Synergistic and tunable human gene activation by combinations of synthetic transcription factors

Pablo Perez-Pinera[1], David G Ousterout[1], Jonathan M Brunger[1,2], Alicia M Farin[3], Katherine A Glass[1,2], Farshid Guilak[1,2], Gregory E Crawford[4,5], Alexander J Hartemink[4,6] & Charles A Gersbach[1,4]

Mammalian genes are regulated by the cooperative and synergistic actions of many transcription factors. In this study we recapitulate this complex regulation in human cells by targeting endogenous gene promoters, including regions of closed chromatin upstream of silenced genes, with combinations of engineered transcription activator–like effectors (TALEs). These combinations of TALE transcription factors induced substantial gene activation and allowed tuning of gene expression levels that will broadly enable synthetic biology, gene therapy and biotechnology.

Synthetic biology aims to study the control of gene expression by constructing gene regulation systems from the 'bottom up' to better understand natural biological systems and develop useful tools for biotechnology[1]. Despite many substantial accomplishments, this field has been limited primarily to studying artificial promoter-transgene systems with one or two transactivators, typically in microorganisms[1–6]. In contrast, the natural regulation of mammalian gene expression is extraordinarily complex and typically achieved through the combinatorial control of each gene by many regulatory factors. This level of complexity has not yet been achieved in synthetic gene regulation systems and has not been possible for the regulation of endogenous genes. However, the recent emergence of technologies for engineering transcription activator–like effectors (TALEs) targeted to almost any DNA sequence[7–14] provides a unique opportunity for recapitulating this natural complexity. In the current study, we achieved the combinatorial regulation of endogenous mammalian genes in their natural chromosomal context by engineering several TALE transcription factors (TALE-TFs) to bind nearby sites upstream of the transcription start site (TSS) for a target gene. The composition of these combinations of independent TALE-TFs can be manipulated to control gene activation. Synergistic regulation of gene expression by multiple transcriptional activators is known to occur through simultaneous binding and stabilization of components of the preinitiation complex[15,16]. Building on this model, we activated endogenous genes with combinations of engineered transcription factors and were able to tune gene expression levels by systematically varying these combinations.

Each TALE-TF has two distinct protein domains that carry out individual molecular functions: (i) the repeat variable diresidue region binds to DNA at user-specified sequences[7,8], and (ii) the VP64 effector domain recruits transcriptional machinery[9,10] (**Fig. 1a**). This design permits the rapid construction of synthetic transcription factors[9,11,12]. Several TALE-TFs have been recently reported to regulate native mammalian gene expression[9,10,14,17–20]. However, the levels of gene activation in these studies were modest, and several genes could not be induced (**Supplementary Table 1**). Therefore, there is a clear need for improvements to gene-activation strategies that capitalize on the synthetic TALE-TF technology.

We designed six, seven or eight TALE-TFs targeted to the promoter regions of *IL1RN*, *KLK3* (also known as the prostate-specific antigen gene, *PSA*), *CEACAM5* (also known as *CEA*) and *ERBB2*, which are implicated in immunomodulation, inflammation and cancer (**Supplementary Fig. 1**). The target sites for these TALE-TFs were distributed within 600 bp of the TSS (**Fig. 1b**). We transfected TALE-TF expression plasmids into HEK293T cells either individually or as a combination of all the TALE-TFs targeted to a particular promoter and confirmed the expression of the TALE-TFs by western blot (**Supplementary Fig. 2**). We first measured TALE-TF activity in reporter assays in which luciferase was under the control of the respective gene promoter (**Fig. 1c–f**). Most individual TALE-TFs activated the co-transfected plasmid reporters but only modestly, as in previous studies (**Supplementary Table 1**)[9,10,14,17–20]. However, the delivery of combinations of TALE-TFs led to substantial synergistic effects on gene activation. Notably, the synergistic activation of the plasmid-based reporters was recapitulated in the upregulation of the native genes in their natural chromosomal context as determined by quantitative reverse transcription PCR (qRT-PCR), including increases in mRNA abundance greater than 10,000-fold (**Fig. 1g–j**). Detection of induced protein expression of IL-1ra (encoded by *IL1RN*), KLK3, CEACAM5 and ERBB-2 by ELISA and western blotting validated the functional outcome of the activation

[1]Department of Biomedical Engineering, Duke University, Durham, North Carolina, USA. [2]Department of Orthopaedic Surgery, Duke University Medical Center, Durham, North Carolina, USA. [3]Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA. [4]Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, USA. [5]Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, North Carolina, USA. [6]Department of Computer Science, Duke University, Durham, North Carolina, USA. Correspondence should be addressed to C.A.G. (charles.gersbach@duke.edu).

RECEIVED 26 SEPTEMBER 2012; ACCEPTED 27 DECEMBER 2012; PUBLISHED ONLINE 3 FEBRUARY 2013; DOI:10.1038/NMETH.2361

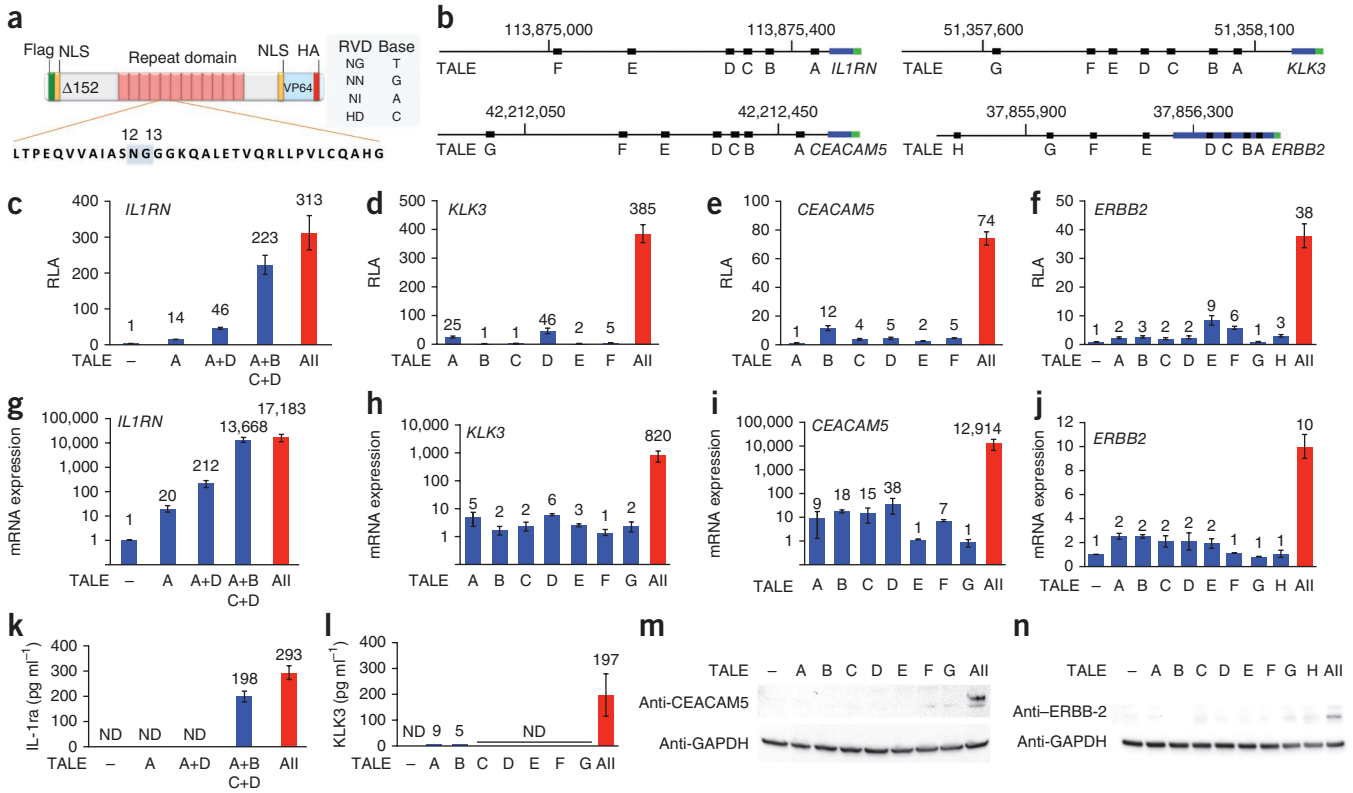

Figure 1 | Synergistic activation of gene expression by combinations of TALE-TFs. (**a**) Structure and sequence of the TALE-TFs in this study. NLS, nuclear localization signal; RVD, repeat variable diresidue; Δ152, 152-residue deletion; HA, hemagglutinin. (**b**) Genomic positions of TALE-TF target sites in *IL1RN*, *KLK3*, *CEACAM5* and *ERBB2* (hg19 coordinates; black boxes). Transcribed and coding regions are shown in blue and green, respectively. (**c**–**f**) Relative luciferase activity (RLA) in promoter reporter assays. (**g**–**n**) Relative mRNA expression measured by quantitative RT-PCR (**g**–**j**) and protein expression assayed by ELISA (**k**,**l**) or western blot (**m**,**n**) for each target gene in human cells transfected with the indicated TALE-TFs. Data for each gene is organized by column. $n = 3$ independent experiments for luciferase reporter assays and quantitative RT-PCR. For ELISA data, $n = 6$ biological replicates for *IL1RN* and $n = 4$ biological replicates for *KLK3*. ND, not detected. Data (**c**–**l**) are shown as the mean ± s.e.m. $P < 0.0001$ by analysis of variance (ANOVA) for all bar graphs.

of these genes (**Fig. 1k**–**n**). In particular, we only reproducibly detected expression of IL-1ra, KLK3 and CEACAM5 protein in samples with combinations of TALE-TFs. We found low expression of ERBB-2 in control samples and cells transfected with single TALE-TFs, but its expression was substantially enhanced in cells transfected with all TALE-TFs (**Fig. 1n**).

These results are consistent with a mechanism in which the VP64 acidic activation domain of multiple transcription factors is simultaneously interacting with and stabilizing components of the preinitiation complex[15,16]. We confirmed this mechanism by demonstrating that the VP64 domain, rather than nucleosome displacement by TALEs, was essential to achieving the synergistic effect (**Supplementary Fig. 3**). Alternative acidic activation domains could also synergistically activate gene expression (**Supplementary Fig. 4**). The expression of other genes near *IL1RN* did not increase, indicating that this large synergistic activation was specific to the target gene (**Supplementary Fig. 5**).

We did not specifically design the TALE-TFs used in this study to target DNase-hypersensitive regions (**Supplementary Fig. 6**), which is in contrast to many other reports of synthetic transcription factors that only target open chromatin. In fact, *IL1RN*, *KLK3* and *CEACAM5* are not expressed in HEK293T cells. Notably, targeting chromatin that is inaccessible to DNase did not prevent gene activation by the engineered TALE-TFs (**Fig. 1g**–**i**). These results suggest that targeting open chromatin may not be a prerequisite to successful TALE-TF engineering and that activation of silenced genes is possible in the absence of chromatin-modifying drugs[17], particularly when using combinations of TALE-TFs. In contrast to these three genes, *ERBB2* is moderately expressed in HEK293 cells, and the TALE-TFs for *ERBB2* regulation were targeted to open chromatin (**Supplementary Fig. 6**). Combinations of these TALE-TFs also led to synergistic *ERBB2* activation, although the effect was not as substantial relative to the other genes as a result of higher basal expression (**Fig. 1f**,**j**,**n**).

To comprehensively characterize the effects of combinatorial regulation of mammalian genes by engineered TALE-TFs, we co-transfected all 63 combinations of six TALE-TFs targeting three different genes with a corresponding luciferase reporter into HEK293T cells (**Fig. 2a**–**c**). Various combinations of TALE-TFs could reproducibly achieve tunable levels of gene expression over a large dynamic range. Many TALE-TFs that did not activate the reporter when delivered alone contributed to synergistic activation of expression when combined with other TALE-TFs (**Supplementary Table 2**). In some cases, the addition of a TALE-TF decreased gene expression. However, for all three genes there was an increase in the average gene expression with increasing numbers of TALE-TFs (**Fig. 2d**), and the average contribution of each additional TALE-TF decreased as the number of TALE-TFs increased (**Fig. 2e**).

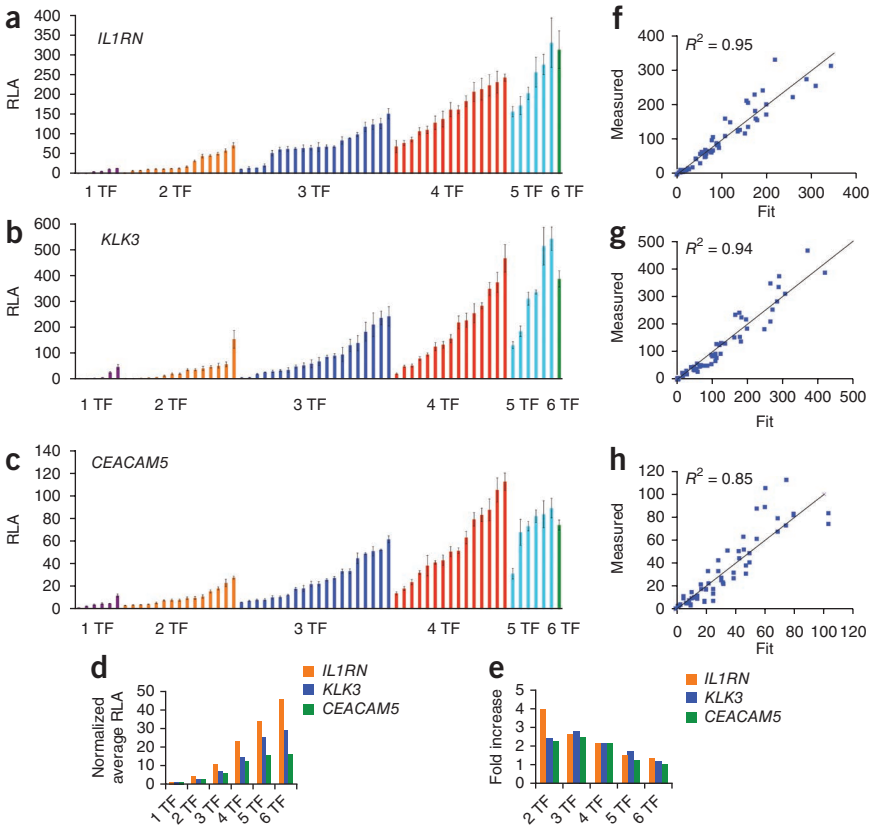

Figure 2 | Combinatorial regulation of gene expression by TALE-TFs. (**a**–**c**) All possible 63 combinations of six TALE-TFs targeting *IL1RN*, *KLK3* and *CEACAM5* were tested for activation of a luciferase reporter plasmid and ordered according to number of TALE-TFs and magnitude of relative luciferase activity (RLA). Samples receiving the same number of TALE-TFs are colored identically. Data are shown as the mean ± s.e.m. $n = 3$ independent experiments. $P < 0.0001$ by ANOVA for all three data sets (**Supplementary Table 2**). (**d**) The average RLA for the indicated number of TALE-TFs for each gene. (**e**) The fold increase of RLA for each number of TALE-TFs relative to the average RLA for one less TALE-TF for each gene. (**f**–**h**) The measured values for all 63 combinations of TALE-TFs versus the values fit by the polynomial model, along with $y = x$ (solid line).

To assign quantitative parameters to the relative contribution of each TALE-TF to the synergistic effect across the 63 data points in these experiments, we applied a polynomial model to the data set of each gene of the form

$$y_j = \left(\sum_{i=1}^{6} w_i x_{i,j} \right)^2$$

where y_j is the relative luciferase activity for the jth combination of the six TALE-TFs. The value of $x_{i,j}$ is 0 if the ith TALE-TF is not included in the jth combination and is 1 if it is included. The effect coefficient w_i is a fit parameter that represents the relative contribution of the ith TALE-TF to the regulation of its target promoter in the context of all combinations of the six TALE-TFs. We used multiple regression to solve for values of w_i for all TALE-TFs for each of the three target genes. These coefficients generated an excellent fit of the experimental data (**Fig. 2f**–**h**) and were highly significant ($P < 2 \times 10^{-3}$) in accurately describing the relative contribution of each TALE-TF (**Supplementary Table 3**). Notably, the polynomial model provides a stronger description of the data than the corresponding additive and multiplicative models (**Supplementary Fig. 7**). This is because the additive model does not account for the synergy of TALE-TF activity (**Fig. 2d**), and the multiplicative model does not account for the diminishing contribution of each additional TALE-TF (**Fig. 2e**). The superior fit of the polynomial model relative to the additive model can be mathematically explained by the second-order terms that are the product of the effect coefficients for the different TALE-TFs. This suggests the presence of some form of cooperativity but does not reveal the underlying mechanism. As discussed above, the simultaneous binding and stabilization of components of the preinitiation complex by VP64 probably has a role[15,16], in addition to other secondary effects of VP64-mediated gene activation on local epigenetics and chromatin structure.

Previous studies have suggested that TALE-TF activity may correspond to proximity to the TSS[17] or to TALE repeat variable diresidue composition[13]. In this study, we found no clear correlation of the effect coefficient with TALE array length, composition or distance to the TSS that was consistent for all genes (**Supplementary Fig. 1**). This suggests that these TALE-TF design parameters cannot be used independently to predict highly effective TALE-TFs. It is probable that other biological and structural components of these gene promoters, including genome folding and competition with endogenous regulatory factors, have a dominant role in determining the activity of single TALE-TFs and TALE-TF combinations.

The cooperative activity of TALE-TFs enables the control of gene expression without the need for small molecules used in conventional chemically regulated systems. Unlike previous work in synthetic biology that has focused on the regulation of transgenes by engineered promoters customized with multiple transcription factor binding sites[1–6], the use of TALE-TF combinations that target endogenous promoters begins to recapitulate the complexity of natural systems in a precise and controlled manner. This approach constitutes a powerful experimental system for elucidating the fundamental mechanisms of natural gene regulation that are currently not well understood. The capacity for combinatorial regulation also provides a new framework for engineering biocomputation systems that control endogenous genes in mammalian cells, similarly to recently developed genetic logic gates that control engineered transgenes[2–6]. Precise control of gene expression with multiple tunable inputs may lead to greater potency, robustness and predictability in bioengineered systems in the context of cell-machine interfaces and gene- and cell-based therapies.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by a US National Institutes of Health (NIH) Director's New Innovator Award (DP2-OD008586), a National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award (CBET-1151035), NIH R03-AR061042, The Hartwell Foundation Individual Biomedical Research Award and a March of Dimes Basil O'Connor Starter Scholar Award to C.A.G.; grants from the NIH (P50-GM081883) and the Defense Advance Research Projects Agency (HR0011-09-1-0040) to A.J.H.; and grants from the NIH to G.E.C. (U54-HG004563) and F.G. (R01-AR048852). D.G.O. was supported by a predoctoral fellowship from the American Heart Association. K.A.G. was supported by an NSF Graduate Research Fellowship.

AUTHOR CONTRIBUTIONS

P.P.-P., A.M.F., G.E.C., A.J.H. and C.A.G. designed experiments. P.P.-P., D.G.O., J.M.B., A.M.F. and K.A.G. performed the experiments. P.P.-P., F.G., G.E.C., A.J.H. and C.A.G. analyzed the data. P.P.-P. and C.A.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nmeth.2361.

1. Guido, N.J. *et al. Nature* **439**, 856–860 (2006).
2. Guet, C.C., Elowitz, M.B., Hsing, W. & Leibler, S. *Science* **296**, 1466–1470 (2002).
3. Cox, R.S. III., Surette, M.G. & Elowitz, M.B. *Mol. Syst. Biol.* **3**, 145 (2007).
4. Khalil, A.S. *et al. Cell* **150**, 647–658 (2012).
5. Ausländer, S., Ausländer, D., Müller, M., Wieland, M. & Fussenegger, M. *Nature* **487**, 123–127 (2012).
6. Tamsir, A., Tabor, J.J. & Voigt, C.A. *Nature* **469**, 212–215 (2011).
7. Moscou, M.J. & Bogdanove, A.J. *Science* **326**, 1501 (2009).
8. Boch, J. *et al. Science* **326**, 1509–1512 (2009).
9. Zhang, F. *et al. Nat. Biotechnol.* **29**, 149–153 (2011).
10. Miller, J.C. *et al. Nat. Biotechnol.* **29**, 143–148 (2011).
11. Cermak, T. *et al. Nucleic Acids Res.* **39**, e82 (2011).
12. Reyon, D. *et al. Nat. Biotechnol.* **30**, 460–465 (2012).
13. Streubel, J., Blucher, C., Landgraf, A. & Boch, J. *Nat. Biotechnol.* **30**, 593–595 (2012).
14. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. *Nat. Commun.* **3**, 968 (2012).
15. Carey, M., Lin, Y.S., Green, M.R. & Ptashne, M. *Nature* **345**, 361–364 (1990).
16. Joung, J.K., Koepp, D.M. & Hochschild, A. *Science* **265**, 1863–1866 (1994).
17. Bultmann, S. *et al. Nucleic Acids Res.* **40**, 5368–5377 (2012).
18. Tremblay, J.P., Chapdelaine, P., Coulombe, Z. & Rousseau, J. *Hum. Gene Ther.* **23**, 883–890 (2012).
19. Geissler, R. *et al. PLoS ONE* **6**, e19509 (2011).
20. Garg, A., Lohmueller, J.J., Silver, P.A. & Armel, T.Z. *Nucleic Acids Res.* **40**, 7584–7595 (2012).

ONLINE METHODS

Cell culture and transfection. HEK293T cells were obtained from the American Tissue Collection Center (ATCC) through the Duke Cancer Center Culture Facility and were maintained in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin at 37 °C with 5% CO_2. HEK293T cells were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Transfection efficiencies were routinely >95% as determined by flow cytometry after delivery of a control enhanced GFP expression plasmid. The amount of DNA used for lipofection was 800 ng per well in 24-well plates or 200 ng per well in 96-well plates. For luciferase reporter assays in 24-well plates, 100 ng of reporter plasmid was included with 700 ng of TALE-TF expression plasmid. When comparing single TALE-TFs to the combination of all TALE-TFs (**Fig. 1**), the total amount of TALE-TF expression plasmid was held constant (800 ng of single TALE-TFs and 800 ng of total TALE-TF expression plasmid divided equally among each factor). When assessing the individual contribution of each TALE-TF (**Fig. 2**), the amount of each TALE-TF was held constant at 116 ng, with empty expression plasmid added to a total of 700 ng. Amounts of DNA for transfections in 96-well plates were scaled accordingly.

Plasmids and TALE-TF. TALE-TFs were assembled using the Golden Gate TALEN and TAL effector kit obtained from Addgene[11]. A destination vector for the final assembly step was created to include a Flag epitope tag and an SV40 NLS at the N terminus, a 152-residue deletion from the N terminus of the wild-type TALE protein that has been previously shown to preserve the DNA-binding ability of TALEs[10], 63 wild-type TAL amino acids after the repeat domain[9], a C-terminal SV40 NLS, a VP64 domain that contains four repeats of the minimal activation domain of VP16 and an HA tag at the C terminus (**Fig. 1a**). TALE-TFs were originally designed to target within the 600 bp upstream of the transcriptional start site (**Fig. 1b**) on the basis of the criteria described by Cermak *et al.*[11], although a subsequent study showed that these guidelines did not affect TALE functionality[12]. TALE-TFs were designed downstream of the transcription start site for *ERBB2* but upstream of the translation start site on the basis of previous studies showing high activity of synthetic zinc finger transcription factors targeting this region[21]. The compositions of all TALE-TFs used in this study are provided in **Supplementary Figure 1**.

The reporter plasmids were built by cloning PCR-amplified genomic DNA sequences upstream of the genes of interest, *IL1RN* (chromosome 2, 113874366–113875462), *KLK3* (chromosome 19, 51357466–51358177), *CEACAM5* (chromosome 19, 42211804–42212651) and *ERBB2* (chromosome 17, 37855857–37856492), into the vector pGL3-Basic (Promega). Coordinates are provided based on the hg19 reference genome.

Luciferase assays. Forty-eight hours after transfection, cells were collected into 96-well plates, washed with PBS once and lysed with 100 mM monobasic sodium phosphate and 0.2% Triton X-100. The lysate was incubated with Bright-Glo Substrate (Promega) in a 1:1 ratio, and luciferase activity was measured using a Synergy 2 Multi-Mode Microplate Reader (BioTek). The results are expressed as relative luciferase activity (RLA), which is the average luciferase activity normalized to the luciferase activity in samples transfected with the reporter vector and the empty TALE-TF expression vector. Data are presented from three independent experiments performed with two biological replicates per experiment.

Western blot analysis. Cells were lysed in 50 mM Tris-Cl (pH 7.4), 150 mM NaCl, 0.5% Triton X-100 and 0.1% SDS. Protein concentrations in cell lysates were measured by the BCA Protein Assay (Pierce). Lysates were mixed with loading buffer and boiled for 5 min, and equal amounts of protein were run in NuPAGE Novex 4–12% Bis-Tris Gel polyacrylamide gels and transferred to nitrocellulose membranes. Nonspecific antibody binding was blocked with 50 mM Tris, 150 mM NaCl and 0.1% Tween-20 (TBS-T) with 5% nonfat milk for 30 min. The membranes were incubated with primary antibodies (horseradish peroxidase (HRP)-conjugated anti-HA (Roche, clone 3F10) in 5% milk in TBS-T diluted 1:5,000 for 30 min; anti-CEACAM5 (Cell Signaling Technology, clone CB30) in 5% milk in TBS-T diluted 1:1,000 overnight; anti-GAPDH (Cell Signaling Technology, clone 14C10) in 5% milk in TBS-T diluted 1:5,000 for 30 min; and anti-ERBB2 (Cell Signaling Technology, clone 29D8) in 5% BSA in TBS-T diluted 1:2,000 for 2 h). The membranes were then washed with TBS-T for 30 min. Membranes labeled with primary antibodies were incubated with rabbit HRP-conjugated antibody (Sigma-Aldrich, catalog number A6154) diluted 1:5,000 for 30 min and washed with TBS-T for 30 min. Membranes were visualized using the Immun-Star WesternC Chemiluminescence Kit (Bio-Rad), and images were captured using a ChemiDoc XRS+ System and processed using ImageLab software (Bio-Rad).

Enzyme-linked immunosorbent assay. Serum-free culture medium (OPTI-MEM) was collected and frozen at −80 °C. Human IL-1ra and KLK3 secretion into culture medium was quantified by ELISA according to the manufacturer's protocols (R&D Systems, catalog numbers DY280 and DKK300, respectively). For the IL-1ra ELISA, the standard curve was prepared by diluting recombinant human IL-1ra in OPTI-MEM, and the IL-1ra in culture medium was measured undiluted. For the KLK3 ELISA, the standard curve was prepared by diluting recombinant KLK3 in the manufacturer's calibrator diluent, and the samples were concentrated approximately eightfold by centrifugation through 3-kDa MWCO filters for 20 min (Amicon Ultra, catalog number UFC500396). Reported values were corrected by the concentration factor for each sample.

For both assays, optical density was measured at 450 nm with a wavelength correction at 540 nm. Each standard and sample was assayed in duplicate. The duplicate readings were averaged and normalized by subtracting the average zero standard optical density. A standard curve was generated by log transforming the data and performing a linear regression of the IL-1ra or KLK3 concentration versus the optical density. The lower limit of detection was 50 pg/ml for human IL-1ra and 32 pg/ml for human KLK3. Data reported are the mean and s.e.m. of these individual values combined from multiple experiments ($n = 6$ biological replicates for IL-1ra, $n = 4$ biological replicates for KLK3).

Quantitative RT-PCR. Total RNA was isolated using the RNeasy Plus RNA isolation kit (Qiagen). cDNA synthesis was performed

using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). Real-time PCR using SsoFast EvaGreen Supermix (Bio-Rad) was performed with the CFX96 Real-Time PCR Detection System (Bio-Rad) with 45 cycles, melting for 2 s at 95 °C and annealing and extension for 2 s at 55 °C. Real-time PCR oligonucleotide primers (ERBB2, 5′-AGCCGCGAGCACCCAAGT-3′ and 5′-TTGGTGGGCAGGTAGGTGAGTT-3′; CEACAM5, 5′-TCCCCACAGATGGTGCAT-3′ and 5′-GAACGGCGTGGATTCAATAG-3′; KLK3, 5′-CTCGTGGCAGGGCAGTCT-3′ and 5′-AGCTGTGGCTGACCTGAAAT-3′; IL1RN, 5′-GACCCTCTGGGAGAAAATCC-3′ and 5′-GTCCTTGCAAGTATCCAGCA-3′; PSD4, 5′-GCAGCACCTCCTGGTCAC-3′ and 5′-ATCCGACACATCCTGATTCC-3′; IL1F10, 5′-CCTCCCCATGGCAAGATACT-3′ and 5′-AGCAGTTGTCTGCAACAGGA-3′; and GAPDH, 5′-CAATGACCCCTTCATTGACC-3′ and 5′-TTGATTTTGGAGGGATCTCG-3′) were designed using Primer3Plus software and purchased from IDT. Primer specificity was confirmed by agarose gel electrophoresis and melting curve analysis. Reaction efficiencies over the appropriate dynamic range were calculated to ensure linearity of the standard curve. Data are presented from three independent experiments performed with two biological replicates per experiment.

Statistics. Statistical analyses were performed by single-factor ANOVA with $\alpha = 0.05$ in Microsoft Office Excel 2007. Effect coefficients (**Supplementary Table 3**) were determined using the regression tool in the data analysis add-in to Microsoft Office Excel 2007, with the relative luciferase activities (**Fig. 2a–c** and **Supplementary Table 2**) serving as the *y* input and an array of zeros and ones representing each TALE-TF combination as the *x* input.

21. Beerli, R.R., Dreier, B. & Barbas, C.F. III. *Proc. Natl. Acad. Sci. USA* **97**, 1495–1500 (2000).

doi:10.1038/nmeth.2361