exon (fig. S13). The accessibility of this linked site was modestly more specific to distal convoluted tubule cells than the Slc12a3 promoter. In contrast, the accessible site closest to the Slc12a3 promoter (only 216 bp away) was not linked to the Slc12a3 promoter by our approach, nor is its accessibility specific to distal convoluted tubule cells. Similarly, a marker gene for Loop of Henle cells, Slc12a1, is linked to two distal sites (fig. S14), both of which exhibit accessibility specific to Loop of Henle cells. In contrast, the nearest accessible site (9 kb from the TSS), which was not linked, does not exhibit this specificity.

Links between distal *cis*-regulatory elements and their target genes can be useful for explaining differential expression across cell types. For example, the cell type-specific expression of Slc6a18, a marker gene for type 2 proximal tubule S3 cells, is not mirrored by cell type-specific promoter accessibility (fig. S11C). However, from our covariance approach, its TSS is linked to a site 16 kb away whose accessibility is correlated with Slc6a18 expression (Fig. 4A). To quantify the utility of the links between distal *cis*-regulatory elements and their target genes identified from sci-CAR data, we constructed a linear regression model to predict gene expression differences based on chromatin accessibility at promoters only vs. promoters together with linked distal sites. Including linked distal sites improved predictions by four-fold (*P*-value < 5e-5, paired permutation test based on 20,000 simulations; Fig. 4B).

Our analyses illustrate the advantages of a single cell co-assay over assays that solely profile transcription or chromatin accessibility. Sci-CAR is compatible with fresh or fixed nuclei, and like other sci-seq techniques, can encode multiple samples per experiment. Its throughput can potentially be increased by additional rounds of split-pool indexing (*13*). With 384 x 384 x 384 sci-CAR, one could potentially co-assay millions of single cells per experiment. A limitation of sci-CAR is

the sparsity of the resulting data, particularly with respect to chromatin accessibility. This can potentially be overcome in the future through protocol optimizations, particularly of crosslinking conditions. A second limitation is that although we were able to link distal elements and target genes on the basis of covariance of accessibility and expression, these data remain correlative and involve a minority of DE genes and DA elements.

Notwithstanding these limitations, sci-CAR expands the potential of combinatorial indexing for scalably profiling single cell molecular phenotypes, and may be particularly useful in the context of organism-scale single cell atlases. With further development, we anticipate that additional DNA/RNA co-assays may be realized by simply integrating other sci-seq protocols together with sci-RNA-seq (e.g. methylation + transcripts; chromosome conformation + transcripts; DNA sequence + transcripts) (*8–13*). A longer-term goal is to adapt single cell combinatorial indexing to span the Central Dogma, such that aspects of DNA, RNA and protein species can be concurrently assayed from each of many single cells.

REFERENCES AND NOTES

1. S. J. Clark, R. Argelaguet, C.-A. Kapourani, T. M. Stubbs, H. J. Lee, C. Alda-Catalinas, F. Krueger, G. Sanguinetti, G. Kelsey, J. C. Marioni, O. Stegle, W. Reik, scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018). doi:10.1038/s41467-018-03149-4 Medline

2. C. Angermueller, S. J. Clark, H. J. Lee, I. C. Macaulay, M. J. Teng, T. X. Hu, F. Krueger, S. Smallwood, C. P. Ponting, T. Voet, G. Kelsey, O. Stegle, W. Reik, Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016). doi:10.1038/nmeth.3728 Medline

3. Y. Hou, H. Guo, C. Cao, X. Li, B. Hu, P. Zhu, X. Wu, L. Wen, F. Tang, Y. Huang, J. Peng, Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* **26**, 304–319 (2016). doi:10.1038/cr.2016.23 Medline

4. Y. Hu, K. Huang, Q. An, G. Du, G. Hu, J. Xue, X. Zhu, C.-Y. Wang, Z. Xue, G. Fan, Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol.* **17**, 88 (2016). doi:10.1186/s13059-016-0950-z Medline

5. S. Pott, Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* **6**, e23203 (2017). doi:10.7554/eLife.23203 Medline

6. F. Guo, L. Li, J. Li, X. Wu, B. Hu, P. Zhu, L. Wen, F. Tang, Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res.* **27**, 967–988 (2017). doi:10.1038/cr.2017.82 Medline

7. S. Amini, D. Pushkarev, L. Christiansen, E. Kostem, T. Royce, C. Turk, N. Pignatelli, A. Adey, J. O. Kitzman, K. Vijayan, M. Ronaghi, J. Shendure, K. L. Gunderson, F. J. Steemers, Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat. Genet.* **46**, 1343–1349 (2014). doi:10.1038/ng.3119 Medline

8. D. A. Cusanovich, R. Daza, A. Adey, H. A. Pliner, L. Christiansen, K. L. Gunderson, F. J. Steemers, C. Trapnell, J. Shendure, Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015). doi:10.1126/science.aab1601 Medline

9. V. Ramani, X. Deng, R. Qiu, K. L. Gunderson, F. J. Steemers, C. M. Disteche, W. S. Noble, Z. Duan, J. Shendure, Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017). doi:10.1038/nmeth.4155 Medline

10. Y. Yin, Y. Jiang, J. B. Berletch, C. M. Disteche, W. S. Noble, F. J. Steemers, A. C. Adey, J. A. Shendure, High-throughput mapping of meiotic crossover and chromosome mis-segregation events in interspecific hybrid mice. bioRxiv 338053 [Preprint]. 12 June 2018. https://doi.org/10.1101/338053

11. S. A. Vitak, K. A. Torkenczy, J. L. Rosenkrantz, A. J. Fields, L. Christiansen, M. H. Wong, L. Carbone, F. J. Steemers, A. Adey, Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat. Methods* **14**, 302–308 (2017). doi:10.1038/nmeth.4154 Medline

12. R. M. Mulqueen, D. Pokholok, S. J. Norberg, K. A. Torkenczy, A. J. Fields, D. Sun, J. R. Sinnamon, J. Shendure, C. Trapnell, B. J. O'Roak, Z. Xia, F. J. Steemers, A. C. Adey, Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* **36**, 428–431 (2018). doi:10.1038/nbt.4112 Medline

13. J. Cao, J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, J. Shendure, Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017). doi:10.1126/science.aam8940 Medline

14. T. E. Reddy, F. Pauli, R. O. Sprouse, N. F. Neff, K. M. Newberry, M. J. Garabedian, R. M. Myers, Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res.* **19**, 2163–2171 (2009). doi:10.1101/gr.097022.109 Medline

15. S. John, P. J. Sabo, R. E. Thurman, M.-H. Sung, S. C. Biddie, T. A. Johnson, G. L. Hager, J. A. Stamatoyannopoulos, Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264–268 (2011). doi:10.1038/ng.759 Medline

16. T. E. Reddy, J. Gertz, G. E. Crawford, M. J. Garabedian, R. M. Myers, The hypersensitive glucocorticoid response specifically regulates period 1 and expression of circadian genes. *Mol. Cell. Biol.* **32**, 3756–3767 (2012). doi:10.1128/MCB.00062-12 Medline

17. C. M. Vockley, A. M. D'Ippolito, I. C. McDowell, W. H. Majoros, A. Safi, L. Song, G. E. Crawford, T. E. Reddy, Direct GR binding sites potentiate clusters of TF binding across the human genome. *Cell* **166**, 1269–1281.e19 (2016). doi:10.1016/j.cell.2016.07.049 Medline

18. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi:10.1038/nature11247 Medline

19. X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner, C. Trapnell, Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017). doi:10.1038/nmeth.4402 Medline

20. J. D. Buenrostro, B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzales, M. P. Snyder, H. Y. Chang, W. J. Greenleaf, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015). doi:10.1038/nature14590 Medline

21. Y. Chinenov, M. Coppo, R. Gupte, M. A. Sacta, I. Rogatsky, Glucocorticoid receptor coordinates transcription factor-dominated regulatory network in macrophages. *BMC Genomics* **15**, 656 (2014). doi:10.1186/1471-2164-15-656 Medline

22. L. Chen, J. W. Lee, C.-L. Chou, A. V. Nair, M. A. Battistone, T. G. Păunescu, M. Merkulova, S. Breton, J. W. Verlander, S. M. Wall, D. Brown, M. B. Burg, M. A. Knepper, Transcriptomes of major renal collecting duct cell types in mouse identified by single-cell RNA-seq. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9989–E9998 (2017). doi:10.1073/pnas.1710964114 Medline

23. X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, D. Huang, Y. Xu, W. Huang, M. Jiang, X. Jiang, J. Mao, Y. Chen, C. Lu, J. Xie, Q. Fang, Y. Wang, R. Yue, T. Li, H. Huang, S. H. Orkin, G.-C. Yuan, M. Chen, G. Guo, Mapping the mouse cell atlas by Microwell-seq. *Cell* **172**, 1091–1107.e17 (2018). doi:10.1016/j.cell.2018.02.001 Medline

24. J. Park, R. Shrestha, C. Qiu, A. Kondo, S. Huang, M. Werth, M. Li, J. Barasch, K. Susztak, Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758–763 (2018). doi:10.1126/science.aar2131 Medline

25. The Human Protein Atlas; www.proteinatlas.org.

26. M. Uhlen, C. Zhang, S. Lee, E. Sjöstedt, L. Fagerberg, G. Bidkhori, R. Benfeitas, M. Arif, Z. Liu, F. Edfors, K. Sanli, K. von Feilitzen, P. Oksvold, E. Lundberg, S. Hober, P. Nilsson, J. Mattsson, J. M. Schwenk, H. Brunnström, B. Glimelius, T. Sjöblom, P.-H. Edqvist, D. Djureinovic, P. Micke, C. Lindskog, A. Mardinoglu, F. Ponten, A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017). doi:10.1126/science.aan2507 Medline

27. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). doi:10.1186/gb-2008-9-9-r137 Medline

28. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013). doi:10.1038/nmeth.2688 Medline

29. D. A. Cusanovich, J. P. Reddington, D. A. Garfield, R. M. Daza, D. Aghamirzaie, R. Marco-Ferreres, H. A. Pliner, L. Christiansen, X. Qiu, F. J. Steemers, C. Trapnell, J. Shendure, E. E. M. Furlong, The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018). doi:10.1038/nature25981 Medline

30. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi:10.1093/bioinformatics/bts635 Medline

31. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi:10.1093/bioinformatics/btp352 Medline

32. H. A. Pliner, J. S. Packer, J. L. McFaline-Figueroa, D. A. Cusanovich, R. M. Daza, D. Aghamirzaie, S. Srivatsan, X. Qiu, D. Jackson, A. Minkina, A. C. Adey, F. J. Steemers, J. Shendure, C. Trapnell, Cicero predicts cis-regulatory DNA Interactions from single-cell chromatin accessibility data. *Mol. Cell* S1097-2765(18)30547-1 (2018). doi:10.1016/j.molcel.2018.06.044 Medline

33. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi:10.1093/bioinformatics/btq033 Medline

34. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015). doi:10.1093/bioinformatics/btu638 Medline

35. R. Gentleman, V. Carey, W. Huber, F. Hahne, genefilter: genefilter: methods for filtering genes from high-throughput experiments (2017).

36. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011). doi:10.1093/bioinformatics/btr064 Medline

37. M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F.-Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, T. R. Hughes, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014). doi:10.1016/j.cell.2014.08.009 Medline

38. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018). doi:10.1186/s13059-017-1382-0 Medline

39. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi:10.1186/s13059-014-0550-8 Medline

40. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010). doi:10.18637/jss.v033.i01 Medline

41. The Human Protein Atlas, Cenpp; www.proteinatlas.org/ENSG00000188312-CENPP/tissue/kidney#img.

42. The Human Protein Atlas, Pdgfrb; www.proteinatlas.org/ENSG00000113721-PDGFRB/tissue/kidney#img.

43. The Human Protein Atlas, Slc5a12; www.proteinatlas.org/ENSG00000148942-SLC5A12/tissue/kidney#img.

44. The Human Protein Atlas, GATM; www.proteinatlas.org/ENSG00000171766-GATM/tissue/kidney#img.

45. The Human Protein Atlas, Alpl; www.proteinatlas.org/ENSG00000162551-ALPL/tissue/kidney#img.

46. The Human Protein Atlas, Slc34a1; www.proteinatlas.org/ENSG00000131183-SLC34A1/tissue/kidney#img.

47. The Human Protein Atlas, Atp11a; www.proteinatlas.org/ENSG00000068650-ATP11A/tissue/kidney#img.

48. The Human Protein Atlas, Rnf24; www.proteinatlas.org/ENSG00000101236-RNF24/tissue/kidney#img.

49. The Human Protein Atlas, Mki67; www.proteinatlas.org/ENSG00000148773-

MKI67/tissue/kidney#img.

50. The Human Protein Atlas, Kit; www.proteinatlas.org/ENSG00000157404-KIT/tissue/kidney#img.

51. The Human Protein Atlas, Daam2; www.proteinatlas.org/ENSG00000146122-DAAM2/tissue/kidney#img.

52. The Human Protein Atlas, Calcr; www.proteinatlas.org/ENSG00000004948-CALCR/tissue/kidney#img.

53. X. Zhou, T. Wang, Using the Wash U Epigenome Browser to examine genome-wide sequencing data. *Curr. Protoc. Bioinformatics* **40**, 10.10.1–10.10.14 (2012). doi:10.1002/0471250953.bi1010s40

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aau0730/DC1
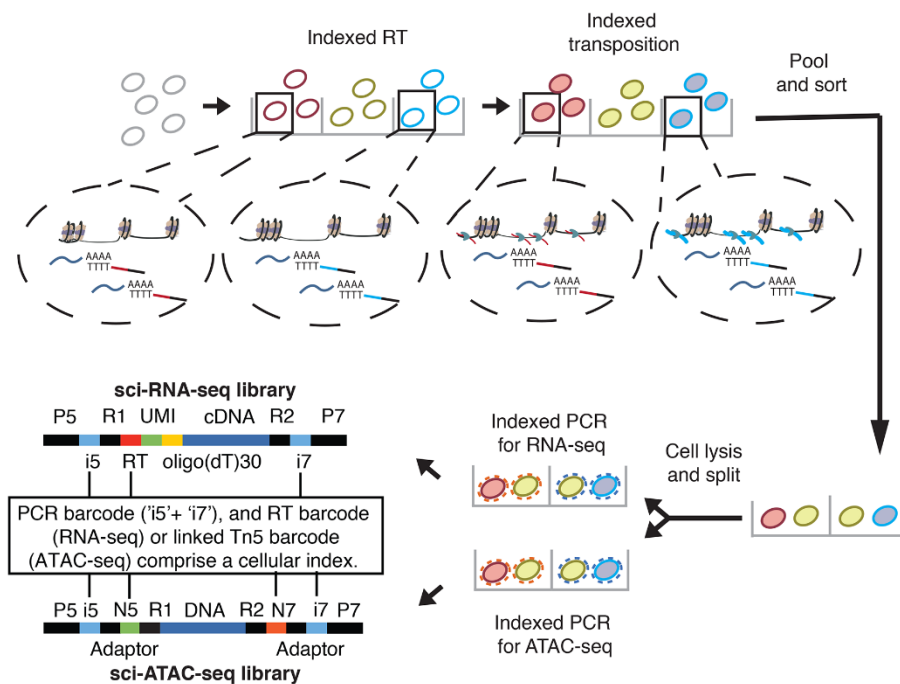Materials and Methods
Figs. S1 to S14
References (*28–53*)
Tables S1 to S13
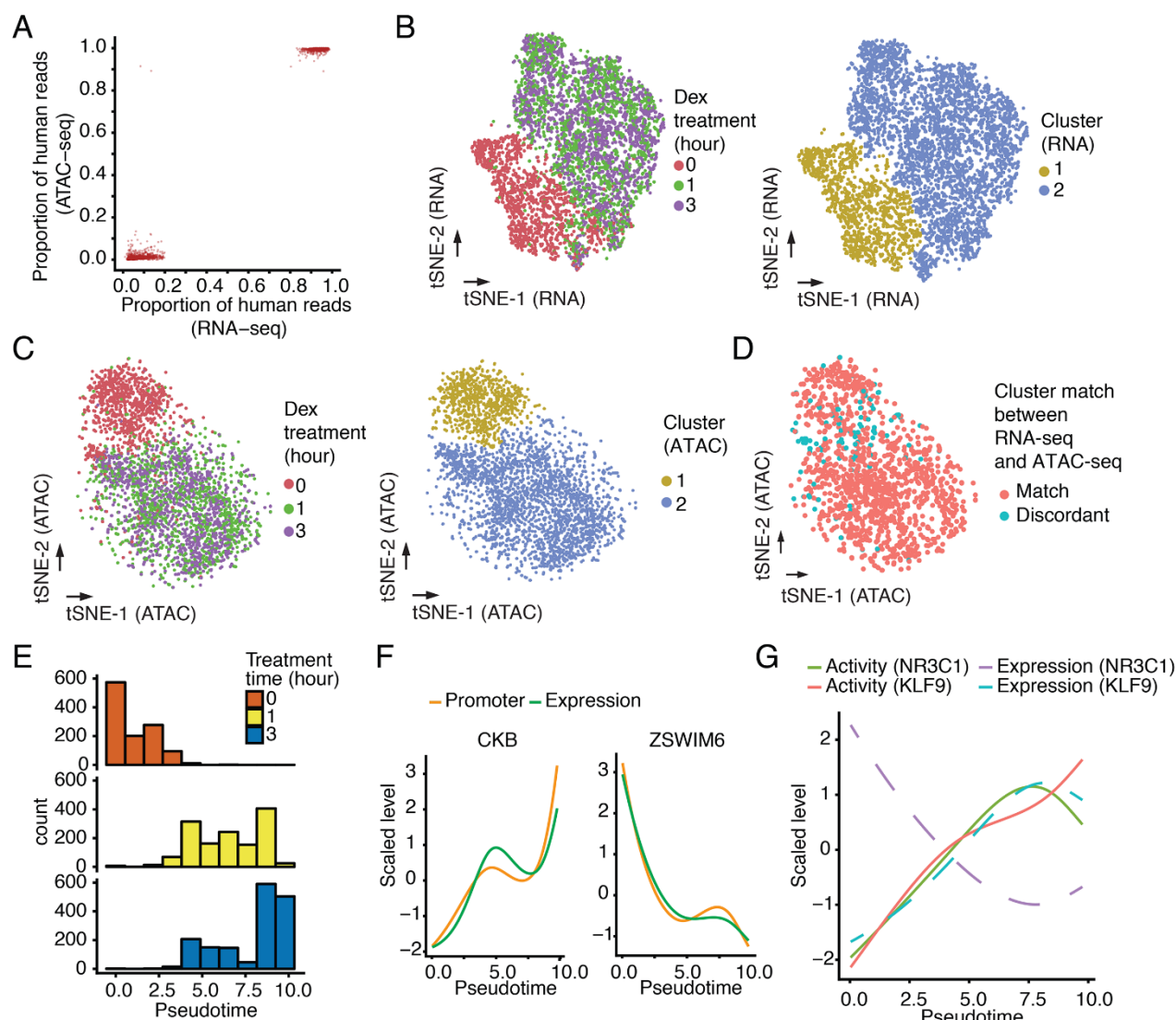
3 May 2018; accepted 22 August 2018
Published online 30 August 2018
10.1126/science.aau0730

Fig. 1. sci-CAR workflow. Key steps outlined in text. RNA-seq: index2 and read1 cover the i5 index, UMI and RT barcode; index1 and read2 cover the i7 index and cDNA fragment. ATAC-seq: read1 and read2 cover genomic DNA sequence. Index 1 and index 2 cover the Tn5 and PCR barcodes.

Fig. 2. Joint profiling of chromatin accessibility and transcription in dexamethasone treated A549 cells. (**A**) Scatter plot showing the proportion of human reads, out of all reads mapping uniquely to the human or mouse reference genomes, for cells in which both RNA-seq profiles and ATAC-seq profiles were obtained. Only HEK293T (human) and NIH/3T3 (mouse) cells are plotted. (**B**) t-SNE visualization of A549 cells (RNA-seq) including cells from both sci-CAR and sci-RNA-seq-only plates, colored by DEX treatment time (left) or unsupervised clustering id (right). (**C**) t-SNE visualization of A549 cells (ATAC-seq) including cells from both sci-CAR and sci-ATAC-seq-only plates, colored by DEX treatment time (left) or unsupervised clustering id (right). (**D**) t-SNE visualization of A549 cells (ATAC-seq) with linked RNA-seq profiles. If the cell is in cluster 1 (or cluster 2) in both RNA-seq and ATAC-seq, then it is labeled as "Match", otherwise it is labeled "Discordant". (**E**) Distribution of cells from different DEX treatment timepoints in gene expression pseudotime inferred by trajectory analysis. (**F**) Smoothed line plot showing scaled (with the R function scale) gene expression and promoter accessibility of *CKB* and *ZSWIM6* across pseudotime. Unscaled, unsmoothed data shown in fig. S5, F and G. (**G**) Smoothed line plot showing the scaled mRNA level and activity change of transcription factors *NR3C1* and *KLF9* across pseudotime. Unscaled, unsmoothed data shown in fig. S6, D and E.
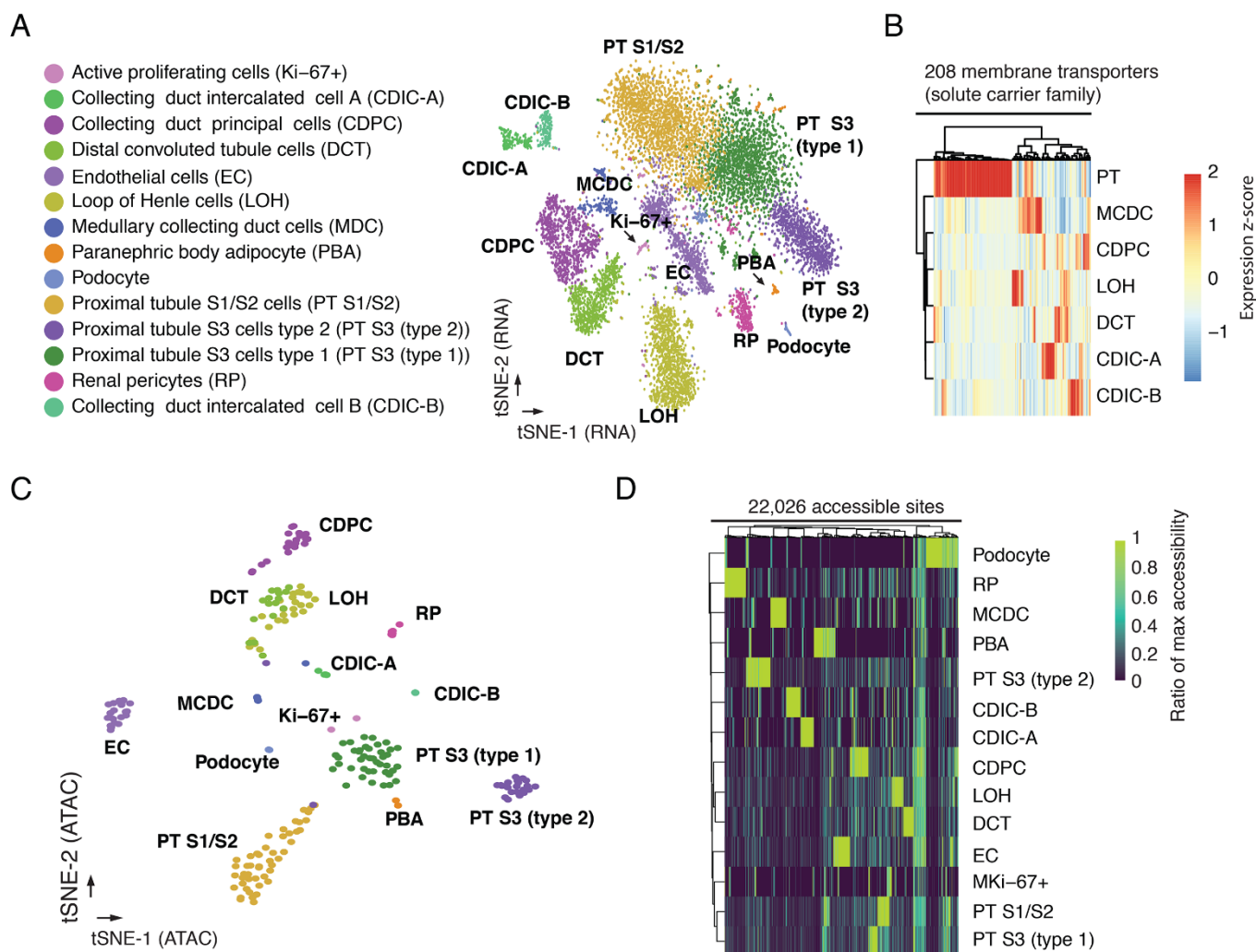
Fig. 3. sci-CAR enables joint profiling of chromatin accessibility and transcription in mouse kidney. (A) t-SNE visualization of mouse kidney nuclei (RNA-seq). Cell types are assigned based on established marker genes. (B) Heatmap showing the relative expression of genes from the solute carrier group of membrane transport proteins in consensus transcriptomes of each cell type estimated by RNA-seq data from the co-assay. The raw expression data (UMI count matrix) was log-transformed, column centered and scaled (using the R function scale), and the resulting values clamped to [-2, 2]. (C) t-SNE visualization of mouse kidney nuclei (ATAC-seq) after aggregating cells with highly similar transcriptomes ('pseudocells'), colored by cell types identified from RNA-seq. (D) Heatmap showing the relative chromatin accessibility of cell type-specific sites for each cell type estimated by ATAC-seq data from the co-assay. The raw aggregated ATAC-seq data (read count matrix) was normalized first by the total number of reads for each cell type then by the maximum accessibility score across all cell types.
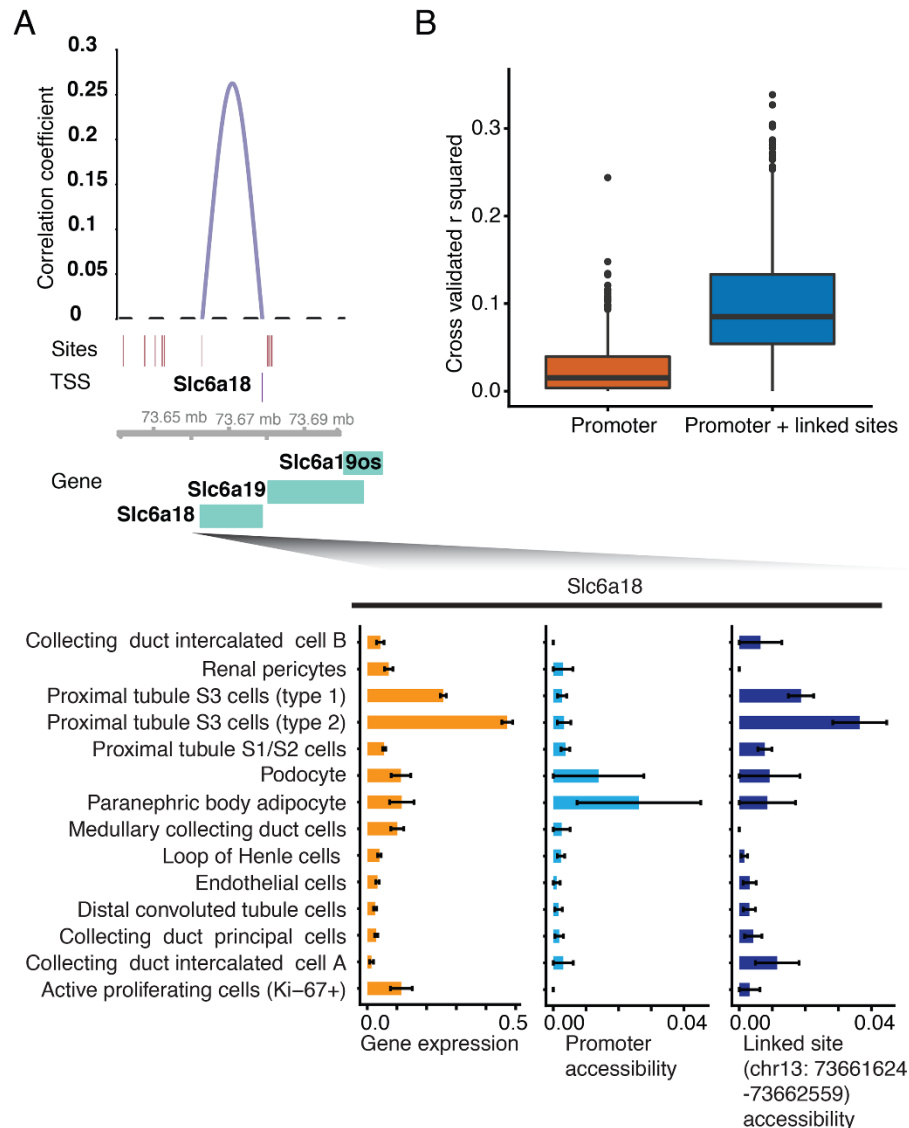
Fig. 4. Linking cis-regulatory elements to regulated genes based on covariance in single cell co-assay data. (A) Top: genome browser plot showing links between accessible distal regulatory sites and the gene Slc6a18. The height corresponds to the correlation coefficient. Bottom: barplots showing the average expression, promoter accessibility and linked site accessibility for cell type-specific marker gene Slc6a18 across different cell types. Gene expression values for each cell were calculated by dividing the raw UMI count by cell-specific size factors. Site accessibilities for each cell were calculated by dividing the raw read count by cell-specific size factors. Error bars represent standard errors of the means. (B) Two linear regression models were built to predict gene expression differences between cell types. The first model predicts changes on the basis of promoter accessibility alone. The second model predicts changes based on the chromatin accessibility of the promoter and distal sites that are linked to it. The boxplot shows the cross-validated r-squared calculated for each gene from the two models.

Science

Joint profiling of chromatin accessibility and gene expression in thousands of single cells

Junyue Cao, Darren A. Cusanovich, Vijay Ramani, Delasa Aghamirzaie, Hannah A. Pliner, Andrew J. Hill, Riza M. Daza, Jose L. McFaline-Figueroa, Jonathan S. Packer, Lena Christiansen, Frank J. Steemers, Andrew C. Adey, Cole Trapnell and Jay Shendure

ARTICLE TOOLS	http://science.sciencemag.org/content/early/2018/08/29/science.aau0730
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/08/29/science.aau0730.DC1
REFERENCES	This article cites 39 articles, 8 of which you can access for free http://science.sciencemag.org/content/early/2018/08/29/science.aau0730#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service