

In Depth Characterization of the Promoter Proximal Proteome of Single Copy Locus *FOXP2*

Authors

Tim M. G. MacKenzie, Lucia Ramirez, Ruiqi Jian, Lihua Jiang, and Michael P. Snyder

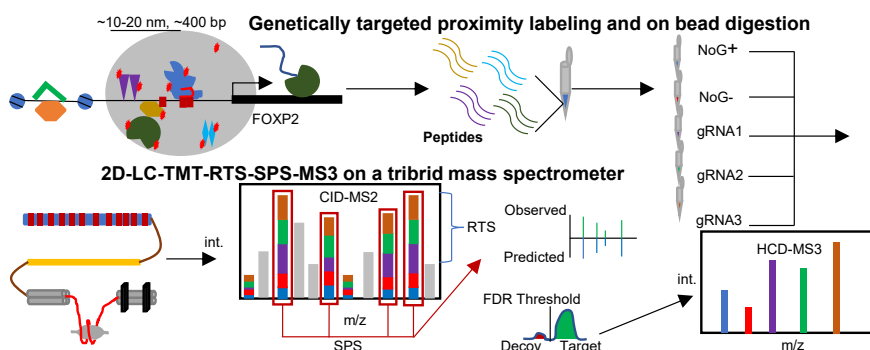
Correspondence

mpsnyder@stanford.edu

Graphical Abstract

In Brief

Genetically targeted proximity labeling with dCas9-APEX2 coupled to quantitative mass spectrometry proteomics via 2D-LC-TMT-RTS-SPS-MS3 on a tribrid mass spectrometer uncovers promoter interactors at the single copy locus *FOXP2*. Candidate proteins are cross-validated with extensive multi-omic integration, including computationally predicted locus binders and hundreds of ChIP-Seq experiments performed by ENCODE. IRF2BP2 and glucocorticoid signaling are revealed as regulators of *FOXP2* expression. MS transcription factor detection is significantly correlated with expression levels, with more lowly expressed proteins less likely to be detected.



Highlights

- dCas9-APEX2 proximity labeling reveals promoter proteome of single copy locus *FOXP2*.
- Deep proteome coverage with 2D-LC and TMT-RTS-SPS-MS3 on a tribrid instrument.
- Multi-omic integration including ENCODE ChIP-Seq to confirm dozens of locus binders.
- IRF2BP2 and glucocorticoid signaling discovered as negative regulators of *FOXP2*.
- Lowly expressed transcription factors significantly less likely to be detected by MS.

In Depth Characterization of the Promoter Proximal Proteome of Single Copy Locus *FOXP2*

Tim M. G. MacKenzie¹, Lucia Ramirez, Ruiqi Jian, Lihua Jiang², and Michael P. Snyder^{1*}

Identifying the proteins that interact with sequence-defined chromatin segments is a critical step in understanding gene expression. Most procedures perform bulk analysis of samples to identify the general interactions that occur in a cellular population and thereby do not detect the proteins that operate at a single locus. To circumvent this limitation, we developed a modified method that uses genetically targeted proximity labeling with dCas9-APEX2 to specifically biotinylate the promoter proximal proteome of the single copy locus *FOXP2* in live HEK293 cells. After capture of the tagged proteins with streptavidin and isobaric labeling of the peptides produced from on-bead digestion with tandem mass tags, we used quantitative 2D-LC-MS3 on a tribrid mass spectrometer to identify 373 significantly enriched proteins at the active promoter relative to control samples (Storey- $q < 0.05$, fold change > 1.2). These proteins were enriched for transcription factors (TFs) and components of the spliceosome. To validate our candidate transcriptional regulators, we utilized computationally predicted TF binding and the >200 ChIP-Seq experiments performed in HEK293 cells by Encyclopedia of DNA Elements. In addition to validating dozens of candidate TFs as binders of the targeted genomic locus, we newly identify IRF2BP2 and glucocorticoid signaling as negative regulators of *FOXP2* transcription, suggesting they each play a key role in *FOXP2* gene expression. We further demonstrate that MS detects approximately one third of both binders and nonbinders, with more highly expressed genes significantly more likely to be detected regardless of binding status or locus specificity. Encyclopedia of DNA Elements ChIP-Seq binders not detected by MS show significantly lower expression compared to nonbinders only at the targeted *FOXP2* promoter and not at off target loci.

Deciphering the factors that regulate gene expression at important loci is crucial for linking genotype and phenotype. These transcription factors (TFs) bind to DNA and recruit transcriptional machinery and chromatin remodeling proteins

to impact gene expression. The composition of proteins at both promoters and enhancers integrates environmental and cell-intrinsic signals to determine the activation or silencing of a gene at steady state and in response to external cues (1–3). Determining the proteins proximal to a gene promoter is a key step in building a mechanistic understanding of gene expression and developing experimental interventions in the process.

There is a growing appreciation for the role of alternative promoter selection in regulating gene expression (4–9). Results from both long read (10) and traditional short read (11) sequencing have indicated that alternative promoters resulting in different transcription start sites (TSSs) and 5' UTRs are the major driver of transcriptome diversity in many cancers and can provide a better patient prognosis than gene-level expression. The choice of TSS and the resulting 5' UTR can impact the translation efficiency of the resulting gene, even when the same protein isoform is encoded (12–14). These noncoding regions are enriched in ultraconserved elements, regulatory regions with greater than 90% sequence identity for more than 100 bp that emerged more than 400 million years ago, before the evolution of lobe-finned fish and amphibians (15). Some promoters are ubiquitously active, whereas others are tissue, developmental time point, or disease-state specific (16, 17). Efforts to identify promoter-proximal proteomes to understand the molecular mechanisms governing alternative promoter selection require tools that can detect minute analytes across a wide dynamic range and specific targeting to resolve a ~ 1 kbp promoter.

The isolation of sequence-defined chromatin regions represents a major biochemical challenge due to low intrinsic signal coupled to high background (18–20). There are only two counts of a single copy locus per diploid cell, and a 1 kbp promoter represents only $\sim 0.0001\%$ of the Gbp sized human genome. Compounding the challenge is the biophysical similarity of chromatin—negatively charged DNA and positively charged nucleosomes with a large dynamic range of

From the Department of Genetics, Stanford University, Stanford, California, USA

*For correspondence: Michael P. Snyder, mpsnyder@stanford.edu.

proteins of interest. The copy number of TFs per animal cell is on the order of 10,000 to 300,000 per nucleus (21), and there are tens of millions of histones per cell (22, 23). Early efforts to isolate sequence-defined chromatin regions focused on yeast to reduce background with its smaller genome (24, 25) or telomeres to increase signal due to their sequence abundance (26, 27). Recent efforts have utilized nuclease-dead dCas9-based strategies to target and isolate single copy loci in model organisms from human cells (28, 29) to plants (30, 31), typically relying on formaldehyde-crosslinking proteins to chromatin and isolating them. Complementary approaches have relied on proximity labeling enzymes like BioID or the engineered peroxidase APEX2 that can biotinylate proteins nearby their targeted region (32, 33). Several examples using dCas9-APEX2 demonstrate specific biotinylation of proteins physically near a targeted promoter region (34–44). These proximity labeling approaches preclude the need to cross-link chromatin, thereby enabling capture of chromatin-associated proteins that are not covalently linked to DNA by formaldehyde and allow for stringent wash steps using streptavidin-based purification of the labeled local proteome (45).

Mass spectrometry proteomics is the analytical tool often used for proteomic profiling of chromatin (46). While some approaches have compared the whole chromatin proteome between cell states (47, 48), advances in protocols and instrumentation have enabled deep proteome coverage even on material-limited samples including single cells (14, 49, 50) and subcellular spatial proteomes with organellar resolution (51, 52). However, these analyses do not uncover the proteins operating at single copy loci. Highly accurate quantitation and multiplexing can be achieved with tandem mass tags (TMTs) that barcode individual samples before combining them together for LC-MS analysis (53). A typical “bottom-up” data-dependent acquisition shotgun proteomics approach digests the protein fraction of samples to the corresponding peptides which are separated by liquid chromatography, measured on the mass spectrometer, and fragmented to produce characteristic patterns that are also measured on a mass analyzer (MS2) (54, 55). Extensive fractionation using an online 2D-LC system resolves complex chromatography peaks to allow for deeper proteome coverage (56). A third mass spectrometry step (MS3) to measure TMT signals ensures greater fidelity to the underlying ground truth quantitation by accounting for reporter ion ratio compression, and real-time database searching and synchronous precursor selection (RTS-SPS) economizes on instrument time and increases signal, respectively (57–59). The Orbitrap Ascend tribrid mass spectrometer design allows for deep proteome coverage using a TMT-based RTS-SPS-MS3 proteomics experiment (60, 61).

The TF *FOXP2* is among the most highly conserved genes across vertebrates in terms of sequence and location and developmental time of expression (62–65). Originally

disclosed as the first gene identified to be linked to Mendelian speech disorders, it has drawn a great deal of attention as a model to study evolution across vertebrates broadly and related to speech and vocalization in particular (66–74). The *FOXP2* locus harbors a cluster of 27 ultraconserved regions shared across vertebrates but not invertebrates, with only a dozen genes in the entire genome having a higher number of such conserved clusters (75). While some research has sought out downstream targets of *FOXP2* (76–78) or identified its role in developing brain (79–81), lung (82), and other tissues (83, 84), there have been fewer reports on regulation of *FOXP2* expression itself (85, 86). This may be partially explained by the complex patterns of regulation and expression: at least four separate alternative promoters can be active depending on the system under study, and further regulation can occur in *cis* from genomic elements 3 Mbp away from the upstream promoter (87). Understanding the molecular factors contributing to this complex regulation has relevance to human health as altered *FOXP2* expression, both upregulation (88, 89) and downregulation (90–92), has been tied to various types of cancer and poor patient prognosis (93). The high degree of conservation *FOXP2* shows in its transcriptional regulation, coupled to its phenotypic relevance in vertebrate development and human health, make it an ideal model system to study the principles governing alternative promoter selection.

In an effort to unravel the highly conserved *FOXP2* regulatory module, we sought to deeply characterize in live HEK293 cells the proteins interacting with the active, upstream promoter of this single-copy locus (TSS1). We used genetically targeted proximity labeling with dCas9-APEX2 to covalently tag the promoter-proximal proteome with biotin. After streptavidin magnetic bead-based purification with stringent washing and on-bead digestion, we utilized online 2D-LC deep fractionation with TMT-enabled quantitative proteomics *via* RTS-SPS-MS3 on a tribrid mass spectrometer. Bioinformatic analysis on the 6029 proteins inferred by Proteome Discoverer (94) identified 373 proteins significantly enriched at the *FOXP2* promoter compared with the untargeted no gRNA control [Storey-*q* < 0.05, fold change (FC) > 1.2]. This set of proteins was significantly enriched for TFs and spliceosome components. To benchmark our candidate proteins and provide a comprehensive characterization of potential regulators of the upstream promoter of *FOXP2*, we compared our proteomic results to multiple orthogonal approaches, including computationally predicted TFs recognizing the targeted region and the 223 Encyclopedia of DNA Elements (ENCODE) chromatin immunoprecipitation (ChIP)-Seq experiments performed in HEK293(T) cells (218 TFs, 5 histone modifications) (95–98). We discovered and verified dozens of candidate regulators of the chromatin region at the active promoter of *FOXP2*. We also show that IRF2BP2 and glucocorticoid signaling each negatively regulate *FOXP2* expression in siRNA experiments and

treatment with dexamethasone, respectively. We further demonstrate a uniform MS detection rate of approximately one third for binders and nonbinders at on target and off target loci, with more highly expressed genes significantly more likely to be detected by MS regardless of binding status or locus specificity. Proteins not detected by MS show significantly lower expression for binders *versus* nonbinders only at the on target TSS1, while there is no significant expression difference for undetected proteins at off target loci.

EXPERIMENTAL PROCEDURES

Experimental Design and Statistical Rationale

For proteomics experiments, we employed two negative controls: one without labeling occurring to account for endogenously biotinylated proteins and nonspecific interaction with streptavidin magnetic beads (NoG-) and one with untargeted labeling to account for proximity labeling background from dCas9-APEX2 molecules not associating with targeted locus within the cell (NoG+). To increase signal of the promoter proximal proteome, we tiled the locus with three gRNAs in separate cell lines as pseudo-technical replicates. To account for biological and technical batch-based variation, we required a minimum of three biological replicates, totaling 15 separate conditions for analysis. These requirements are achievable with TMTpro which is capable of 18-plexing. Each of the five conditions were processed in parallel for each biological replicate until all samples were combined after TMT labeling. For ChIP, each of the four conditions (untargeted NoG negative control plus three on target sgRNAs) had three biological replicates. Each biological replicate produced two technical replicates for ChIP for a total of $n = 6$ per condition. As with proteomics experiments, the targeted sgRNA cell lines were pooled together as pseudo technical replicates, and conditions within a biological replicate were processed simultaneously. For protein bioinformatic analysis, we report adjusted p -values to account for the multiple hypothesis comparison problem. When filtering out background, we used strict Bonferroni correction to limit false positives being inappropriately called contaminants (99). When testing for enrichment, we used Storey correction to be more permissive in calling hits (100, 101). Global quantification data approximated a negative binomial distribution before $\log_2$ transformation, after which it approximated a normal distribution. To test statistical significance of protein types in enriched *versus* total sample, we performed a χ^2 test on the categorical data (in class/not in class). We performed Cochran's Q test to ascertain whether the proportion of ENCODE TFs binding at analyzed loci significantly differed. We used McNemar's test for pairwise comparisons of individual loci as a follow up to Cochran's Q test with Bonferroni correction to call significance.

Cell Line Construction and Culture

HEK cells were cultured in Dulbecco's Modified Eagle Medium containing 10% FBS and 1% penicillin-streptomycin and passaged every 3 days. Cells were maintained at 37 °C in a humidified atmosphere containing 5% CO₂. To stably incorporate the Caspex construct, cells were transfected with Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions and selected for 2 weeks with 4 μ g/ml puromycin in the media. Inducible Caspex expression was a gift from Steven Carr & Samuel Myers (Addgene plasmid #97421) (38). Single colonies were selected by fluorescence activated cell sorting (FACS) at the Stanford Shared FACS Facility

(SF1a) and tested for doxycycline inducibility of Caspex monitored by GFP visualization and anti-FLAG Western blotting with Rabbit M2 (Sigma F2555). Individual sgRNA constructs were transfected into the best responding Caspex cell line and selected with 200 μ g/ml hygromycin. Cell lines were maintained with 4 μ g/ml puromycin and/or 200 μ g/ml hygromycin to maintain selection pressure for the Caspex and sgRNA plasmids, respectively, until use in ChIP-PCR or proximity labeling experiments.

Plasmid Construction

The UCSC Genome browser was utilized to select sgRNA sequences targeting the promoter of *FOXP2* that conformed to requirements for expression from the U6 promoter while minimizing off target effects. Selected guides were required to start with G for transcription initiation and were rejected if they contained a run of 4 or more T's in a row to prevent early termination. The MIT specificity score for all guides was greater than 90. Appropriate gRNA sequences were cloned into the lenti-sgRNA hygro backbone, a gift from Brett Stringer (Addgene plasmid #104991) (102). Briefly, the lenti-sgRNA hygro plasmid was digested with *BsmBI* (New England Biolabs) to create overhangs that could hybridize with Fwd-5'-ACACCGN₂₀G-3' and Rev-5'-AAAACN₂₀G-3' sequences, which were ligated with the designed oligonucleotides using T4 polymerase (New England Biolabs). The vector was transformed into One Shot Stbl3 chemically competent *E. coli* (Invitrogen), single colonies were picked, and plasmid DNA was purified with a QIAGEN EndoFree Maxiprep Kit. Successful gRNA incorporation was confirmed by Sanger sequencing (Integrated DNA Technologies). Primers for gRNA incorporation were synthesized by Integrated DNA Technologies and were as follows: g1 Fwd-5'-ACACCGCAGACACCTTCGGTGATAG-3', Rev-5'-AAAATATCACCGAAAGGTGTCTGCG-3'; g2 Fwd-5'-ACACCGACACCTTCGGTGATAGGGG-3', Rev-5'-AAAACCCCTATCACCGAAAGGTGTCTGCG-3'; g3 Fwd-5'-ACACCGTTATCCCGAAGC GTCAGTAG-3', Rev-5'-AAAATACTGACGCTTCGGGATAACG-3' to target sequences in chr7 (hg38) of g1:114087491-114087513, g2:114087494-114087516, g3:114087279-114087301. Target sequences were chosen for proximity to the TSS and overlapping labeling radii (~400 bp (38)) within the *FOXP2* TSS1 promoter while conforming to U6 expression requirements.

Cross-Linking and Sonication for ChIP

Doxycycline (70% in ethanol) was added to cells at a final concentration of 1 μ g/ml in a 15-cm² plate for 21 h so Caspex expression would be induced when cells were approximately 90% confluent (~10⁷ cells). Caspex expression was visually confirmed by fluorescence of the GFP marker. A single cell suspension was crosslinked with 1% formaldehyde at room temperature with rotation for 10 min. To quench formaldehyde, 2M glycine was added to a final concentration of 125 mM and incubated for 5 min at room temperature with rotation. Cells were washed twice with ice-cold PBS, pelleted, snap-frozen in liquid nitrogen, and stored at -80 °C until use. Cells were thawed at 4 °C in ice-cold PBS with rotation for 30 min. Pelleted cells were treated with 3 ml hypotonic buffer (20 mM Hepes pH 7.9, 10 mM KCl, 1 mM EDTA pH 8.0, 10% glycerol) with protease inhibitors [Roche cOmplete protease inhibitor (1 tablet/50 ml), 0.5 mM PMSF] and 1 mM DTT added just before use. The plasma membrane was allowed to swell for 10 min before shearing with 30 strokes of a Dounce homogenizer. Total time for swelling and homogenization was limited to 15 min. Nuclei were pelleted and washed once with ice-cold hypotonic buffer before lysis in 3 ml ice-cold radio-immunoprecipitation (RIPA) buffer (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, and 0.1% SDS) with protease inhibitors, DTT, and phosphatase inhibitors (1 mM Na₂P₂O₄, 2 mM Na₃VO₄, 10 mM NaF) added just before use. Lysed

nuclear pellets were sonicated on ice to shear chromatin for solubilization with a SFX250 Sonifier (Branson) set to an intensity (output control) of 3.5. Lysates were sonicated for 16 rounds of 30 s on and 30 s off, taking care to prevent foaming. To prevent sample overheating, lysate was allowed to rest on ice for at least 2 min every four cycles. Lysate was clarified by centrifugation at 14,000 rpm for 15 min at 4 °C. Supernatant was transferred to a 15 ml Falcon tube, diluted to 4 ml total in RIPA buffer, flash frozen in liquid nitrogen, and stored at –80 °C until use. Chromatin fragmentation was assessed after de-crosslinking (*vide infra*) using a Bioanalyzer (Agilent). Typical results produced an average size of ~530 bp with 55 to 60% of fragments in the 200–1000 bp range (SF1b).

Chromatin Immunoprecipitation

For chromatin immunoprecipitation, each biological replicate produced two technical replicates for pulldown. For one technical ChIP replicate, 2 ml of sonicated lysate was used. Before pulldown, 20 µl input (1%) was removed to compare nucleic acid enrichment before and after ChIP. Sonicated lysate was incubated at 4 °C overnight with 5 µg FLAG Rabbit M2 monoclonal antibody (Sigma F2555) with rotation. A 1:1 mixture of 150 µl total Protein A:Protein G magnetic beads (Invitrogen Dynabeads, 30 mg/ml each) was washed twice with 1 ml ice-cold RIPA buffer to pre-clear the beads. The beads were transferred fully to the antibody-complex-chromatin mixture and incubated for 1 h at 4 °C with rotation. The beads were washed three times with ice-cold RIPA buffer with inhibitors added followed by a wash with ice-cold PBS. The beads were transferred from the 15 ml Falcon tube to a 2 ml DNA lo-bind Eppendorf tube with 800 µl + 2 times 400 µl ice-cold PBS to ensure complete transfer. The PBS was removed, and beads were incubated with 100 µl of freshly made 1% SDS, 10 mM Tris pH 8.0, and 1 mM EDTA at 65 °C for 10 min with gentle mixing by vortex every 2 min. The supernatant was collected, and beads were incubated with 150 µl of 0.67% SDS, 10 mM Tris pH 8.0, and 1 mM EDTA at 65 °C for 10 min with gentle mixing every 2 min before combining both eluates. Input DNA (stored at 4 °C overnight) was diluted 1.5x in 1% SDS, 10 mM Tris pH 8.0, and 1 mM EDTA before both ChIP DNA and input DNA had crosslinks reversed overnight at 65 °C. An equal volume of 1% SDS, 10 mM Tris pH 8.0, and 1 mM EDTA with 100 µg RNase A (QIAGEN) was added to de-crosslinked DNA and incubated for 30 min at 37 °C followed by addition of 5.0 µl of 20 mg/ml Proteinase K (QIAGEN) and incubation at 45 °C for 30 min to remove RNA and proteins. DNA was purified with QIAGEN QIAquick purification columns and used for qPCR. De-crosslinked and purified input DNA (1 µl) was used for the Bioanalyzer assay to assess chromatin fragmentation.

qPCR

Each pulldown was analyzed in quadruplicate using SYBR Green qPCR (Applied Biosystems, A46012). Input DNA was typically diluted 10x to reach similar concentration to ChIP DNA. Forward and reverse primers synthesized by Integrated DNA Technologies at 400 nM were combined with 2 µl DNA and 10 µl SYBR Green Mix in a 20 µl final reaction volume. Enzyme was activated at 95 °C for 2 min before 40 cycles of denaturation and annealing (95 °C for 5 s, 60 °C for 30 s). The following primers were used (proximity of g1 and g2 allowed the same primer pair): g1/g2 Fwd-5'-TGGCTGTTTGTGGGTGGGTTT-3', Rev-5'-GAAGCCCTCCCTATCACCGAA-3'; g3 Fwd-5'-GGAGTCAA-GAACTCCTGGGC-3', Rev-5'-TCAAGGCAGCAGTCATCCCT-3'; GAPDH control Fwd-5'-TTGGCTACAGCAAGAGGGTG-3', Rev-5'-GGGGAGATTTCAGTGTGGTGG-3'. Enrichment was determined using the $2^{-\Delta\Delta Ct}$ method (103). Input DNA was normalized to 100% by subtracting the appropriate dilution factor from the raw Ct value (e.g., 9.966 for 10x dilution of 1% input [$\log_2\{1000\}$]). The Ct values for each individual gRNA from a single technical replicate were averaged.

The degree of enrichment after ChIP was calculated by comparing the Ct of ChIP DNA for the condition of interest normalized to the off-target GAPDH control with the Ct of adjusted input DNA normalized to off target GAPDH, i.e., $\text{ChIP}(\text{Ct}[\text{pooled gRNAs or NoG}] - \text{Ct}[\text{GAPDH}]) - \text{adjusted input}(\text{Ct}[\text{pooled gRNAs or NoG}] - \text{Ct}[\text{GAPDH}]) = \Delta\Delta Ct$. The FC was calculated as $2^{-\Delta\Delta Ct}$. FC for biological and technical replicates were averaged and presented numerically as mean $\pm$ SEM and visually as a box and whisker plot.

Proximity Labeling

Cells were treated with 1 µg/ml doxycycline 21 h before labeling, and Caspex expression was confirmed by fluorescence of the GFP marker. Biotin tyramide phenol (APEX-BIO) was diluted in external media from a 500 mM stock solution in DMSO to solubilize before being added to cells at a final concentration of 500 µM. After at least 30 min of incubation to allow time for the rate-limiting step of biotin diffusion across membranes (104), hydrogen peroxide was diluted in media to 100 mM before addition to the cells at a final concentration of 1 mM to induce biotinylation. After 60 s of very gentle swirling, the media were decanted as quickly as possible, and cells were washed three times with ice-cold quenching solution (100 mM sodium azide and 100 mM sodium ascorbate in PBS), taking care to avoid dislodging the cells from the dish. Cells were washed with once ice-cold PBS to help remove excess biotin phenol, transferred to 15 ml Falcon tubes, pelleted, flash frozen in liquid nitrogen, and stored at –80 °C until use. For confirmation of proximity labeling, one 10-cm² dish was used per biological replicate. For mass spectrometry studies, eight 15-cm² dishes were used per replicate per condition.

Western Blot

Labeled cell pellets were thawed at 4 °C with gentle rotation in ice-cold PBS. Pelleted cells were lysed in ice-cold RIPA with freshly added protease and phosphatase inhibitors, DTT, and 10 mM sodium ascorbate and 10 mM sodium azide added to inhibit APEX2 to prevent excess labeling with any remaining adventitious biotin phenol. Lysate was sonicated on ice for three 10 s cycles. Forty micrograms of protein were denatured in NuPAGE LDS sample buffer (Invitrogen) supplemented with 50 mM DTT, followed by SDS-PAGE using 4 to 12% Bis-Tris WedgeWell Gel (Invitrogen) in MOPS running buffer (Invitrogen). Proteins were transferred to nitrocellulose membranes (0.2 µm, Bio-Rad) with a Trans-Blot Turbo Transfer System (Bio-Rad), blocked with 5% milk in TBST (30 min), and incubated with GAPDH Rabbit monoclonal RM114 antibody (Sigma SAB5600208) at 4 °C overnight at a 1:2000 dilution. Following primary antibody incubation, membranes were washed three times with TBST, incubated with horseradish peroxidase-conjugated secondary anti-rabbit IgG (Cell Signaling 70745) and Streptavidin (Abcam AB279315) for 30 min (1:1000 and 1:10000 dilutions, respectively). Membranes were washed four times, and signals were developed with chemiluminescence and imaged using the ChemoDoc Imaging system (Bio-Rad). Images were processed with the Fiji implementation of ImageJ (105). Changes in contrast were applied to the entire blot.

Enrichment of Biotinylated Proteins and on Bead Digestion

Cell pellets were thawed at 4 °C with gentle rotation in cold PBS. Pelleted cells were lysed in 3 ml ice-cold RIPA buffer with protease and phosphatase inhibitors, DTT, sodium ascorbate, and sodium azide added fresh before use. Lysate was sonicated as described above for ChIP. Typical sonication conditions resulted in an average size of ~450 bp with 85 to 90% of chromatin between 200 and 1000 bp as assessed by the Bioanalyzer assay (SF1c). Protein concentration was determined by the Bradford assay at 595 nm (Abcam AB102535)—the addition of redox quenchers to prevent excess

biotin labeling precludes use of the common BCA assay. For each condition, 500 μ l of a Streptavidin M280 magnetic bead slurry (Dynabeads Invitrogen, 10 mg/ml) was utilized. The magnetic beads were washed twice with ice-cold RIPA buffer. Lysates of equal protein amounts for each condition were incubated with precleared beads for 120 min at room temperature. After capture of biotinylated proteins in the Falcon tube, beads were fully transferred to a 2 ml Eppendorf protein lo-bind tube. Contaminants were removed by washing twice with 1 ml lysis buffer, once with 1M KCl, once with 100 mM Na_2CO_3 , and twice with 6M guanidinium chloride (GdnCl) in 50 mM triethylammonium bicarbonate (TEAB) (all ice-cold). Proteins were denatured, reduced, and alkylated by incubation with 100 μ l 6M GdnCl, 10 mM TCEP, 40 mM chloroacetamide in 100 mM Tris pH 8.5 for 5 min at 95 °C followed by 55 min at room temperature. Beads were washed four times with ice cold 50 mM TEAB to remove excess GdnCl to prevent incomplete protease digestion. 1% of the bead mixture was removed to check capture efficiency via biotin elution (*vide infra*) before digestion. To a 20 μ l slurry of bead-bound proteins in TEAB was added 400 ng (2 μ l) of trypsin/LysC (Promega) (106), and the reaction was incubated at 37 °C overnight. Digestion was quenched by the addition of 10% trifluoroacetic acid (TFA) to a final concentration of 1%. Supernatant was collected, and the beads were washed with 25 μ l of 1% TFA, and supernatants were combined. Peptides were stored at -80 °C until use. Remaining beads were eluted with biotin buffer to check digestion efficiency (45). Briefly, beads were incubated with 50 μ l 2x LDS loading buffer, 200 mM DTT, and 15 mM biotin for 15 min at 95 °C two times. Combined supernatants were diluted 2x in LDS loading buffer and used for Western blot.

TMT Labeling

Peptides were desalted with a Waters Oasis HLB Cartridge before labeling with TMTpro 18-plex reagent (Thermo) according to the manufacturer's protocol. Briefly, TMT label reagents in anhydrous acetonitrile were added to each sample. The TMT isotopic labels were randomized across the different conditions according to the following scheme: g2 Rep1:126; g2 Rep2:127N; g1 Rep3:127C; g1 Rep1:128N; NoG + Rep3:128C; NoG + Rep1:129N; NoG-Rep1:129C; empty:130N; NoG + Rep2:130C; NoG-Rep2:131N; g3 Rep3:131C; g3 Rep1:132N; g3 Rep2:132C; sample pool: 133N; g1 Rep2:133C; NoG-Rep3:134N; g2 Rep3:134C; empty:135N, where Rep# refers to the biological replicate. The labeling reaction was allowed to proceed for 1 h at room temperature before quenching with 5% hydroxylamine for 15 min. Samples were dried with a SpeedVac (Thermo) before resuspension in 100 mM ammonium formate for LC-MS analysis.

Online 2D-LC and Data Acquisition via RTS-SPS-MS3

Data were collected with a Waters Acquity UPLC M-Class 2D-LC system directly connected to an Ascend Tribrid Mass Spectrometer (Thermo). A 12 multifraction gradient over 220 min was utilized with injection of 15 μ g of peptides. Peptide amount was estimated using calibration runs on the MS with equal volume of each TMT-labeled sample injected to ensure equal peptide loading in the data acquisition run. The first dimension of separation at high pH across a BEH column consisted of buffer A (20 mM ammonium formate at pH 10) and buffer B (acetonitrile) using 12 discontinuous steps of buffer B at 10.8%, 13.1%, 14.9%, 16.7%, 17.7%, 18.9%, 19.9%, 20.4%, 22.2%, 25.8%, 28.9%, and 45% at a flow rate of 2 μ l/min before transfer to a low pH C18 analytical column (75 μ m ID/10 μ m tip ID x 28 cm C18-AQ 1.8 μ m resin). For each step, a 5 min isogradients of %B was used (SF2). TICs in SF2 were produced with the RaMS package in R (107). In the second dimension, a linear gradient from 5% to 30% buffer B (0.1% formic acid in acetonitrile) in buffer A

(0.1% formic acid in water) at a flow rate of 300 nl/min was used and directly injected to the mass spectrometer. The 3s duty cycle scan sequence began with an Orbitrap MS1 spectrum with the following parameters: resolution 120,000, scan range 400 to 1500 m/z , automatic gain control (AGC) target of 400,000 (100%), maximum injection time 50 ms, and RF lens 50% with a minimum of 6 points desired across the peak. Precursors were filtered using monoisotopic peak determination, charge state 2 to 7, and dynamic exclusion of 60 s with a ± 10 ppm tolerance excluding isotopes and different charge states. MS2 spectra were collected in the linear ion trap at a rapid scan rate with precursor fit of 50% in a 0.7 m/z window with a scan range of 400 to 1500 m/z (*mode auto*). Ions were fragmented with collision induced dissociation (CID) at a collision energy of 33% (activation time 10 ms, $Q = 0.25$) with a maximum injection time of 40 ms and AGC target of 25,000 (250%). MS2 ions were selected by quadrupole across a 300 to 1500 m/z range for RTS-SPS analysis against the UniProt 2022 human protein database. A maximum of four peptides per protein were selected for RTS-SPS analysis. Enzyme specificity was set to Trypsin/P with a maximum of one missed cleavage and 35 ms maximum search time. A maximum of two variable modifications were allowed. Static modifications of cystine carbamidomethylation (57.0215 Da) and lysine and n-terminal TMTpro modification (304.2071 Da) plus variable methionine oxidation (15.9949 Da) were included for selection to MS3. Scoring thresholds for each charge state were set to Xcorr = 2.5(2+), 2.6(3+), 3.2(4+), and 3.2(5+) with dCn = 0.1 for all states. MS3 was measured in the Orbitrap at a resolution of 60,000 with fragmentation achieved via higher-energy collisional dissociation (HCD) at 65% collision energy. The scan range consisted of 100 to 500 m/z with a maximum injection time of 150 ms, AGC target of 100,000 (200%), and eight precursors being isolated.

Protein Inference

Protein groups consistent with identified peptides were inferred with Proteome Discoverer 2.5 (Thermo). Raw files were searched against the UniProt 2022 human proteome database consisting of 20,324 sequences. Mass tolerance of 10 ppm was used for precursor ions and 0.6 Da for fragment ions. The search included cysteine carbamidomethylation (57.0215 Da) as a static modification. Peptide N-terminal and lysine TMTpro 18plex modification (304.2071 Da), protein N-terminal acetylation (42.0106 Da), and methionine oxidation (15.9949 Da) were set as dynamic modifications. Up to two missed cleavages were allowed for trypsin digestion. The peptide false discovery rate (FDR) was set as <1% using Percolator (108) with a reversed-sequence decoy database (109). Only peptide level FDR was applied in the search parameters. For protein identification, at least one peptide with a minimum six amino acid length was required. A total of 339 decoys and 6029 target proteins were detected for a protein level FDR of 5.6%. Peptide spectral matches (PSMs), peptide groups, and inferred proteins resulting from Proteome Discoverer analysis are presented in supplemental Tables S1–S3, respectively. Grouped abundance for proteins from each TMT channel in supplemental Table S3 was used for bioinformatic analysis in Excel.

Bioinformatic Analysis

Grouped abundance of proteins was first median normalized within each TMT channel, after which biological replicates were averaged across TMT channels. Individual gRNA cell lines were treated as pseudo-technical replicates and averaged together for analysis. Raw quantification values approximated a negative binomial distribution before and after median normalization and were $\log_2$ transformed to approximate a normal distribution. The difference between $\log_2$ quantified conditions of interest for individual proteins was converted to a Z-score to enforce normality for statistical analysis, and a

corresponding two-tailed p -value was calculated. The Z-score was calculated as x/σ , where x = difference in $\log_2$ (median normalized protein quantification) and σ = standard error of the mean. Quantification error of the SEM for the $\log_2$ transformation was propagated via the delta method ($d/dx \log_2(\sigma) = \sigma^*(1/y*\ln(2))$, where y = median normalized quantification). Error of summation was propagated in the standard way (square root of the sum of squared errors). Proteins that were not quantified were excluded from analysis (965 total). Proteins that were not detected across all four conditions (NoG-, NoG+, pooled gRNAs, and sample pool of all replicates) were excluded from analysis to prevent undefined ratios after manual inspection to assess potential biological significance, removing a total of 49 proteins. A contaminant list was created by comparing NoG- to the pooled labeling conditions (all gRNAs + NoG+) and proteins enriched by 1.2FC and Bonferroni adj. $p < 0.05$ were excluded from downstream analysis. Differences between pooled on-target conditions versus NoG+ from the resulting list were then analyzed. Proteins with a FC > 1.2 and a Storey- q value < 0.05 were considered enriched. Calculated p -values of zero were replaced with the lowest calculated p -value to prevent an undefined value in the volcano plot.

Gene Ontology Analysis

Gene ontology analysis was performed with DAVID, GORilla, and Pantherdb. For each tool, the list of 6029 inferred proteins was uploaded as a background list rather than using the whole human proteome to test for GO enrichment. Proteins enriched by FC > 1.2 in on target versus NoG + conditions with (373) or without (775) the q -value filter were tested for enrichment of GO Biological Process, GO Cellular Component, and GO Molecular Function in each tool. For PantherdbGO, GO complete terms were searched, and Fisher's exact test with Benjamini-Hochberg correction was used to determine significance. For DAVID, the top three Functional Annotation Clusters were recorded alongside the Functional Annotation Chart with Benjamini-Hochberg correction used for all comparisons. Benjamini-Hochberg correction was used for GORilla analysis in each category. Results are presented in [supplemental Table S4](#).

Predicted TF Binding

For TFBSPred, the first TSS of FOXP2 (hg 38:7:114086327(+) - 114087534) was chosen for analysis (accessed August 2024). Every box was checked for lineage, tissue, karyotype, and sex. Every cell line was chosen for analysis, and the default TFFM (TF Flexible Model) search threshold of 0.90 was used. Every predicted TF present within the whole sequence was compared against our mass spectrometry results. For PROMO, *homo sapiens* was selected for the species, and the same sequence analyzed by TFBSPred was used for input (accessed August 2024). All predicted TFs were recorded and compared to mass spectrometry data in [supplemental Table S6](#).

Comparison with ENCODE ChIP-Seq Database

All published TF and histone experiments that passed ENCODE data quality standards in the ChIP-Seq data matrix performed in HEK293 or HEK293T cells available at encodeproject.org were analyzed. The ENCSR experiment identifiers and associated laboratories for the 223 analyzed experiments are recorded in [supplemental Table S7](#). To determine if a factor binds to the FOXP2 genomic locus, the web-based genome browser tool for each experiment was utilized. Only signals that were called as significant by ENCODE in the combined replicate experiments that reached the irreproducible discovery rate threshold were considered as true hits—a significant signal in one or both individual replicates that did not persist in the combined experiments was discarded. Each experiment was analyzed at the following genomic loci (hg38 chr7): TSS1: 114084641-114087534;

e330: 114411165-114418591; upstream intergenic region: 114075920-114084640; FOXP2 gene body: 114086327-114693773. Any peaks that passed the irreproducible discovery rate threshold within the analyzed region were recorded as a binary Yes/No in [supplemental Table S7](#). Multiple significant peaks within a defined analyzed region were individually counted for the $n > 3$ condition. Locus specificity of binding was determined by analyzing off target regions at the following genomic loci: GAPDH (hg38 chr12): 6533595-6536613; IRX6 (hg38 chr16): 55322957-55326110; MAGI1 (hg38 chr3): 66037224-66040189; and NANOG (hg38 chr12): 7787127-7789975.

Statistical Analysis

To assess statistical significance of enrichment values for ChIP-qPCR, expression changes for siRNA, and mean expression for MS detected versus not detected TFs, a two-sided t test was performed. To determine enrichment of TFs and components of the spliceosome, the χ^2 test was performed on pairwise comparisons after creating a contingency table. For protein quantification, the two-tailed p -value corresponding to the Z-score calculated for difference in protein quantification between states of interest was calculated. To correct for multiple hypothesis testing across the thousands of quantified proteins, Bonferroni or Storey methods were used as indicated in the text. For Bonferroni correction, the significance threshold was set at α/n , where α = uncorrected significance threshold (set to 0.05) and n = number of comparisons. Storey correction is a modification of Benjamini-Hochberg correction (110) where the proportion of null hypotheses is estimated to be less conservative in calling significance. The results from the n statistical tests were listed in ascending order of p -value and assigned a rank order of 1,2,...,n. The proportion of null hypotheses, π_0 , was estimated according to the formula $\pi_0 = (\#p\text{-values} > \lambda) / (n * [1 - \lambda])$, where $\lambda = 0.4$ as estimated from the flat portion of the p -value histogram. The value λ represents uniform distribution of truly null values considering the fact that truly significant values will skew toward lower p -values. The q -value for element i was calculated as the lower of $(\pi_0 * n * p\text{-value}_i) / \text{rank}_i$ and $q\text{-value}(i + 1)$. Cochran's Q test was calculated according to the formula $(k * [k - 1]) * (\sum \{ \text{column total} - N/k \}^2 / \sum \{ \text{row total} * [k - \text{row total}] \})$, where k = number of loci and N = grand total, with genomic loci in columns, TFs in rows, ENCODE positive = 1, and ENCODE negative = 0 and compared to the χ^2 distribution. McNemar's test on pairwise comparisons was calculated according to the formula $(b - c)^2 / (b + c)$, where b and c are discordant entries in the contingency table (1Y2N and 1N2Y) and compared to the χ^2 distribution with Bonferroni correction.

RESULTS

Preparing the FOXP2 Promoter-Proximal Proteome for Mass Spectrometric Analysis by Genetically Targeted Proximity Labeling

The experimental approach is outlined in [Figure 1A](#) and used a modified protocol. We began our investigation by generating stable HEK293 cells expressing dox-inducible dCas9-APEX2. Only a small fraction of cells responded to doxycycline treatment after puromycin selection, so we used FACS to generate a clonal population to minimize background signal from unlabeled chromatin in nonexpressing cells for downstream experiments ([supplemental Fig. S1A](#)). After confirming competence in proximity labeling ([Fig. 1B](#)), we generated three stable cell lines each expressing a different gRNA within the active FOXP2 promoter region. We performed ChIP-qPCR against the FLAG tag on the proximity

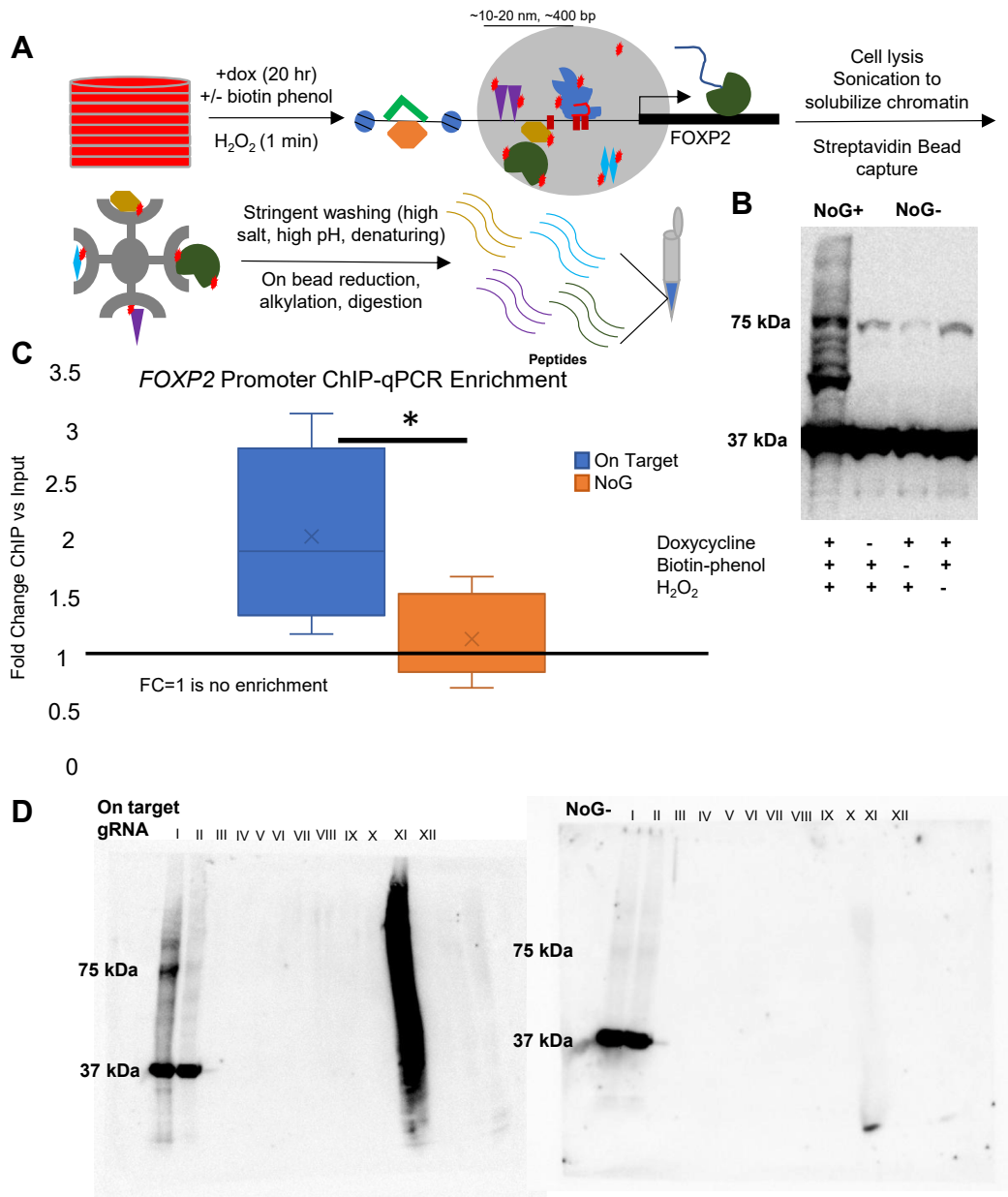


FIG. 1. Preparing the *FOXP2* promoter-proximal proteome for mass spectrometric analysis by genetically targeted proximity labeling. A, schematic overview of experimental approach for promoter-pulldown proteomics. A large number of input cells (8 15-cm² plates per condition per replicate) are used to generate sufficient material for MS analysis. Activation of genetically targeted proximity labeling by dCas9-APEX2 creates a cloud of reactive phenoxyl radicals (gray circle) that covalently biotinylate proteins (red stars) within 10 to 20 nm of the promoter targeted by the designed sgRNAs (designated by red boxes). Sonication to solubilize chromatin enables capture of biotinylated proteins by streptavidin magnetic beads. Stringent washing removes background contaminants before on-bead digestion for mass spectrometric analysis of generated peptides. B, Western blot showing proximity labeling positive and negative controls. All samples show the GAPDH loading control and endogenous 75 kDa biotinylated proteins. Systematic exclusion of necessary chemicals prevents proximity labeling, visible as the smear in the condition with all reagents included. The condition with all labeling reagents and no gRNA (NoG+) and the condition with exclusion of biotin phenol (NoG-) are representative of negative controls in the MS experiment. C, enrichment of targeted *FOXP2* promoter using ChIP-qPCR by the 2^{-ΔΔCt} method. The region targeted by the sgRNAs was enriched by 2.06 ± 0.31, while the untargeted NoG + condition showed no enrichment (1.15 ± 0.15) compared to an off-target *GAPDH* control. **p* < 0.05, *n* = 6. D, Western blot demonstrating capture and retention of biotinylated proteins during stringent washing for targeted labeling (gRNA2, left) and untargeted no labeling (NoG-, right) conditions. GAPDH loading control is visible in both conditions, while biotin tagging visualized by Streptavidin-HRP is only present in labeling positive conditions. The biotinylated proteins from input are efficiently captured by the streptavidin beads and show no elution during wash steps. 1% of beads predigestion demonstrates labeled proteins remain captured, while the absence of signal in the final

labeling construct to confirm accurate targeting using the $2^{-\Delta\Delta Ct}$ approach to compare enrichment of targeted *versus* untargeted conditions relative to a background locus (103). The no gRNA negative control showed no enrichment after ChIP compared to an off target *GAPDH* control locus ($FC = 1.15 \pm 0.15$, where $FC = 1$ is no change) (Fig. 1C). By contrast, the pooled on-target gRNAs targeting the *FOXP2* promoter showed significant enrichment after ChIP ($FC = 2.06 \pm 0.31$, $p < 0.05$) (mean $\pm$ SEM, $n = 6$). This degree of enrichment is consistent with other locus-specific chromatin isolation qPCR results (40, 43, 44). We anticipate this to represent a lower bound of enrichment given that streptavidin-based purification can enrich on-target loci by up to 284-fold when compared head-to-head with FLAG-based enrichment (38, 111).

After confirming successful targeting to the *FOXP2* promoter, we covalently tagged the proximal proteome with biotin using standard proximity labeling conditions in which 1 min of H_2O_2 treatment catalyzed biotinylation of surface exposed tyrosine residues within an approximately 20 nm radius. We sonicated chromatin to an average size of ~450 bp to solubilize labeled chromatin proteins (supplemental Fig. S1, B and C) and captured biotinylated proteins with streptavidin magnetic beads. To reduce capture of nonspecific background, we increased the stringency of wash steps, replacing the standard 2M urea wash with 6M GdnCl. Proteins adopt a random coil conformation at 6M GdnCl, whereas the midpoint of unfolding is at 3M urea (112). Given that the half time for biotin-bound streptavidin unfolding in 6M GdnCl is 50 days (113), we hypothesized a wash with 6M GdnCl would remove more contaminants while maintaining capture of biotinylated proteins. Accordingly, we observed no loss of material eluting in flow through by Western blot before the final step (Fig. 1D). We performed on-bead denaturation, reduction, alkylation, and digestion to generate peptides for mass spectrometric analysis.

Extensive Online Fractionation and Quantitative Proteomics via 2D-LC-TMT-RTS-SPS-MS3

The full mass spectrometry data acquisition workflow is outlined in Figure 2A. In addition to the three individual gRNA cell lines tiling the *FOXP2* promoter, we included two negative controls: no gRNA with exclusion of biotin phenol to account for endogenously biotinylated proteins and nonspecific bead interactors to generate a high confidence contaminant list (NoG-) and a no gRNA condition with biotin phenol to control for background labeling (NoG+) (see Fig. 1B). Each condition was comprised of eight 15-cm² plates and had three independent biological replicates, comprising more than a billion

cells (10^9 , although only 60% represent on target chromatin). The five conditions for each biological replicate were processed in parallel. The 15 conditions plus a pool of equal amounts of each sample were combined together after labeling with TMTs for a 16-plex experiment, and 15 μ g of protein was injected for LC-MS.

In order to reduce the signal loss from repeated sample handling steps (114) while maintaining the benefits of deep proteome coverage from extensive fractionation, we utilized online 2D-LC directly connected to the mass spectrometer. A 12-fraction high-pH gradient separated peptides before transfer to a low pH nanoflow column (supplemental Fig. S2). Using RTS-SPS-MS3 for quantitative proteomics (see methods), we detected 49,570 PSMs *via* database search (Supplemental Table S1). We utilized Proteome Discoverer to infer proteins consistent with the 31,835 identified peptide groups (Supplemental Table S2), generating a list of 6029 proteins (Fig. 2B) (Supplemental Table S3). Since this was a discovery-based effort aimed at proteins expressed at a low level (e.g., TFs), we did not require proteins to be identified by multiple peptides to be included in downstream analysis.

We excluded nonquantified proteins from downstream analysis rather than impute missing values, resulting in a list of 5074 proteins. We first normalized the grouped abundance of each protein by the median value within a TMT channel (115) and then averaged biological replicates (Fig. 2C). Pairwise comparisons of global protein quantification across different conditions revealed the NoG-measurements poorly correlated with all conditions in which labeling occurred, as expected (Fig. 2D). In contrast, there was a high degree of correlation between the NoG+ conditions and the pooled on target conditions (Fig. 2E). We considered all proteins that were enriched by at least 1.2-fold in NoG- *versus* the pooled labeling conditions and an adjusted p -value < 0.05 to be contaminants and removed them from downstream analysis, resulting in 4377 proteins. We chose the FC threshold based on observations that isobaric tags can measure 1.2FC at the 10s-100s of femtomolar level in MS2-based quantification (116) compared with nM TF concentrations in eukaryotic cells. Our RTS-SPS-MS3 data acquisition is expected to have higher accuracy and precision (117). We used Bonferroni correction in generating the contaminant list in order to minimize the number of false positives identified as contaminants while correcting for multiple hypothesis testing (99).

FOXP2 Promoter Proximal Proteome Is Enriched for TFs and Splicing Machinery

To determine the proteins enriched at the active *FOXP2* promoter, we compared the quantification between the

lane indicates on-bead digestion went to completion. I: input; II: flow-through; III: RIPA wash 1; IV: RIPA wash 2; V: 1M KCl wash; VI: 100 mM Na_2CO_3 wash; VII: 6M GdnCl wash; VIII: denaturation, reduction, alkylation; IX: TEAB wash 1; X TEAB wash 4; XI: 1% bead elution predigestion; XII: remaining bead elution postdigestion. ChIP, chromatin immunoprecipitation.

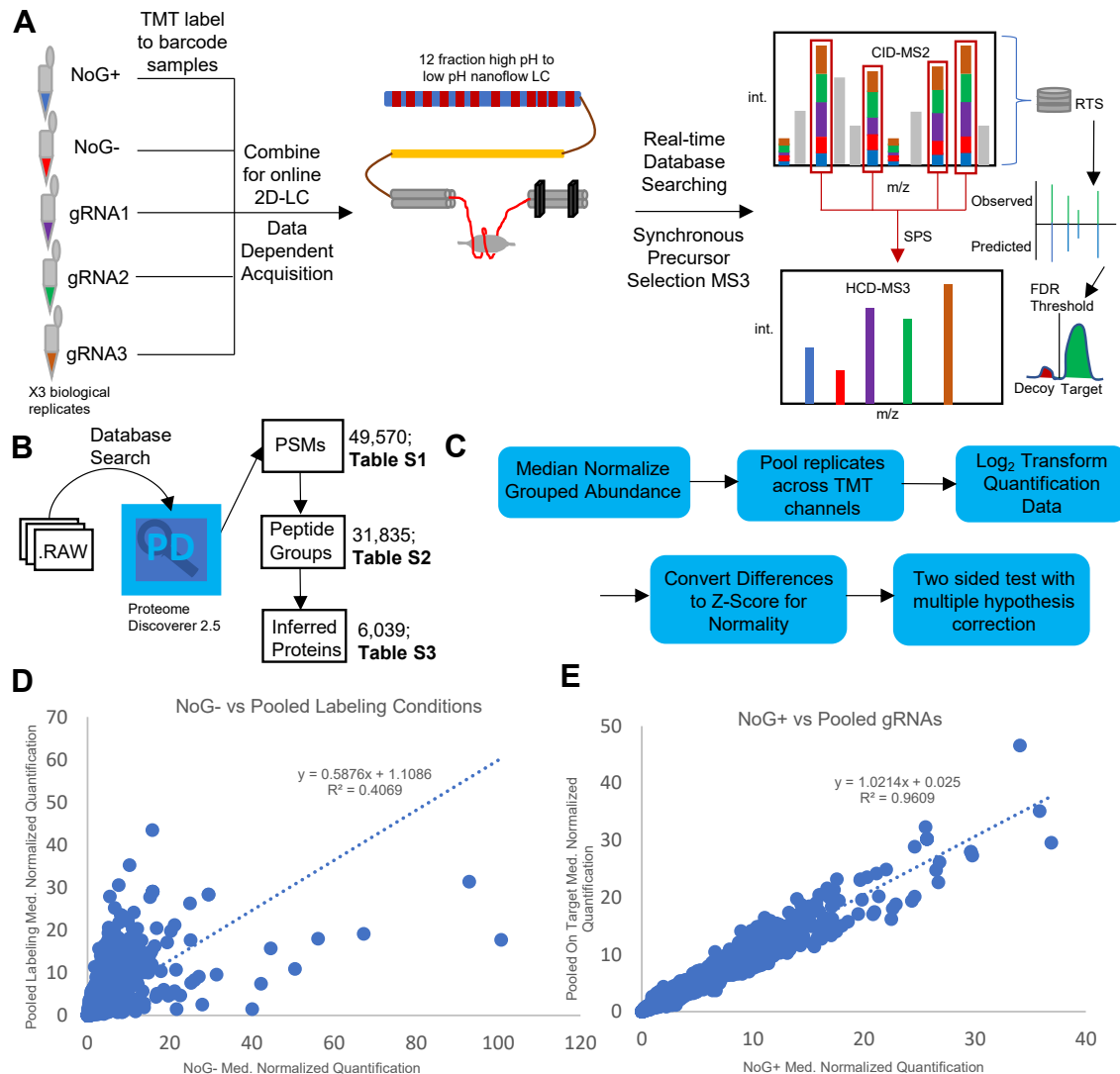


FIG. 2. Extensive online fractionation and quantitative proteomics via 2D-LC-TMT-RTS-SPS-MS3. A, overview of 2D-LC-TMT-RTS-SPS-MS3 mass spectrometry workflow. Three biological replicates representing five conditions (three on target sgRNAs, plus two negative controls) are barcoded with TMTs before combination and injection on the LC-MS. A high pH 12-fraction gradient followed by an analytical low pH column separates peptides by retention time before direct connection to the tribrid MS instrument. A quadrupole selects ions in a defined m/z range for a high resolution orbitrap scan to capture MS1. Ions are sent to the linear ion trap for MS2 and eventually return to the orbitrap for MS3 quantification. Fragments in MS2 are compared in real-time on the instrument to a protein database to ensure only signals that can be matched to peptides with high confidence are sent to MS3 to economize on the instrument duty cycle, and multiple precursors are selected. Peptide-level FDR is controlled with a decoy database strategy. B, raw files from the mass spectrometer were processed using a database search via Proteome Discoverer. Total number of identified PSMs (49,570; [supplemental Table S1](#)), peptide groups (31,835; [supplemental Table S2](#)), and inferred proteins (6029; [supplemental Table S3](#)) are indicated. C, bioinformatic workflow to analyze protein quantification. D, global pairwise protein quantification comparing NoG-against all conditions in which labeling occurred (pooled gRNAs + NoG+). There is low correlation in protein quantification ($R^2 = 0.41$) between NoG- and conditions where labeling occurred. E, global pairwise protein quantification comparing NoG + against pooled on target conditions. There is a high degree of correlation in protein quantification ($R^2 = 0.96$) between on target and untargeted conditions. FDR, false discovery rate; MS3, third mass spectrometry step; PSM, peptide spectral match; RTS, real-time database search; SPS, synchronous precursor selection; TMT, tandem mass tag.

pooled on-target conditions with the untargeted negative control (NoG+) using the Storey method to correct for multiple hypothesis testing (100, 101), generating 596 proteins with $q < 0.05$. Of these proteins, 373 were enriched with $FC > 1.2$. A

recent report utilizing dCas9-APEX2 to identify a promoter-proximal proteome identified biologically relevant proteins that were nonsignificantly enriched (40). Ignoring significance, 775 proteins were enriched with $FC > 1.2$. After the filtration

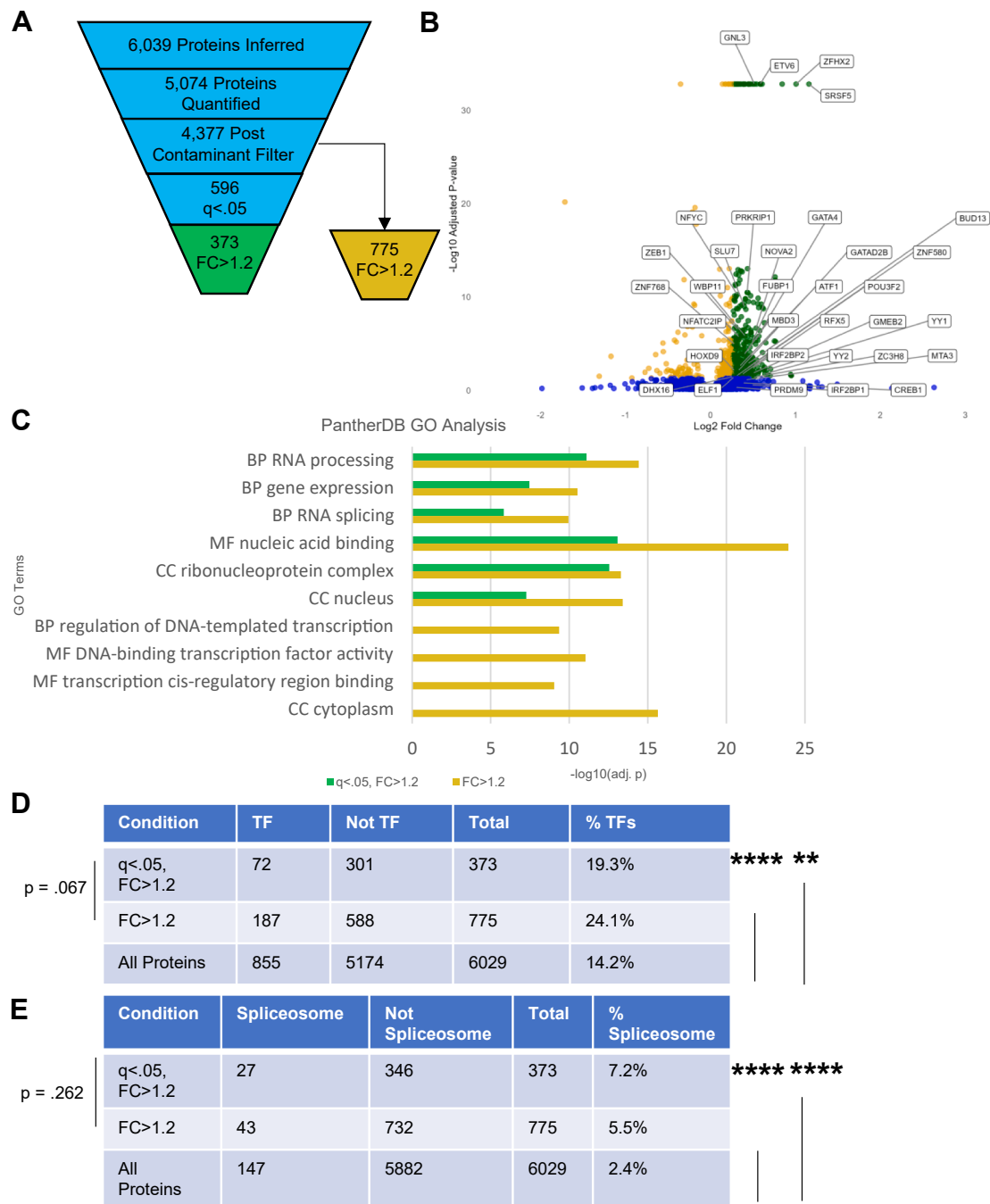


FIG. 3. **FOXP2 promoter proximal proteome is enriched for transcription factors and splicing machinery.** A, number of proteins remaining after each data analysis filter (quantification, NoG- contaminant list, *q*-value filter, and *FC* > 1.2). The number of proteins with *FC* > 1.2 post contaminant filter ignoring the *q*-value filter is also shown. B, volcano plot showing Log₂FC(pooled on target versus NoG+) on the x-axis and -log₁₀(*q*-value) on the y-axis. Proteins enriched with *FC* > 1.2 and *q* < 0.05 are highlighted in green, while proteins with *q* < 0.05 and *FC* < 1.2 are golden, and proteins with *q* > 0.05 are blue. Proteins presented in Table 1 are labeled. C, Gene Ontology enrichment analysis using PantherdbGO for proteins enriched by *FC* > 1.2 with (green) and without (golden) the *q*-value filter. D, contingency table for transcription factors identified in proteins enriched by *FC* > 1.2 with and without the *q*-value filter compared with all identified proteins. ***p* < 0.01 *****p* < 0.00001. E, contingency table for components of the spliceosome identified in proteins enriched by *FC* > 1.2 with and without the *q*-value filter compared with all identified proteins. *****p* < 0.00001. BP, biological process; CC, cellular component; FC, fold change; MF, molecular function; TF, transcription factor.

steps (Fig. 3A), we created a volcano plot to visualize the results (Fig. 3B).

We analyzed both sets of enriched proteins ($FC > 1.2$ with or without q -value filter) via gene ontology analysis using multiple tools (118–121). Representative results from PantherdbGO are shown in Figure 3C, and results from all tools with $-\log_{10}(\text{adjusted } p) > 6$ are presented in Supplemental Table S4. Expected terms such as nuclear localization and mRNA processing indicate nuclear enrichment was successful. RNA splicing and related terms like ribonucleoprotein complex were enriched as well, potentially indicative of cotranscriptional splicing regulation (122) or accumulation of dCas9-APEX2 in the nucleolus (42). There generally was greater statistical significance in terms identified in the list of proteins with $FC > 1.2$ without the q -value filter; this list also had greater enrichment of terms related to sequence-specific TF binding and activity of RNA polymerase II, but it also included likely false positive terms like cytoplasm. We conclude that analyzing proteins enriched below statistical significance can unveil true promoter proximal proteins, but that results must be interpreted with caution and validated by orthogonal approaches.

In order to determine if we were able to successfully enrich TFs at the *FOXP2* locus, we compared our enriched proteins to the set of all human TFs. We first combined two previously compiled lists of human TFs and removed duplicates, generating a set of 2562 proteins (Supplemental Table S5) (123, 124). We then generated a contingency table to determine if the fraction of TFs in the enriched lists were statistically different from the entire set of proteins inferred in our mass spectrometry experiment (Fig. 3D). The fraction of TFs for the enriched proteins with ($q < 0.05$, $FC > 1.2$; 19.3%) and without the q -value filter ($FC > 1.2$; 24.1%) was significantly different from the total list of proteins (14.2%) by the χ^2 test ($p < 0.01$ and $p < 0.00001$, respectively). The percent of components of the spliceosome (125) in the enriched lists with ($q < 0.05$, $FC > 1.2$; 7.3%, $p < 0.00001$) and without the q -value filter ($FC > 1.2$; 5.5%, $p < 0.00001$) was also significantly different from the spliceosome components in the whole set of inferred proteins (2.4%) (Fig. 3E) (Supplemental Table S5).

Within the set of significantly enriched TFs, we detected POU3F2, which has been shown to drive reporter gene expression from an intronic *FOXP2* enhancer and has a binding site highly conserved across vertebrates upstream of TSS1 (126, 127). We conclude that genetically targeted proximity labeling is able to successfully detect promoter-proximal proteomes, including classic sequence-specific TFs.

Orthogonal Confirmation of Identified Proteins

Selected proteins enriched by $FC > 1.2$ at TSS1 are indicated in Table 1. To validate our findings, we utilized orthogonal approaches to benchmark our candidate transcriptional regulators identified by mass spectrometry.

Computationally Predicted TF Binding to the FOXP2 Locus

We first used two separate computational tools that predict potential TF binding sites at DNA regions, TFBSPred (128) and PROMO (129, 130). TFBSPred predicted potential binding of the mass spectrometry-identified hits NFYA*, NFYC*, KLF9*, RFX5*, ATF1*, CREB1*, GMEB2*, VEZF1*, CEBPA*, and NR2F2* (* $q < 0.05$, $FC > 1.2$). The protein NFATC1 was also predicted by TFBSPred. Both NFATC1 and NFATC2 have been shown to impact β -cell proliferation through a mechanism relying on FOXP proteins from a FOXP1/2/4 triple knockout model (131). Interestingly, although NFATC1 quantification was not significant and unchanged ($q > 0.05$, $FC = 1.05$) and we did not detect NFATC2, the related NFATC2IP* was significantly enriched.

The mass spectrometry-identified hits NFIX*, MAZ*, NFYC*, NFYA*, YY1*, ATF1*, CREB1*, HOXD9*, and ELF1* were identified as potential binders by the PROMO web tool. Our mass spectrometry results detected the PROMO-predicted protein IRF2, but it was not enriched. However, related proteins IRF2BP1* and IRF2BP2* were both enriched via quantitative mass spectrometry (1.40FC and 1.30FC, respectively). The output of predicted binders from both tools is included as Supplemental Table S6. We note that the tools considered all possible human TFs across all cell types, including factors not expressed in HEK293 cells. Therefore, we would not expect to have high coverage of the predicted binders.

The known coregulators ATF1 and CREB1 were identified by both computational tools and our mass spectrometry results. Consistent with these observations, a ChIP-chip analysis of CREB1 in HEK cells showed binding at the *FOXP2* locus (132). Notably, the sequence in the microarray analysis was included before identification of the active upstream promoter targeted in our experiments. It instead represents a downstream region comprising two alternative promoters and a conserved enhancer that has been demonstrated to form an active chromatin loop with TSS1 in HEK293 cells (86) (*vide infra*).

ENCODE ChIP-Seq Database

We sought to further validate our mass spectrometry results by identifying true positive proteins that bind to the *FOXP2* promoter in HEK293 cells. The gold standard for identifying protein–DNA interactions in cells is ChIP-Seq and related approaches. The ENCODE database has a collection of ChIP-Seq experiments targeting a diversity of TFs across many cell lines (Fig. 4A). We analyzed the 223 ChIP-Seq experiments performed in HEK293(T) cells to determine TF and histone post-translational modification status at the *FOXP2* gene (Fig. 4B). We analyzed binding of factors at the 2893 bp region of TSS1 that has been validated as an active promoter by PolII binding and luciferase assays. A chromatin conformation capture (3C) study (86) has demonstrated long range

TABLE 1
Selected mass spectrometry-enriched proteins at FOXP2 TSS1

Protein	Class	Notes
POU3F2	TF	^a Conserved binding site upstream of TSS1 (126)
GNL3	TF	^a Member of Wnt/ β -catenin pathway (143)
NFYC	TF	^a ^b
GMEB2	TF	^a ^b
RFX5	TF	^a ^b
HOXD9	TF	^a ^b
ELF1	TF	^a ^b
ETV6	TF	^a
ZEB1	TF	^a
GATA4	TF	^a
FUBP1	TF	^a
ZC3H8	TF	^a
ATF1	TF	^a ^b Co-binds with CREB1
CREB1	TF	^a ^b Co-binds with ATF1, binds to FOXP2 e330 in ChIP-chip analysis (132)
YY1	TF	^a ^b ENCODE ChIP-Seq peak at TSS1
YY2	TF	^a ENCODE ChIP-Seq peak at TSS1
ZFHX2	TF	^a ENCODE ChIP-Seq peak at TSS1
ZNF768	TF	^a ENCODE ChIP-Seq peak at TSS1
ZNF580	TF	^a ENCODE ChIP-Seq peak at TSS1
ZEB1	TF	^a ENCODE ChIP-Seq peak at TSS1
NFATC2IP	TF	^a NFATC1 predicted by TFBSPred; NFATC proteins (1 and 2) impact β -cell proliferation through FOXP-dependent mechanism (131)
IRF2BP2	TF	^a Partner protein IRF2 predicted by PROMO
IRF2BP1	TF	^a Partner protein IRF2 predicted by PROMO
SRSF5	Spliceosome	^a
PRKRIP1	Spliceosome	^a
SLU7	Spliceosome	^a
NOVA2	Spliceosome	^a
WBP11	Spliceosome	^a
BUD13	Spliceosome	^a
DHX16	Spliceosome	^a
PRDM9	Chromatin remodeler	^a Validated H3K4Me ₂ $\rightarrow$ H3K4Me ₃ activity, consistent with promoter chromatin state
GATAD2B	Chromatin remodeler	^a NuRD complex member
MBD2	Chromatin remodeler	^a NuRD complex member
MBD3	Chromatin remodeler	^a NuRD complex member
MTA3	Chromatin remodeler	^a NuRD complex member

^a $q < .05$, ^aFC > 1.2.

^bPredicted by computational TF tool.

contact between TSS1 and the 7426 bp region that comprises TSS2, TSS3, and a highly conserved enhancer 330 kbp away from TSS1, which contains the sequence included in the microarray analysis referenced above; we therefore included 7.4 kbp regulatory region (hereafter: e330) in our analysis. We also considered the intergenic region upstream of TSS1 between *FOXP2* and *PPP1R3A*, which includes a purported human-specific enhancer upstream of TSS1, and whether there were more than three significant peaks in the *FOXP2* gene body ($n > 3$) to more comprehensively characterize chromatin interactors at the genetic locus in HEK293 cells. We compared the presence of significant ChIP-Seq signals at these loci in the ENCODE database with our mass spectrometry data in Supplemental Table S7.

Our targeted promoter showed H3K4Me₃ and H3K27Ac histone marks, consistent with an active TSS (133). These

histone modifications were present at e330 in addition to H3K4Me₁, in line with an active enhancer region. Consistent with the observed ChIP-Seq-based chromatin state of our targeted promoter, we detected by MS the nonsignificantly enriched chromatin remodeler PRDM9^a, which has validated H3K4Me₂ $\rightarrow$ H3K4Me₃ methyltransferase activity (134). We also detected members of the nucleosome remodeling and histone deacetylation (NuRD) complex GATAD2B^a, MBD2^a, MBD3^a, and MTA3^a (135). The NuRD complex has been demonstrated to be associated with H3K4Me₃ histone modifications at poised bivalent sites and to bind at the promoter of active genes (136, 137). By contrast, we failed to detect enrichment of multiple members of the SWI/SNF, ISWI, and INO80 families of chromatin remodelers, suggesting the NuRD complex may regulate the epigenetic state of the *FOXP2* promoter in HEK293 cells.

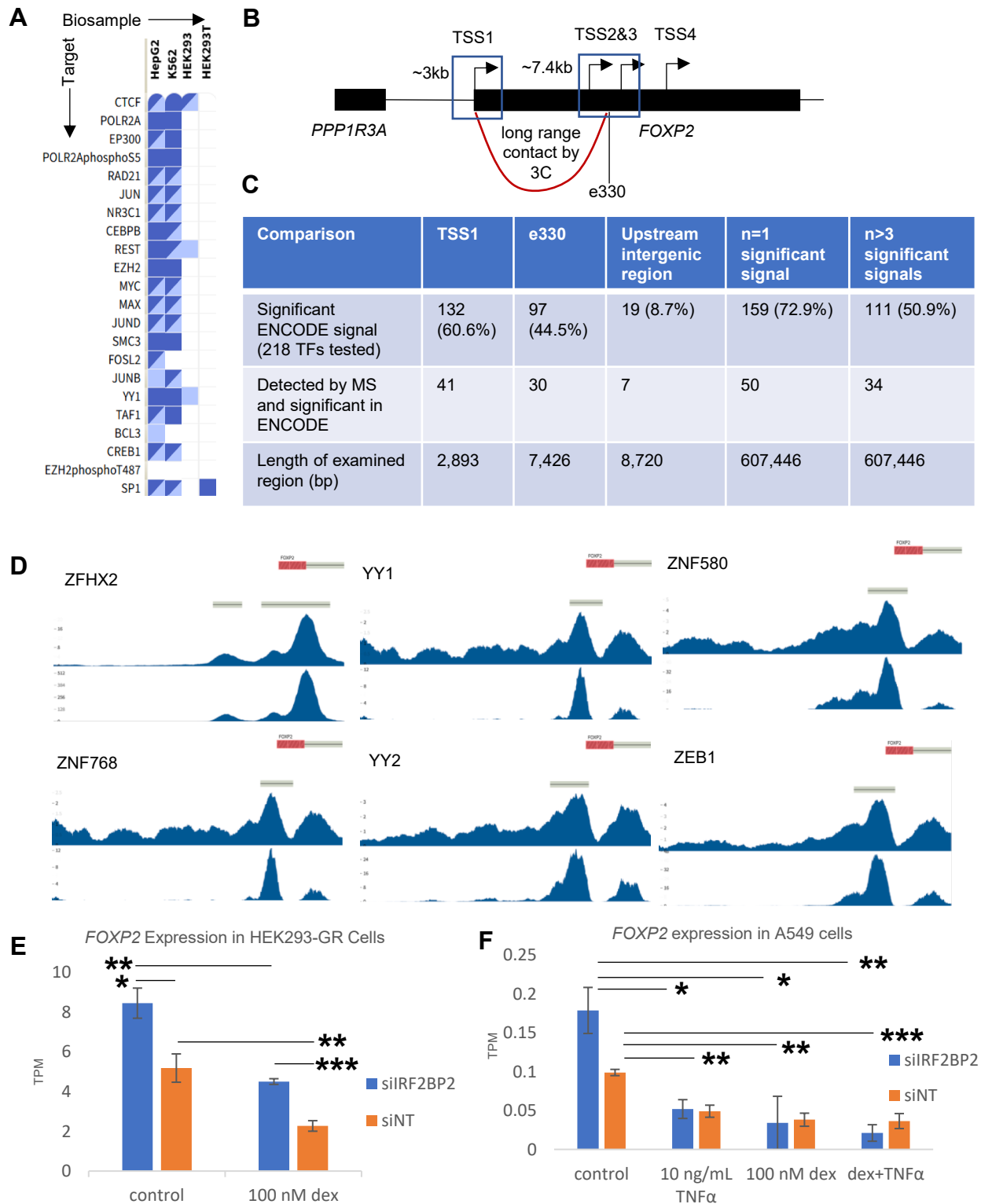


FIG. 4. Orthogonal confirmation of identified proteins. A, ENCODE ChIP-Seq Data Matrix available at encodeproject.org. Filtered to show transcription factors in human cell lines HepG2, K562, HEK293, and HEK293T. B, gene body diagram showing *FOXP2* genomic regions examined in ENCODE ChIP-Seq database. TSS1 = hg38 chr7: 114084641-114087534; e330 = hg38 chr7:114411165-114418591, comprising TSS2&3 and a conserved enhancer. A 3C study has indicated these regions form a long range chromatin loop (86), indicated by a red line. C, number of transcription factors bound at each indicated region of *FOXP2* from ENCODE ChIP-Seq database and corresponding mass spectrometry results. See also [Supplemental Table S7](#) for detailed comparison of ENCODE results to mass spectrometry results. D, ENCODE

Of the 218 TFs studied in HEK cells by ENCODE, 132 (60.6%) showed significant ChIP-Seq signals at TSS1 (Fig. 4C). At e330, 97 (44.5%) TFs showed a significant signal. Only 19 (8.7%) showed binding in the upstream intergenic region between TSS1 and *PPP1R3A*. There were 159 factors (72.9%) with significant signals in the *FOXP2* gene body, and only 111 (50.9%) had more than three significant signals. Most of the TFs detected at e330 overlapped with those found at TSS1, with the exception of 12 proteins that had no promoter peak but were bound to the enhancer (FOXA1, TRIM28, ZNF384, MEIS1, ZFP3, ZNF34, ZNF362, ZNF624, AEBP2, GFI1B, GLI2, and ZNF654). Of the 218 TFs that have ChIP-Seq experiments in HEK293 cells in the ENCODE database, we detected 72 by mass spectrometry. We detected 41 of the TFs with ChIP-Seq signals at TSS1 in our mass spectrometry dataset, suggesting that 91 ChIP-Seq positive TFs in the ENCODE database (68.9%) were undetected as false negatives. Fifty-seven percent of the mass spectrometry detected TFs with ENCODE data (41/72) were true positive binders at TSS1 before performing bioinformatic analysis on the dataset.

Of the 72 ENCODE TFs identified in our data set of 6029 detected proteins, only 58 were enriched in labeled conditions versus NoG-; the majority of proteins eliminated (12/14, 86%) were not quantified at all, indicating they were present near the detection limit of the instrument rather than being enriched in NoG-. There were 13 TFs with $q < 0.05$, of which 11 had $FC > 1.2$. Only eight of those 13 (61.5%) had ChIP-Seq signals, and the two eliminated by the $FC > 1.2$ filter (TARDBP and PKNOX1) were true positives in ENCODE ChIP-Seq. The ChIP-Seq tracks of the six proteins with $q < 0.05$, $FC > 1.2$ and significant ENCODE ChIP-Seq signals are shown in Figure 4D. Comparing ENCODE results to the list of all proteins with $FC > 1.2$ without the q -filter, 25 of 57 were enriched, but only 14 showed signals at TSS1 (56%). Of the 11 false positives, there were three TFs with peaks at e330 (TRIM28, ZNF384, and MEIS1).

Comparison to Known Regulators of FOXP2 Expression

We further compared our mass spectrometry data and results from the ENCODE database with the few known regulators of *FOXP2* expression. One of the most well-known and best characterized regulators of *FOXP2* expression is the Wnt/ β -catenin TF LEF1. Experiments in zebrafish embryos have

demonstrated that *lef1* directly binds to and regulates the expression of *foxp2* (138). In contrast, LEF1 was detected and passed the NoG-contaminant filter, indicating nuclear expression, but it was not enriched in our targeted mass spectrometry results. Furthermore, ENCODE showed no significant ChIP-Seq signal at TSS1 or e330 in HEK293 cells. This result highlights the need for care in transferring observations of transcriptional regulation across cell types. On the other hand, the TCF/LEF TF family member TCF7L2 showed a significant ChIP-Seq signal at TSS1 and e330. Although we detected TCF7L2 in our inferred proteins, it was not significant and was more enriched in NoG + conditions compared to on target. However, analysis of the PSMs showed that the only identified peptide was shared with LEF1. This observation highlights the difficulty in the protein inference problem for which there are multiple approaches and no commonly accepted best practice (139–142). It is noteworthy that one of the top hits in our list (top 10 FC versus NoG + for $q < 0.05$), GNL3, is a component of the Wnt/ β -catenin signaling pathway (143).

The TF FOXP2 was predicted to bind to the *FOXP2* promoter by TFBSPred and showed a significant signal in the ENCODE ChIP-Seq database at TSS1. However, there was no difference between on and off-target conditions in our quantitative dataset ($FC = 1.02$). Furthermore, the protein ZBTB20 showed a significant ChIP-Seq signal, and the mouse homolog Zbtb20 has been shown to bind to and control *FoxP2* expression in the developing mouse brain (144), but we failed to detect it in our mass spectrometry dataset. The homolog of the human protein PAX6 has been identified as binding to the *foxp2* locus and impacting gene expression in zebrafish models (145). We detected this protein, but it was not enriched in on target conditions ($FC = 1.05$), although it was not tested in HEK293 cells by ENCODE. Each of these observations highlights that a lack of detection or enrichment in our mass spectrometry dataset does not preclude a protein from being a member of the *FOXP2* promoter-proximal proteome.

IRF2BP2 and Glucocorticoid Signaling Are Negative Regulators of FOXP2 Transcription

Given that there can be limited overlap between TF binding and regulatory activity (146), we sought to demonstrate a direct impact on transcriptional output rather than simply a binding interaction at the *FOXP2* promoter from within our

ChIP-Seq tracks of the six proteins with called peaks that were enriched in the mass spectrometry dataset ($q < 0.05$, $FC > 1.2$). Each track shows the *FOXP2* gene body, the irreproducible discovery rate called peak thresholds, the fold change of the indicated transcription factor over control, and the associated p -value. Each transcription factor has its own data-dependent scale bar for fold change and p -value. E, expression of *FOXP2* in transcripts per million (TPM) in HEK293 cells constitutively expressing the glucocorticoid receptor (GR). Expression of *FOXP2* transcripts is increased upon siRNA knockdown of IRF2BP2 and decreased upon activation of GR signaling by dexamethasone. Error bars = SEM, $n = 4$. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. F, expression of *FOXP2* in TPM in A549 cells with siRNA knockdown of IRF2BP2 or nontargeting (NT) with or without activation of GR and/or TNF α signaling. Error bars = SEM, $n = 3$. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 3C, chromatin conformation capture; ChIP, chromatin immunoprecipitation; ENCODE, Encyclopedia of DNA Elements; TF, transcription factor; TSS, transcription start site.

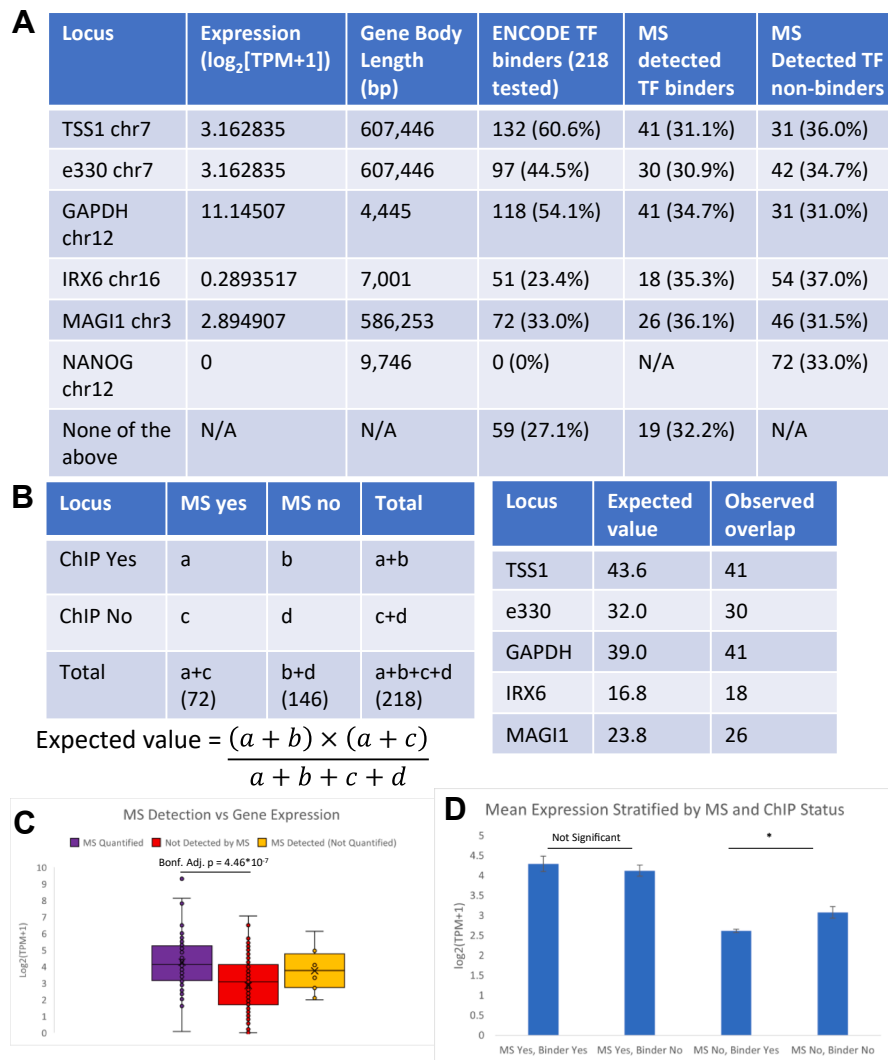


FIG. 5. Assessment of proteomics locus specificity by off target locus ENCODE comparison. A, expression, ENCODE TF binding status in HEK293 cells, and MS TF detection status for FOXP2 TSS1 and e330 and promoters of GAPDH, IRX6, MAGI1, and NANOG. B, formula for calculating expected value for overlap of ChIP- and MS-positive TFs with expected *versus* observed values at each locus. C, mean gene expression for proteins that MS detected and quantified (n = 60), did not detect (n = 146), or detected but did not quantify (n = 12). Quantified *versus* undetected Bonf. Adj. $p = 4.46 \times 10^{-7}$. D, average expression levels measured in log₂(TPM+1) pooled across loci stratified by MS detection and ChIP-Seq binding status. *Bonf. Adj. $p < 0.05$.

proposed candidate regulators. The protein IRF2BP2 was significantly enriched in our proteomic assay and has been identified as an IRF2-dependent transcriptional repressor (147, 148). Given that IRF2BP1 was also enriched (nonsignificantly) and IRF2 was predicted from the PROMO binding tool, we sought to characterize effects of IRF2BP2 on FOXP2 expression. Characterization of IRF2BP2 genome-wide binding and effects of its depletion on transcription have been reported previously in HEK cells that constitutively express the glucocorticoid receptor (149).

As shown in Figure 4E, siRNA knockdown of IRF2BP2 leads to a significant 1.6-fold increase in expression of FOXP2

compared with a nontargeting control (siNT). This effect persisted when glucocorticoid signaling was activated by addition of 100 nM dexamethasone. Notably, activation of glucocorticoid signaling by dexamethasone significantly decreased FOXP2 expression in both siIRF2BP2 and siNT conditions. There were three identified peaks in the FOXP2 gene body for IRF2BP2 binding *via* ChIP-Seq in control conditions and only one peak with dexamethasone treatment, although the control peaks sat between TSS1 and e330 and did not include our targeted promoter. Conspicuously, one of the three peaks in control conditions resides approximately 56 kbp downstream of the targeted promoter, a region that

forms a weak chromatin loop (*i.e.*, close spatial relationship) with TSS1 across neuronal cell types in 3C experiments (86). The peak in dexamethasone-treated cells was more than 400 kbp downstream of TSS1 and did not have evidence for long range chromatin looping in any cell types since it was outside the interrogated enhancer-promoter pairs in the 3C study.

To determine whether these observations were translatable to other cellular contexts, we compared *FOXP2* expression before and after knockdown of IRF2BP2 in the lung cancer cell line A549 from the same study (Fig. 4F). *FOXP2* is only weakly expressed at the transcript and protein levels in A549 cells, lowering power to call statistical significance (note scale bar differences between 4E and 4F). This down-regulation is tied to increased aggressiveness of lung cancer (150). Knockdown of IRF2BP2 led to a 1.8-fold increase in *FOXP2* expression that approached but did not reach significance at the $p < 0.05$ level compared to siNT conditions ($p = 0.0558$). Treatment with dexamethasone, TNF α , or a combination of both significantly decreased *FOXP2* expression. These changes did not differ between siIRF2BP2 and siNT conditions ($p > 0.3$), and there were no detectable differences among the individual or combined treatments. We conclude that IRF2BP2 and glucocorticoid signaling can negatively regulate *FOXP2* expression across cell types.

Assessment of Proteomics Locus Specificity by Off Target Locus ENCODE Comparison

To ascertain the locus specificity of our proteomics data, we repeated our ENCODE analysis at several off target regions and compared the results to our MS-detected proteins. Since long-range DNA contacts are typically intra-chromosomal, we chose genomic loci outside chromosome 7 that, to the best of our knowledge, have no implications for co-regulation with *FOXP2* based on the literature to date. Each analyzed locus was approximately 3 kbp in length and comprised the furthest upstream promoter of a gene to align with the analysis of *FOXP2* TSS1.

We chose genes to represent a variety of transcriptional states. Accordingly, the transcript per million (TPM)-based expression levels in HEK293 cells ($\log_2[\text{TPM}+1]$) according to DepMap (151) ranged from silent heterochromatin in *NANOG* (0) to lowly transcribed *IRX6* (0.2893517) to moderately transcribed *MAGI1* (2.894907) to the house-keeping *GAPDH* (11.14507) as compared to the moderately transcribed *FOXP2* (3.162835). In comparison to the 132 ENCODE TFs binding to *FOXP2* TSS1 (60.6%) and 97 TFs binding e330 (44.5%), 0 TFs bind *NANOG* (0%), 51 bind *IRX6* (23.4%), 72 bind *MAGI1* (33.0%), and 118 bind *GAPDH* (54.1%) (Fig. 5A).

We first performed a locus specificity analysis based on the ENCODE ChIP-Seq data. Global analysis of binary binding data at each locus showed a significant difference in TF

binding proportion for at least one locus, even when removing the no-binder *NANOG* condition (Cochran's $Q(\text{df}, Q(5)) = 306.9$, $p = 3.215 \times 10^{-64}$; $Q(4) = 129.1$, $p = 6.207 \times 10^{-27}$ removing *NANOG*) (supplemental Table S8). We found that the proportion of TF binders at each locus in the ENCODE ChIP-Seq data showed significant difference for all pairwise locus comparisons except for TSS1-*GAPDH* (Bonf. Adj. $p = 0.921$), e330-*GAPDH* (Bonf. Adj. $p = 0.122$), and e330-*MAGI1* (Bonf. Adj. $p = 0.0584$) in follow-up McNemar's tests (Supplemental Table S8).

We created contingency tables for each locus to determine if the 72 TFs detected by MS with ENCODE ChIP-Seq data showed enrichment for binders in either on target or off target conditions (Supplemental Table S8). The expected overlap between MS detection and ChIP-Seq positive TFs did not differ significantly at any analyzed locus (Fig. 5B) (Fisher's Exact test, $p > 0.4$ for all loci). We detected approximately a third of TFs with ENCODE data whether looking at locus specific binders (range 30.9–36.1%), locus-specific non-binders (range 31.0–37.0%), or TFs that did not bind any of the analyzed loci (32.2%) (Fig. 5A).

We hypothesized that the similar proportion of detection across all analyzed loci reflected both inherent MS sensitivity limits and underlying biology. We compared the expression levels for ENCODE TFs based on DepMap $\log_2[\text{TPM}+1]$ in HEK293 cells that our MS data acquisition did not detect ($n = 146$), detected but did not quantify ($n = 12$), or detected and quantified ($n = 60$) (Fig. 5C). Proteins we did not detect by MS showed significantly lower expression (mean = 2.861387, SD = 1.6202) compared to those we detected and quantified (mean = 4.282906, SD = 1.644012) (Bonf. Adj. $p = 4.46 \times 10^{-7}$). There was not a significant difference in average expression for TFs that were detected but not quantified (mean = 3.748952, SD = 1.194371) versus TFs that were either quantified (Bonf. Adj. $p = 0.652$) or not detected (Bonf. Adj. $p = 0.109$). For the 13 lowest expression TFs we quantified by MS ($\log_2[\text{TPM}+1] < 3$), TSS1 showed the highest proportion of binders (9, 69.2%), followed by e330 and *GAPDH* (8, 61.5%), then *MAGI1* (6, 46.2%) and *IRX6* (4, 30.8%).

To assess locus specificity of our proteomics data in light of expression-based detection limitations, we compared averaged $\log_2[\text{TPM}+1]$ expression data for TFs at each locus stratified by MS-detection status and ChIP-Seq binding status, excluding *NANOG* due to a lack of binders (Fig. 5D) (Supplemental Table S8). Average expression across all loci for the four conditions significantly differed for all pairwise comparisons with a difference in MS detection status, as expected. TFs we detected by MS did not show a significant difference in expression based on ChIP-Seq binding status. On the other hand, TFs we did not detect by MS showed significantly lower expression comparing binders and non-binders (Bonf. Adj. $p < 0.05$). In a locus-specific analysis, mean expression for TFs not detected by MS showed lower

expression for binders compared to nonbinders only at the targeted TSS1 locus (Bonf. Adj. $p < 0.05$), while expression for other loci did not differ significantly with binding status (Bonf. Adj. $p > 0.1$).

DISCUSSION

The presence of multiple promoters within a gene permits regulatory systems that can incorporate complex logic operations for precise operation of single copy gene dosage (152). While synthetic biologists are just beginning to harness these principles for designed cellular phenotypes, evolution has continuously optimized gene circuits as long as they have existed (153). These tightly regulated genetic circuits can lead to discordance between mRNA and protein levels, emphasizing the need for proteomic measurements to accurately capture cell states (12–14). The *FOXP2* locus is an outlier within the genome, being enriched with more than 10 ultra-conserved noncoding regulatory regions conserved across vertebrates but not invertebrates (75) and harboring multiple alternative promoters active across different conditions (85). This high conservation with multiple promoters allows *FOXP2* to fill many roles, from neuronal expression related to vocal learning (74) to body plan development and jaw formation (82, 154).

We used genetically targeted proximity labeling with dCas9-APEX2 to covalently tag with biotin the promoter-proximal proteome of the single copy locus *FOXP2* at the upstream regulatory region active in that and many other cell types (TSS1) (85). The extensive washing to remove background contaminants coupled with quantitative mass spectrometry via online 2D-LC with TMT-RTS-SPS-MS3 enabled by a tribrid mass spectrometer (Fig. 2A) allowed us to attain deep proteome coverage and identify 373 proteins significantly enriched at the *FOXP2* promoter (Storey- $q < 0.05$, FC > 1.2), including 72 TFs. Many of these candidate transcriptional regulators have orthogonal evidence of biological relevance—from previous literature evidence to computationally predicted TF binding to experimental ENCODE ChIP-Seq data. A great deal of effort has gone into generating mice with disruptions to *FoxP2* expression in a bid to understand its biological function (155–158). The candidate regulators identified in this study can be further used for hypothesis generation to uncover the upstream mediators of *FoxP2* effects in different conditions and developmental time points.

All identified proteins are only candidate regulators until confirmed through orthogonal lines of evidence. We leveraged the ENCODE database to provide direct evidence for or against factors that were present in both datasets, giving an important benchmark for our proteomic results and bioinformatic analysis pipeline. Although there were more than 200 TF ChIP-Seq experiments in HEK293 cells, some of our candidate regulators we detected and enriched by MS that

were not tested in HEK cells had been tested in other cell lines. Future locus-specific chromatin isolation studies would benefit from utilizing one of the ENCODE ‘depth’ cell lines, like HepG2 or K562 (Fig. 4A), if they accurately recapitulate the transcriptional conditions of interest to researchers.

Confirmation of candidate regulators as binders via ChIP-Seq or ChIP-qPCR is a resource and labor-intensive process that also requires the existence of ChIP-grade antibodies. Contemporary locus-specific chromatin studies often follow up on only a few biologically informed hits and deeply characterize them (28, 39, 40, 44). A fruitful frontier is determining changes in proteome composition at a targeted locus in response to experimental perturbations (41–43, 159). Mass spectrometry provides rich, multidimensional data of measured peptides and their quantification to the protein level which can be used to prioritize hits for follow up in a discovery-based manner. Some of the false positives identified in our mass spectrometry dataset that did not show signals in the ENCODE database had weak underlying PSM data (e.g., identified by a single shared peptide), highlighting the utility of incorporating the mass spectrometry data at all levels for decision making. There is much to be gained from systematic analysis of proteomics datasets in addition to targeted follow ups. Caution must be taken, however, especially with results from labs without deep MS-based proteomics experience—it is not uncommon in the proteomics field to observe uncorrected p values at the protein quantification level (101) despite stringent, data-driven FDR at the PSM, peptide, and protein identification levels. With calls from active practitioners (54, 99, 160) and options for multiple hypothesis correction in most MS data analysis tools, the proteomics community is beginning to adopt statistical strategies for multiple hypothesis testing in protein bioinformatic analysis, but it is by no means universal.

It is noteworthy that LEF1 showed no signal at TSS1 or e330 in ChIP-Seq given its demonstrated role in regulating *foxp2* expression in developing zebrafish (138). This was not due to lack of expression—we detected LEF1, and it was enriched relative to NoG-conditions. Nuclear expression of a TF coupled to open chromatin at an actively transcribed gene that it is known to regulate under certain conditions is not sufficient to induce binding. It is possible that LEF1 can be recruited to the *FOXP2* locus in HEK293 cells under the appropriate signaling conditions. Alternatively, LEF1 may need coregulators to bind to *FOXP2* that are not expressed in HEK293 cells. Another possibility is that LEF1 binding and regulation of *FOXP2* occurs at genomic elements other than the upstream transcribed promoter or active enhancer 330 kb downstream or that LEF1 regulates *FOXP2* expression without a direct binding interaction (146). Other components of the Wnt/ β -catenin signaling pathway do have evidence for interaction with *FOXP2* TSS1 in HEK cells—TCF7L2 binding at TSS1 was significant according to ENCODE ChIP-Seq, and GNL3 was one of our top significantly enriched hits.

We identified two novel negative regulators of *FOXP2*, IRF2BP2 and glucocorticoid signaling. Knockdown of IRF2BP2 with siRNA increases expression of *FOXP2* at the transcript level, while activation of glucocorticoid signaling by dexamethasone decreases it. These effects appear additive, with the lowest expression levels observed in conditions with expression of IRF2BP2 and activation of glucocorticoid signaling. In addition to IRF2BP2 being significantly enriched in our proteomics results, the partner proteins IRF2BP1 and IRF2 were nonsignificantly enriched and predicted from TF binding tools, respectively. When choosing hits for labor-intensive follow-up validation, it is prudent to pursue proteins that are part of complexes with multiple identified members. The known coregulators CREB and ATF1, both enriched and with computational evidence, have literature evidence for binding to *FOXP2* (132). Multiple members of the NuRD complex were enriched, while other families of chromatin remodeling complexes were not identified. Leveraging prior knowledge on association in shared protein complexes can improve the likelihood of biological relevance.

We found that expression levels for ENCODE TFs significantly correlated with MS-detection status, with more lowly expressed genes less likely to be detected. There was no difference in mean expression based on ChIP-Seq binding status for proteins detected by MS. For ENCODE proteins we did not detect by MS, the mean expression of binders was significantly lower than for nonbinders. Stratifying by locus, only the on target TSS1 showed lower expression for binders, while the expression levels for TF binders and nonbinders did not differ significantly at any of the other analyzed loci. These results indicate that genetically targeted proximity labeling is specific for a targeted region, but sensitivity is still limited. Our off target ENCODE analysis also suggests that highly transcribed genes are not ideal controls for locus specificity analysis given that ChIP-Seq data showed no significant difference in proportion of TF binders between TSS1-*GAPDH* and e330-*GAPDH*. This observation is in line with results from the 4D Nucleome project, which indicate highly expressed housekeeping genes form multiple long-range enhancer-promoter interactions (161). It is interesting to note that *FOXP2* has the highest number of ENCODE TF binders at analyzed loci, exceeding the housekeeping *GAPDH* and giving further evidence for its complex layers of regulation.

Limitations of the Study

One limitation of this study is the use of an untargeted dCas9 as a negative control rather than a nontargeting gRNA containing a sequence not found in the human genome. The dynamics of dCas9 interrogating the genome differs in the presence and absence of a gRNA (162). An off-target locus (36, 159) or an alternative locus of interest for comparative studies (41) can also be used as a control to account for the high background in proximity labeling studies. Despite the

high background, proximity labeling is a valuable approach that provides complementary information to affinity purification, including the ability to detect lower abundance proteins (163). A snapshot of chromatin regulators within an approximately 400 bp radius (38) over the course of a minute helps unravel the dynamic processes underlying gene regulation, including the ability to detect proteins that may not be cross-linked by formaldehyde or that associate with the targeted locus transiently during the labeling period.

Our on-bead digestion protocol leaves behind the residues and associated proteolytic fragments that were biotinylated in our sample preparation (164) (canonically tyrosine, but cysteine and lysine labeling have been observed as well with peroxidase-based methods (165)), compounding the protein inference problem by removing potentially informative peptides from what we injected on the mass spectrometer (166, 167). Using desthiobiotin (26, 168) or a clickable biotin analog with a cleavable linker to release labeled proteins from the streptavidin beads before digestion could capture those peptides while preventing the unacceptable sample loss for low-input studies that occurs with repeated manipulations on the microscale. Such an approach could be adapted in principle to top-down proteomics protocols to study alternative proteoforms interacting with a promoter (169).

Another shortcoming is that this is necessarily a bulk measurement averaged across many cells. It is unlikely that all the identified regulators co-occupy the *FOXP2* promoter at the same time in the same cell, obscuring the important biological insights that can be obtained at the single cell level (14, 170). Understanding which factors cobind or which are mutually exclusive to each other is not possible with this dataset alone. Split enzymes or FRET assays could be an approach to test cobinding of specific factors.

Another issue is that cells were not synchronized. The TF YY1, significantly enriched in our dataset, predicted by PROMO, and with a significant ChIP-Seq signal at TSS1 according to ENCODE (Fig. 4D), has been shown to have differing effects on its role in establishing and maintaining chromatin loops based on the cell cycle at some genomic locations (171). The interaction of *FOXP2* itself with DNA is also tightly regulated and controlled throughout the cell cycle (172). Synchronizing cells could help ensure promoters of interest have similar chromatin states.

There were many false negatives (nearly 70% of ENCODE-positive TFs) despite the high initial input of cells ($\sim 6 \times 10^8$ on-target chromatin equivalents). Some TFs we detected were not quantified, indicating they were present below the limit of quantification, and detection was significantly correlated with TF expression level with a uniform detection rate of approximately one-third (Fig. 5C). Utilizing mass spectrometry approaches like parallel reaction monitoring could help identify candidate regulators at an isolated genomic locus in a hypothesis-driven, targeted approach (173).

It is well appreciated that TFs can undergo extensive posttranslational modifications that impact subcellular localization, interaction partners, and regulatory function (132, 133, 172, 174, 175). We used a closed search strategy and did not specify functionally relevant posttranslational modifications like phosphorylation, methylation, or lysine acetylation in analyzing our data. An open search, spectrum-centric strategy can identify and localize posttranslational modifications even if they have not previously been identified (176–178). Such an approach could in principle identify peptides and proteins within this dataset containing posttranslational modifications related to transcriptional regulatory function, potentially including histone modifications (179).

Generating enough starting material for deep proteome coverage is labor intensive. Contemporary locus-specific proteomics experiments often target repetitive regions like telomeres (92 copies of repetitive sequence per cell) (26, 27, 34, 36, 37, 42), centromeres (46 copies of repetitive sequence per cell) (34, 37, 42, 180, 181), or LINE-1 transcribed promoters (80–100 copies per cell) (181, 182). Another approach is to use reporter plasmids present at a higher copy number to increase intrinsic signal (40, 183) or to use a model system with a smaller genome to reduce background (184), examples of single-copy locus enrichment in human systems (two copies per cell) notwithstanding (28, 29, 38, 39, 42–44, 159). Strategies to reduce necessary input volume are needed to enable more widespread adoption of promoter-pulldown proteomics. Single cell proteomics approaches have utilized a TMT-based signal boosting approach to generate enough material for analysis (185). Adaptation of that paradigm for promoter-proximal proteomes is a promising avenue for development. The data-dependent acquisition approach used in this study samples only a fraction of the ions that elute from the LC for MS analysis stochastically. Data-independent acquisition approaches (159, 181), especially when coupled with parallel accumulation-serial fragmentation enabled by ion mobility separation (186, 187), can sample all ions eluting off the LC and provide deep proteome coverage on material-limited samples, giving another potential avenue to reduce sample input requirements. For promoter occupancy studies on very highly conserved genes, Epi-Decoder provides an interesting proteomics-by-sequencing approach that can be used in yeast (188).

CONCLUSION

The majority of disease-risk variants identified in genome-wide association studies map to noncoding regions (189). Understanding how these genetic variants give rise to different phenotypes and impact human health requires being able to determine chromatin-associated proteins and how they differ across sequence diversity (190). Forward genetic studies relying on targeted antibodies to determine

genome-wide association of disease relevant factors have been leveraged to great effect. Even with increasingly sophisticated protocols that allow individual labs to collect data at a scale previously accessible only to multi-institution consortia (191), these forward genetic studies are intrinsically limited by the availability of high-quality antibodies. Genetically targeted proximity labeling coupled to quantitative mass spectrometry proteomics is a powerful approach for discovery-based reverse genetics to determine the phenotypic impact of sequence variation without the need for ChIP-grade antibodies or a prior hypothesis of molecular mediators. The biochemical and analytical challenges can also serve as a Muse for development of new and improved proteomics instrumentation, sample preparation, data acquisition, and analysis protocols (192, 193).

DATA AVAILABILITY

Mass spectrometry data described in this communication have been deposited to the MassIVE repository partner of the ProteomeXchange Consortium available at massive.ucsd.edu with the identifier MSV000098217. Experiment identifiers and contributing labs for the 223 ENCODE ChIP-Seq experiments utilized in this study are enumerated in Table S7. Data for IRF2BP2 knockdown and ChIP-Seq were obtained through the NCBI GEO with accession identifier [GSE124636](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124636).

Supplemental data—This article contains [supplemental data](#).

Acknowledgments—This work was performed on the ancestral and unceded land of the Muwekma Ohlone Tribe. Cell sorting for this project was done on instruments in the Stanford Shared FACS Facility (RRID: SCR_017788). The Stanford University Department of Statistics Consulting Services and the Vincent Coates Foundation Stanford University Mass Spectrometry facility (RRID: SCR_017801) provided helpful feedback. The authors would like to acknowledge members of the Snyder lab at Porter Drive, particularly Y Zhu, S White, and M Nshanian, for helpful discussions. The authors appreciate comments on early drafts provided by S Saad, S Nageswaran, and G Erwin. This work was supported by NIH grants 3UM1HG009442-04S1 and 5RM1HG00773510 and 1S10OD025212-01 to M. P. S.

Author contributions—T. M. G. M. conceptualization; T. M. G. M. data curation; T. M. G. M. formal analysis; M. P. S. funding acquisition; T. M. G. M., L. R., and R. J. investigation; T. M. G. M. and L. J. methodology; T. M. G. M., L. R., and L. J. project administration; M. P. S. resources; M. P. S. supervision; T. M. G. M. visualization; T. M. G. M. writing—original draft; T. M. G. M., M. P. S., L. R., L. J., and R. J. writing—review and editing.

Conflict of interest—M. P. S. is a co-founder and scientific advisor for Crosshair Therapeutics, Exposomics, Filtricine, Fodsel, Iollo, InVu Health, January AI, Marble Therapeutics, Mirvie, Next Thought AI, Orange Street Ventures, Personalis, Protos Biologics, Qbio, RTHM, SensOmics. M. P. S. serves as a scientific advisor for Abbratech, Applied Cognition, Eno-vone, Jupiter Therapeutics, M3 Helium, Mitrix, Neuvivo, Onza, Sigil Biosciences, TranscribeGlass, WndrHLTH, Yuva Research. M. P. S. is a co-founder of NiMo Therapeutics. M. P. S. is an investor and scientific advisor of R42 and Swaza. M. P. S. is an investor in Repair Biotechnologies. The other authors declare no competing interest.

Abbreviations—The abbreviations used are: 3C, chromatin conformation capture; ChIP, chromatin immunoprecipitation; ENCODE, Encyclopedia of DNA Elements; FACS, fluorescence activated cell sorting; FC, fold change; GdnCl, guanidinium chloride; MS3, triple stage mass spectrometry; PSM, peptide spectral match; RIPA, radioimmunoprecipitation assay; RTS, real-time database search; SPS, synchronous precursor selection; TBST, Tris-buffered saline with Tween 20 detergent; TEAB, triethylammonium bicarbonate; TF, transcription factor; TMT, tandem mass tag; TPM, transcripts per million; TSS, transcription start site.

Received October 16, 2025, and in revised form, April 10, 2026
Published, MCPRO Papers in Press, April 21, 2026, <https://doi.org/10.1016/j.mcpro.2026.101570>

REFERENCES

- Haberle, V., and Stark, A. (2018) Eukaryotic Core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Bio.* **19**, 621–637
- Andersson, R., and Sandelin, A. (2020) Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Gen.* **21**, 71–87
- Schoenfelder, S., and Fraser, P. (2019) Long-Range enhancer-promoter contacts in gene expression control. *Nat. Rev. Gen.* **20**, 437–455
- Ayoubi, T. A. Y., and Van De Yen, W. J. M. (1996) Regulation of gene expression by alternative promoters. *FASEB J.* **10**, 453–460
- Kamat, A., Hinshelwood, M. M., Murry, B. A., and Mendelson, C. R. (2002) Mechanisms in tissue-specific regulation of Estrogen biosynthesis in humans. *Trends Endocrinol. Metab.* **13**, 122–128
- Kornblihtt, A. R. (2005) Promoter usage and alternative splicing. *Curr. Opin. Cell Biol.* **17**, 262–268
- Kolathur, K. K. (2021) Role of promoters in regulating alternative splicing. *Gene* **782**, 1–11
- Alfonso-Gonzalez, C., and Hilgers, V. (2024) (Alternative) transcription start sites as regulators of RNA processing. *Trends Cell Biol.* **34**, 1018–1028
- Ziabari, O. S., Liu, F., Deem, K. D., Liu, X., Kholwadwala, A., and Brisson, J. A. (2025) Gene duplication captures morph-specific promoter usage in the evolution of aphid wing dimorphisms. *PNAS* **122**, e2420893122
- Huang, K. K., Huang, J., Wu, J. K. L., Lee, M., Tay, S. T., Kumar, V., et al. (2021) Long-read transcriptome sequencing reveals abundant promoter diversity in distinct molecular subtypes of gastric cancer. *Genome Biol.* **22**, 1–25
- Demircioğlu, D., Cukuroğlu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N. A., et al. (2019) A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. *Cell* **178**, 1465–1477.E17
- Floor, S. N., and Doudna, J. A. (2016) Tunable protein synthesis by transcript isoforms in human cells. *eLife* **5**, 1–25
- Byeon, G. W., Cenik, E. S., Jiang, L., Tang, H., Das, R., and Barna, M. (2021) Functional and structural basis of extreme conservation in vertebrate 5' untranslated regions. *Nat. Genet.* **53**, 729–741
- Wu, T., Jiang, L., Mukhtar, T., Wang, L., Jian, R., Wang, C., et al. (2026) Single-Cell proteomic landscape of the developing human brain. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-025-02980-7>
- Cummins, M., Watson, C., Edwards, R. J., and Mattick, J. S. (2024) The evolution of ultraconserved elements in vertebrates. *Mol. Biol. Evol.* **41**, 1–17
- Vacik, T., and Raska, I. (2017) Alternative intronic promoters in development and disease. *Protoplasma* **254**, 1201–1206
- Zhang, M., Sjöström, M., Cui, X., Foye, A., Farh, K., Shrestha, R., et al. (2024) Integrative analysis of ultra-deep RNA-Seq reveals alternative promoter usage as a mechanism of activating oncogenic programmes during prostate cancer progression. *Nat. Cell Biol.* **26**, 1176–1186
- MacKenzie, T. M. G., Cisneros, R., Maynard, R. D., and Snyder, M. P. (2023) Reverse-ChIP techniques for identifying locus-specific proteomes: a key tool in unlocking the cancer regulome. *Cells* **12**, 1–38
- Vermeulen, M., and Déjardin, J. (2020) Locus-Specific chromatin isolation. *Nat. Rev. Mol. Cell Bio.* **21**, 249–250
- Gauchier, M., van Mierlo, G., Vermeulen, M., and Déjardin, J. (2020) Purification and enrichment of specific chromatin loci. *Nat. Methods* **17**, 380–389
- Biggin, M. D. (2011) Animal transcription networks as highly connected, quantitative continua. *Developmental Cell* **21**, 611–626
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) Molecular biology of the cell. In: *Chromosomal DNA and Its Packaging in the Chromatin Fiber*, 4th edition, Garland Science, New York
- Wiśniewski, J. R., Hein, M. Y., Cox, J., and Mann, M. (2014) A “Proteomic Ruler” for protein copy number and concentration estimation without Spike-in standards. *Mol. Cell. Proteomics* **13**, 3497–3506
- West, K. L., Byrum, S. D., Mackintosh, S. G., Edmondson, R. D., Taverna, S. D., and Tacket, A. J. (2019) Proteomic characterization of the arsenic response locus in *S. cerevisiae*. *Epigenetics* **14**, 130–145
- Weiß, M., Chanou, A., Schauer, T., Tvardovskiy, A., Meiser, S., König, A.-C., et al. (2023) Single-Copy locus proteomics of Early- and late-firing DNA replication origins identifies a role of Ask1/DASH complex in replication timing control. *Cell Rep.* **42**, 1–23
- Déjardin, J., and Kingston, R. E. (2009) Purification of proteins associated with specific genomic loci. *Cell* **136**, 175–186
- Fujita, T., Asano, Y., Ohtsuka, J., Takada, Y., Saito, K., Ohki, R., et al. (2013) Identification of telomere-associated molecules by engineered DNA-Binding molecule-mediated chromatin immunoprecipitation (enChIP). *Sci. Rep.* **3**, 1–7
- Alkhayer, R., Ponath, V., Frech, M., Ashikary, T., Graumann, J., Neubauer, A., et al. (2023) KLF4-Mediated upregulation of the NKG2D ligand MICA in Acute Myeloid leukemia: a novel therapeutic target identified by enChIP. *Cell Commun. Signal.* **21**, 1–14
- Alkhayer, R., Ponath, V., and von Strandmann, E. P. (2024) Protocol to target a promoter Region in Human embryonic kidney cells using the CRISPR-dCas9 System for single-locus proteomics. *STAR Protoc.* **5**, 1–15
- Wang, Z., He, Z., Liu, Z., Qu, M., Gao, C., Wang, C., et al. (2022) A reverse chromatin immunoprecipitation technique based on the CRISPR-dCas9 System. *Plant Physiol.* **191**, 1505–1519
- Zhang, L., Cai, C., Chen, Q., Tan, X., Chen, S., Zhang, K., et al. (2026) A CRISPR-Based sequence proximity binding protein labelling System for scanning upstream regulatory proteins. *Nat. Plants.* <https://doi.org/10.1038/s41477-025-02212-5>
- Bosch, J. A., Chen, C.-L., and Perrimon, N. (2020) Proximity-Dependent labeling methods for proteomic profiling in living cells: an update. *WIREs Develop. Biol.* **10**, 1–17
- Ummerhum, H., and Hamperl, S. (2020) Proximity labeling techniques to study chromatin. *Front. Genet.* **11**, 1–13
- Schmidtman, E., Anton, T., Rombaut, P., Herzog, F., and Leonhardt, H. (2016) Determination of local chromatin composition by CasID. *Nucleus* **7**, 476–484
- Ugur, E., Bartoschek, M. D., and Leonhardt, H. (2020) Locus-Specific chromatin proteome revealed by Mass spectrometry-based CasID. *The Nucleus. Methods in Molecular Biology* Hancock, R., ed **2175**, Humana, New York, NY

36. Qiu, W., Xu, Z., Zhang, M., Zhang, D., Fan, H., Li, T., *et al.* (2019) Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 System. *Nucleic Acids Res.* **47**, 1–14
37. Gao, X. D., Tu, L.-C., Mir, A., Rodriguez, T., Ding, Y., Leszyk, J., *et al.* (2018) C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat. Methods* **15**, 433–436
38. Myers, S. A., Wright, J., Peckner, R., Kalish, B. T., Zhang, F., and Carr, S. A. (2018) Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-Mediated proximity labeling. *Nat. Methods* **15**, 437–439
39. Gao, S., Menendez, M., Kurylowicz, K., and Griffin, C. T. (2021) Genomic locus proteomic screening identifies the NF- κ B signaling pathway components NF κ B1 and IKBKG as transcriptional regulators of *Ripk3* in endothelial cells. *PLoS ONE* **16**, 1–16
40. Pizzolato, G., Moparthi, L., Pagella, P., Cantù, C., D'Arcy, P., and Koch, S. (2023) The Tumor suppressor p53 is a negative Regulator of the carcinoma-associated transcription factor FOXQ1. *JBC* **300**, 1–9
41. [Preprint] Yu, G., Zhang, D. Y., Aguilan, J. T., Sidoli, S., and Scharff, M. D. (2022) Locus-Specific proteomics identifies new aspects of the chromatin context involved in V Region somatic hypermutation. *bioRxiv*. <https://doi.org/10.1101/2022.09.08.507190>
42. Tchara, P.-E. K., Loehr, J., and Lambert, J.-P. (2025) Coupling proximity biotinylation with genomic targeting to characterize locus-specific changes in chromatin environments. *J. Proteome Res.* **24**, 1845–1860. *Note that this reference uses UltraID-based proximity labeling, not APEX2*
43. Cenik, B. K., Aoi, Y., Iwanaszko, M., Howard, B. C., Morgan, M. A., Andersen, G. D., *et al.* (2024) TurboCas: a method for locus-specific labeling of genomic regions and isolating their associated protein interactome. *Mol. Cell* **84**, 4929–4944.E8. *Note that this reference uses TurboID-based proximity labeling, not APEX2*
44. Torres, M., Kirchner, M., Marks, C. G., Mertins, P., and Kramer, A. (2024) Proteomic insights into circadian transcription regulation: novel E-Box interactors revealed by proximity labeling. *Genes Dev.* **38**, 1020–1032
45. Tan, B., Peng, S., Yatim, S. M. J. M., Gunaratne, J., Hunziker, W., and Ludwig, A. (2020) An optimized protocol for proximity biotinylation in confluent epithelial cell cultures using the peroxidase APEX2. *STAR Protoc.* **1**, 1–14
46. van Mierlo, G., and Vermeulen, M. (2021) Chromatin proteomics to Study epigenetics – challenges and opportunities. *Mol. Cell. Proteomics* **20**, 1–13
47. Ji, X., Dadon, D. B., Abraham, B. J., Lee, T. I., Jaenisch, R., Bradner, J. E., *et al.* (2015) Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *PNAS* **112**, 3841–3846
48. Ugur, E., de la Porte, A., Qin, W., Bultmann, S., Livanova, A., Drukker, M., *et al.* (2023) Comprehensive chromatin proteomics resolves functional phases of pluripotency and identifies changes in regulatory components. *Nucleic Acids Res.* **51**, 2671–2690
49. Specht, H., and Slavov, N. (2018) Transformative opportunities for single-cell proteomics. *J. Proteome Res.* **17**, 2565–2571
50. Gatto, L., Aebbersold, R., Cox, J., Demichev, V., Derks, J., Emmott, E., *et al.* (2023) Initial recommendations for performing, benchmarking, and reporting single-cell proteomics experiments. *Nat. Methods* **20**, 375–386
51. Hein, M. Y., Peng, D., Todorova, V., McCarthy, F., Kim, K., Liu, C., *et al.* (2025) Global organelle profiling reveals subcellular localization and remodeling at Proteome Scale. *Cell* **188**, P1137–P1155.E20
52. McCaskie, K., Hutchings, C., Feret, R., Kim, Y.-I., Breckls, L., Deery, M., *et al.* (2025) Localization of Organelle proteins using data-independent acquisition (DIA-LOP). *Mol. Cell. Prot.* **24**, 1–16
53. Zhang, L., and Elias, J. E. (2017) Relative protein quantification using Tandem Mass tag Mass spectrometry. In: Comai, L., Katz, J., Mallick, P., eds. *Proteomics, Methods in Molecular Biology*, Humana, New York, NY: 1550
54. Jiang, Y., Rex, D. A. B., Schuster, D., Neely, B. A., Rosano, G. L., Volkmar, N., *et al.* (2024) Comprehensive overview of Bottom-Up proteomics using mass spectrometry. *ACS Meas. Sci. Au.* **4**, 338–417
55. Shuken, S. R. (2023) An introduction to mass spectrometry-based proteomics. *J. Proteome Res.* **22**, 2151–2171
56. Liu, Y.-C., and Chen, C.-J. (2023) Online 2D High-pH and Low-pH reversed-phase Nano-LC-MS/MS System for deep proteome analysis. *Anal. Chem.* **95**, 5850–5857
57. Ting, L., Rad, R., Gygi, S. P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **8**, 937–940
58. Schweppe, D. K., Eng, J. K., Yu, Q., Bailey, D., Rad, R., Navarrete-Perea, J., *et al.* (2020) Full-Featured, real-time database searching platform enables fast and accurate multiplexed quantitative proteomics. *J. Proteome Res.* **19**, 2026–2034
59. McAlister, G. C., Nusinoq, D. P., Jedrychowki, M. P., Wühr, M., Huttlin, E. L., Erickson, B. K., *et al.* (2014) MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158
60. He, Y., Shishkova, E., Peters-Clarke, T. M., Brademan, D. R., Westphall, M. S., Bergen, D., *et al.* (2023) Evaluation of the orbitrap ascend Tribrid Mass spectrometer for shotgun proteomics. *Anal. Chem.* **95**, 10655–10663
61. Shuken, S. R., McAlister, G. C., Barshop, W. D., Canterbury, J. D., Bergen, D., Huang, J., *et al.* (2023) Deep proteomic compound profiling with the orbitrap ascend Tribrid Mass spectrometer using Tandem Mass tags and real-time search. *Anal. Chem.* **95**, 15180–15188
62. den Hoed, J., Devaraju, K., and Fisher, S. E. (2021) Molecular networks of the FOXP2 transcription factor in the brain. *EMBO Rep.* **22**, 1–15
63. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., *et al.* (2004) Ultraconserved elements in the human genome. *Science* **304**, 1321–1325
64. Ferlan, R. J., Cherry, T. J., Preware, P. O., Morrissey, E. E., and Walsh, C. A. (2003) Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *J. Comp. Neurol.* **460**, 266–279
65. Co, M., Anderson, A. G., and Konopka, G. (2020) FOXP transcription factors in vertebrate brain development, function, and disorders. *WIREs Develop. Biol.* **9**, 1–25
66. Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., and Monaco, A. P. (2001) A forkhead-domain gene is mutated in a severe speech and Language disorder. *Nature* **413**, 519–523
67. Fisher, S. E., and Scharff, C. (2009) FOXP2 as a molecular window into speech and Language. *Trends Genet.* **25**, 166–177
68. Scharff, C., and Haesler, S. (2005) An evolutionary perspective on FoxP2: strictly for the birds? *Curr. Opin. Neurobiol.* **15**, 694–703
69. Li, G., Wang, J., Rossiter, S. J., Jones, G., and Zhang, S. (2007) Accelerated FoxP2 evolution in echolocating Bats. *PLoS ONE* **2**, 1–10
70. Staes, N., Sherwood, C. C., Wright, K., de Manuel, M., Guevara, E. E., Marques-Bonet, T., *et al.* (2017) FOXP2 variation in great ape populations offers Insight into the evolution of communication skills. *Sci. Rep.* **7**, 1–10
71. Atkinson, E. G., Audesse, A. J., Palacios, J. A., Bobo, D. M., Webb, A. E., Ramachandran, S., *et al.* (2018) No evidence for recent selection at FOXP2 among diverse human populations. *Cell* **174**, 1424–1435.e15
72. Kuhlwlilm, M. (2018) The evolution of FOXP2 in the light of admixture. *Curr. Opin. Behav. Sci.* **21**, 120–126
73. Vaglietti, S., Villeri, V., Dell'Oca, M., Marchetti, C., Cesano, F., Rizzo, F., *et al.* (2023) PolyQ length-based molecular encoding of vocalization frequency in FOXP2. *iScience* **26**, 108036
74. Heston, J. B., and White, S. A. (2015) Behavior-Linked FoxP2 regulation enables Zebra finch vocal learning. *J. Neurosci.* **35**, 2885–2894
75. Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., *et al.* (2004) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, 0116–0130
76. Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., *et al.* (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and Language, in developing human brain. *AJHG* **81**, 1144–1157
77. Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E., *et al.* (2007) High-Throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and Language disorders. *AJHG* **81**, 1232–1250
78. Vernes, S. C., Oliver, P. L., Spiteri, E., Lockstone, H. E., Puliadi, R., Taylor, J. M., *et al.* (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet.* **7**, 1–17

79. Konopka, G., Bomar, J. M., Winden, K., Coppola, G., Jonsson, Z. O., Gao, F., *et al.* (2009) Human-Specific transcription regulation of CNS development genes by FOXP2. *Nature* **462**, 213–217
80. Tsui, D., Vessey, J. P., Tomita, H., Kaplan, D. R., and Miller, F. D. (2013) FoxP2 regulates neurogenesis during embryonic cortical development. *J. Neurosci.* **33**, 244–258
81. Oswald, F., Klöble, P., Ruland, A., Rosenkranz, D., Hinz, B., Butter, F., *et al.* (2017) The FOXP2-Driven network in developmental disorders and neurodegeneration. *Front. Cell. Neurosci.* **11**, 1–24
82. Shu, W., Lu, M. M., Zhang, Y., Tucker, P. W., Zhou, D., and Morrissey, E. E. (2007) Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Develop. Dis.* **134**, 1991–2000
83. Hernández Vázquez, M. N., Ulvmar, M. H., González-Loyola, A., Kritikos, I., Sun, Y., He, L., *et al.* (2021) Transcription factor FOXP2 is a flow-induced regulator of collecting lymphatic vessels. *EMBO J.* **40**, 1–16
84. Xu, S., Liu, P., Chen, Y., Chen, Y., Zhang, W., Zhao, H., *et al.* (2018) Foxp2 regulates anatomical features that May be relevant for vocal behaviors and bipedal locomotion. *PNAS* **115**, 8799–8804
85. Schroeder, D. I., and Myers, R. M. (2008) Multiple transcription start sites for FOXP2 with varying cellular specificities. *Gene* **413**, 42–48
86. Becker, M., Devanna, P., Fisher, S. E., and Vernes, S. C. (2018) Mapping of human FOXP2 enhancers reveals complex regulation. *Front. Mol. Neurosci.* **11**, 1–15
87. Adegbola, A. A., Cox, G. F., Bradshaw, E. M., Hafler, D. A., Gimelbrant, A., and Chess, A. (2014) Monoallelic expression of the Human FOXP2 speech gene. *PNAS* **112**, 6848–6854
88. Zhu, X., Chen, C., Wei, D., Xu, Y., Liang, S., Jia, W., *et al.* (2023) FOXP2 confers oncogenic effects in prostate cancer. *eLife* **12**, 1–23
89. Chen, F., Byrd, A. L., Liu, J., Flight, R. M., DuCote, T. J., Naughton, K. J., *et al.* (2023) Polycomb deficiency drives a FOXP2-High aggressive State targetable by epigenetic inhibitors. *Nat. Commun.* **14**, 1–18
90. Cui, B. G., and Karnoub, A. E. (2016) Silencing FOXP2 in breast cancer cells promotes cancer stem cell traits and metastasis. *Mol. Cell. Oncol.* **3**, 1–3
91. Chen, M.-T., Sun, H.-F., Li, L.-D., Zhao, Y., Yang, L.-P., Gao, S.-P., *et al.* (2018) Downregulation of FOXP2 promotes breast cancer migration and invasion through TGFβ/SMAD signaling pathway. *Oncol. Lett.* **15**, 8582–8588
92. Yan, X., Zhou, H., Zhang, T., Xu, P., Zhang, S., Huang, W., *et al.* (2015) Downregulation of FOXP2 promoter Human hepatocellular carcinoma cell invasion. *Tumor Biol.* **36**, 9611–9619
93. Herrero, M. J., and Gitton, Y. (2018) The untold stories of the speech gene, the FOXP2 cancer gene. *Genes & Cancer* **9**, 11–38
94. Orsburn, B. C. (2021) Proteome discoverer – a community enhanced data processing suite for protein informatics. *Proteomes* **9**, 1–13
95. The ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74
96. Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., *et al.* (2020) New developments on the encyclopedia of DNA elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889
97. Zhang, J., Lee, D., Dhiman, V., Jiang, P., Xu, J., McGilivray, P., *et al.* (2020) An integrative ENCODE resource for cancer genomics. *Nat. Commun.* **11**, 1–11
98. [Preprint] Hitz, B. C., Lee, J.-W., Jolanki, O., Kagda, M. S., Graham, K., Sud, P., *et al.* (2023) The ENCODE Uniform Analysis Pipelines. *bioRxiv*. <https://doi.org/10.1101/2023.04.04.535623>
99. Shukun, S. R., and McNeerney, M. W. (2023) Costs and benefits of popular P-Value correction methods in three models of quantitative omic experiments. *Anal. Chem.* **95**, 2732–2740
100. Storey, J. D., and Tibshirani, R. (2003) Statistical significance for genomewide studies. *PNAS* **100**, 9440–9445
101. Pascovici, D., Handler, D. C. L., Wu, J. X., and Haynes, P. A. (2016) Multiple testing corrections in quantitative proteomics: a useful but blunt tool. *Proteomics* **16**, 2448–2453
102. Stringer, B. W., Day, B. W., D'Souza, R. C. J., Jamieson, P. R., Ensby, K. S., Bruce, Z. C., *et al.* (2019) A reference collection of patient-derived cell line and xenograft models of proneural, classical, and mesenchymal glioblastoma. *Sci. Rep.* **9**, 1–14
103. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCt method. *Methods* **25**, 402–408
104. Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., *et al.* (2013) Proteomic mapping of Mitochondria in living cells via spatially-restricted enzymatic tagging. *Science* **339**, 1328–1331
105. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682
106. Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., Heck, A. J. R., and Schmidt, A. (2012) Large-Scale quantitative assessment of different In-Solution protein digestion protocols reveals superior cleavage efficiency of Tandem Lys-C/Trypsin proteolysis over trypsin digestion. *J. Proteome Res.* **11**, 5145–5156
107. Kumler, W., and Ingalls, A. E. (2022) Tidy data neatly resolves mass-spectrometry's ragged arrays. *R. J.* **14**, 193–202
108. The, M., MacCoss, M. J., Noble, W. S., and Käll, L. (2016) Fast and accurate protein false discovery rates on large-scale proteomics data sets with percolator 3.0. *J. Am. Soc. Mass Spectrom.* **27**, 1719–1727
109. Elias, J. E., and Gygi, S. P. (2007) Target-Decoy search strategy for increased confidence in large-scale protein identification by mass spectrometry. *Nat. Methods* **4**, 207–214
110. Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodological)* **57**, 289–300
111. Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., *et al.* (2017) In Situ capture of chromatin interactions by biotinylated dCas9. *Cell* **170**, 1028–1043.e19
112. Rashid, F., Sharma, S., and Bano, B. (2005) Comparison of guanidine hydrochloride (GdnHCl) and urea denaturation on inactivation and unfolding of Human Placental Cystatin (HPC). *Protein J.* **24**, 283–292
113. Kurban, G. P., Bayer, E. A., Wilchek, M., and Horowitz, P. M. (1991) The Quaternary structure of Streptavidin in Urea. *J. Biol. Chem.* **266**, 14470–14477
114. Fiest, P., and Hummon, A. B. (2015) Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int. J. Mol. Sci.* **16**, 3537–3563
115. Weiner, S., Sauer, M., Visser, P. J., Tijms, B. M., Vorontsov, E., Blennoq, K., *et al.* (2022) Optimized sample preparation and data analysis for TMT proteomic analysis of cerebrospinal fluid applied to the identification of alzheimer's disease biomarkers. *Clin. Proteomics* **19**, 1–20
116. Dowle, A. A., Wilson, J., and Thomas, J. R. (2016) Comparing the diagnostic classification accuracy of iTRAQ, Peak-Area, Spectral-Counting, and emPAI methods for relative quantification in expression proteomics. *J. Prot. Res.* **15**, 3550–3562
117. He, Y., Yang, K., Li, S., Zeller, M., McAlister, G. C., Stewart, H. I., *et al.* (2025) TMT-Based multiplexed (Chemo)Proteomics on the orbitrap astral mass spectrometer. *Mol. Cell. Proteomics* **24**, 1–12
118. Eden, E., Lipson, D., Yagov, S., and Yakhini, Z. (2007) Discovering motifs in ranked lists of DNA sequences. *Plos Comput. Biol.* **3**, 1–15
119. Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform.* **10**, 1–7
120. Thomas, P. D., Ebert, D., Muruganukun, A., Mushayama, T., Albou, L.-P., and Mi, H. (2021) PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci.* **31**, 8–22
121. Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., *et al.* (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221
122. Rambout, X., Dequiedt, F., and Maquat, L. E. (2018) Beyond transcription: roles of transcription factors in Pre-mRNA splicing. *Chem. Rev.* **118**, 4339–4364
123. Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., *et al.* (2010) An Atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744–752
124. Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., *et al.* (2018) The human transcription factors. *Cell* **172**, 650–655
125. Rodrigues, K. S., Petroski, L. P., Utumi, P. H., Ferrasa, A., and Herai, R. H. (2023) IARA: a complete and curated Atlas of the biogenesis of spliceosome machinery during RNA splicing. *Life Sci. Alliance* **6**, 1–15
126. Maricic, T., Günther, V., Georgiev, O., Gehre, S., Čurlin, M., Schreivels, C., *et al.* (2013) A Recent evolutionary change affects a regulatory element in the Human FOXP2 gene. *Mol. Biol. Evol.* **30**, 844–852

127. Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050
128. Zogopoulos, V. L., Spaho, K., Ntouka, C., Lappas, G. A., Kyranis, I., Bagos, P. G., *et al.* (2021) TFBSPred: a functional transcription factor binding site prediction webtool for humans and mice. *Int. J. Epigen.* **1**, 1–11
129. Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., and Albá, M. M. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18**, 333–334
130. Farré, D., Roset, R., Huerta, M., Adsua, J. E., Roselló, L., Albá, M. M., *et al.* (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res.* **31**, 3651–3653
131. Simonett, S. P., Shin, S., Herring, J. A., Bacher, R., Smith, L. A., Dong, C., *et al.* (2021) Identification of direct transcriptional targets of NFATC2 that promote β cell proliferation. *JCI* **131**, 1–14
132. Zhang, X., Odom, D. T., Koo, S.-H., Conkright, M. D., Canettieri, G., Best, J., *et al.* (2005) Genome-Wide analysis of cAMP-Response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *PNAS* **102**, 4459–4464
133. Gates, L. A., Foulds, C. E., and O'Malley, M. W. (2017) Histone marks in the 'Driver's seat': functional roles in steering the transcription cycle. *Trends Biochem. Sci.* **42**, 977–989
134. Di Tullio, F., Schwarz, M., Zorgati, H., Mzoughi, S., and Guccione, E. (2022) The duality of PRDM proteins: epigenetic and structural perspectives. *FEBS J.* **289**, 1256–1275
135. Basta, J., and Rauchman, M. (2015) The nucleosome remodeling and deacetylase (NuRD) complex in development and disease. *Transl. Res.* **165**, 36–47
136. Xie, W., Ling, T., Zhou, Y., Feng, W., Zhu, Q., Stunnenberg, H. G., *et al.* (2012) The chromatin remodeling complex NuRD establishes the poised State of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. *PNAS* **109**, 8161–8166
137. Kraushaar, D. C., Chen, Z., Tang, Q., Cui, K., Zhang, J., and Zhao, K. (2018) The gene repressor complex NuRD interacts with the histone variant H3.3 at promoters of active genes. *Genome Res.* **28**, 1646–1655
138. Bonkowski, J. L., Wang, X., Fujimoto, E., Lee, J. E., Chien, C.-B., and Dorsky, R. I. (2008) Domain-Specific regulation of foxP2 CNS expression by *lefi*. *BMC Develop. Biol.* **8**, 1–15
139. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data. *Mol. Cell. Proteomics* **4**, 1419–1440
140. Li, Y. F., and Radivojac, P. (2012) Computational approaches to protein inference in shotgun proteomics. *BMC Bioinform.* **13**(Suppl 16:S4), 1–19
141. Locard-Paulet, M., Doncheva, N. T., Morris, J. H., and Jensen, L. J. (2024) Functional analysis of MS-Based proteomics data: from protein groups to networks. *Mol. Cell. Proteomics* **23**, 1–13
142. Hinkle, T. B., and Bakalarski, C. E. (2025) Comprehensive protein inference analysis with PyProteinInference elucidates biological understanding of Tandem Mass spectrometry data. *J. Proteome Res.* **24**, 2135–2140
143. Tang, X., Zha, L., Li, H., Huang, Z., Peng, Z., and Wang, Z. (2017) Upregulation of GNL3 expression promotes Colon cancer cell proliferation, migration, invasion, and epithelial-mesenchymal transition via the Wnt/ β -Catenin signaling pathway. *Oncol. Rep.* **38**, 2023–2032
144. Nielsen, J. V., Thomassen, M., Møllgård, K., Norberg, J., and Jensen, N. A. (2014) Zbtb20 defines a hippocampal neuronal identity through direct repression of genes that control Projection Neuron development in the isocortex. *Cereb. Cortex* **24**, 1216–1229
145. Coutinho, P., Pavlou, S., Bhatia, S., Chalmers, K. J., Kleinjan, D. A., and van Heyningen, V. (2011) Discovery and assessment of conserved Pax6 target genes and enhancers. *Genome Res.* **21**, 1349–1359
146. Mahendrawada, L., Warfield, L., Donceq, R., and Hahn, S. (2025) Low overlap of transcription factor DNA binding and regulatory targets. *Nature* **642**, 796–804
147. Ramalho-Oliveira, R., Oliveira-Vieira, B., and Viola, J. P. B. (2019) IRF2BP2: a new player in the regulation of cell homeostasis. *J. Leukoc. Biol.* **106**, 717–723
148. Pastor, T. P., Peixoto, B. C., and Viola, J. P. B. (2021) The transcriptional Co-factor IRF2BP2: a new player in tumor development and microenvironment. *Front. Cell Dev. Biol.* **9**, 1–9
149. Manjur, A. B. M. K., Lempiäinen, J. K., Malinen, M., Palvimo, J. J., and Niskanen, E. A. (2019) IRF2BP2 modulates the crosstalk between glucocorticoid and TNF signaling. *J. Ster. Biochem. Mol. Bio.* **192**, 1–10
150. Su, W., Hu, S., Zhou, L., Bi, H., and Li, Z. (2024) FOXP2 inhibits the aggressiveness of lung cancer cells by blocking TGF β signaling. *Oncol. Lett.* **27**, 1–7
151. Arafeh, R., Shibue, T., Dempster, J. M., Hahn, W. C., and Vazquez, F. (2025) The present and future of the Cancer Dependency Map. *Nat. Rev. Cancer* **25**, 59–73
152. Doshi, J., Willis, K., Madurga, A., Stelzer, C., and Benenson, Y. (2020) Multiple alternative promoters and alternative splicing enable universal transcription-based logic computation in Mammalian cells. *Cell Rep.* **33**, 1–15
153. Sokolowski, T. R., Gregor, T., Bialek, W., and Tkačik, G. (2025) Deriving a genetic regulatory network from an optimization principle. *PNAS* **122**, e2402925121
154. Cesario, J. M., Almaidhan, A. A., and Jeong, J. (2016) Expression of Forkhead Box Transcription factor genes *Foxp1* and *Foxp2* during Jaw development. *Gene Expr. Patterns* **20**, 111–119
155. Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., *et al.* (2005) Altered ultrasonic vocalization in mice with a disruption in the *Foxp2* gene. *PNAS* **102**, 9643–9648
156. French, C. A., Groszer, M., Preece, C., Coupe, A.-M., Rajewsky, K., and Fisher, S. E. (2007) Generation of mice with a conditional *Foxp2* null allele. *Genesis* **45**, 440–446
157. Enard, W., Gehre, S., Hammerschmidt, K., Höfler, S. M., Blass, T., Somel, M., *et al.* (2009) A humanized version of *Foxp2* affects Cortico-Basal ganglia circuits in mice. *Cell* **137**, 961–971
158. Castellucci, G. A., McGinley, M. J., and McCormick, D. A. (2016) Knockout of *Foxp2* disrupts vocal development in mice. *Sci. Rep.* **6**, 1–12
159. Zhu, J., Li, Z., Xue, D., Meng, Z., Shao, S., Pulecio, J., *et al.* (2025) A chemical epigenetic tool to probe site-specific DNA-Binding protein complexes. *PNAS* **122**, 1–10
160. Doncheva, N. T., Schwämmle, V., and Locard-Paulet, M. (2025) Understanding data analysis steps in mass-spectrometry-based proteomics is key to transparent reporting. *J. Proteome Res.* **24**, 4965–4976
161. Dekker, J., Oksuz, B. A., Zhang, Y., Wang, Y., Minsk, M. K., Kuang, S., *et al.* (2025) An integrated view of the structure and function of the human 4D nucleome. *Nature* **649**, 759–776
162. Knight, S. C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L. B., Bosanac, L., *et al.* (2015) Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**, 823–826
163. Lambert, J.-P., Tucholska, M., Go, C., Knight, J. D. R., and Gingras, A.-C. (2015) Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J. Proteomics* **118**, 81–94
164. Branon, T. C., Qin, W., and Udeshi, N. D. (2025) From vision to reality: a perspective on the innovative and collaborative forces that evolved proximity labeling. *Mol. Cell. Prot.* **24**, 1–9
165. He, J.-X., Fei, Z.-C., Fu, L., Tian, C.-P., He, F.-C., Chi, H., *et al.* (2022) A modification-centric assessment tool for the performance of chemoproteomic probes. *Nat. Chem. Biol.* **18**, 904–912
166. Kim, H. B., and Kim, K.-E. (2025) A straightforward interpretation of proximity labeling through direct biotinylation analysis. *ACS Omega* **10**, 26098–26105
167. Wang, G., Liu, J., Sun, X., Qin, W., Han, S., and Zou, P. (2026) Recent advances in proximity labeling-based subcellular proteomic mapping. *Mol. Cell. Prot.* **25**, 1–15
168. Lee, S.-Y., Kang, M.-G., Shin, S., Kwak, C., Kwon, T., Seo, J. K., *et al.* (2017) Architecture mapping of the inner mitochondrial membrane proteome by chemical tools in live cells. *J. Am. Chem. Soc.* **139**, 3651–3662
169. Po, A., and Eysers, C. E. (2023) Top-Down proteomics and the challenges of true proteoform characterization. *J. Proteome Res.* **22**, 3663–3675
170. Nitz, A. A., Chavez, J. H. G., Eliason, Z. G., and Payne, S. H. (2025) Are there yet? Assessing the readiness of single-cell proteomics to answer biological hypotheses. *J. Proteome Res.* **24**, 1482–1492
171. Lam, J. C., Aboreden, N. G., Midla, S. C., Wang, S., Huang, A., Keller, C. A., *et al.* (2024) YY1-Controlled regulatory connectivity

- and transcription are influenced by the cell cycle. *Nat. Genet.* **56**, 1938–1952
172. Saad, S., Swigut, T., Tabatabaee, S., Lalgudi, P., Jarosz, D. F., and Wysocka, J. (2025) DNA binding and mitotic phosphorylation protect polyglutamine proteins from assembly Formation. *Cell* **188**, 2974–2991. E20
173. Kennedy, J. J., Whiteaker, J. R., Ivey, R. G., Burian, A., Chowdhury, S., Tsai, C.-F., et al. (2022) Internal standard triggered-parallel reaction monitoring mass spectrometry enables multiplexed quantification of candidate biomarkers in plasma. *Anal. Chem.* **94**, 9540–9547
174. Filtz, T. M., Vogel, W. K., and Leid, M. (2014) Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol. Sci.* **35**, 76–85
175. Benayoun, B. A., and Veitia, R. A. (2009) A post-translational modification code for transcription factors: sorting through a Sea of signals. *Trends Cell Biol.* **19**, 189–197
176. Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L., et al. (2015) A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat. Biotechnol.* **33**, 743–749
177. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., and Nesvizhskii, A. I. (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520
178. Yu, F., Teo, G. C., Kong, A. T., Haynes, S. E., Avtonomov, D. M., Geiszler, D. J., et al. (2020) Identification of modified peptides using localization-aware open search. *Nat. Commun.* **11**, 1–9
179. Vai, A., Nuberini, R., Graziadei, A., Polasky, D. A., Yu, F., Nesvizhskii, A. I., et al. (2025) Breaking boundaries in histone modification MS-based detection: a tailored search strategy for unrestricted identification of novel epigenetic marks. *Mol. Cell. Prot.* **24**, 1–15
180. Ma, R., Zhang, Y., Zhang, J., Zhang, P., Liu, Z., Fan, Y., et al. (2024) Targeting pericentric non-consecutive motifs for Heterochromatin initiation. *Nature* **631**, 678–685
181. Qian, W., Jiang, P., Niu, M., Fu, Y., Huang, D., Zhang, D., et al. (2025) Selective identification of epigenetic regulators at methylated genomic sites by SelectID. *Nat. Commun.* **16**, 1–14
182. Briggs, E. M., Mita, P., Sun, X., Ha, S., Vasilyev, N., Leopold, Z. R., et al. (2021) Unbiased proteomic mapping of the LINE-1 promoter using CRISPR Cas9. *Mobile DNA* **12**, 1–12
183. Greaves, D. E., Gabaev, I., Williamson, J. C., Bloor, S., Cuesta, S. M., Matheson, N. J., et al. (2026) Locus-specific proteomics identifies novel regulators of Epstein-Barr virus lytic reactivation. *J. Virol* **100**, e01408-25
184. Svoboda, T., Nierderdöckl-Loibl, D., Schüller, A., Hummel, K., Schlosser, S., Razzazi-Fazeli, E., et al. (2025) Locus-Specific chromatin proteomics using dCas-Guided proximity labeling in *Aspergillus nidulans*. *Fungal Genet. Biol.* **178**, 1–12
185. Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018) SCoPE-MS: Mass spectrometry of single Mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 1–12
186. Szyrwił, L., Gille, C., Müllerder, M., Demichev, V., and Ralser, M. (2023) Fast proteomics with dia-PASEF and analytical flow-rate chromatography. *Proteomics* **21**, 1–9
187. Skowronek, P., and Meier, F. (2022) High-Throughput mass spectrometry-based proteomics with dia-PASEF. *Proteomics in Systems Biology. Methods in Molecular Biology* Geddes-McAlister, J., ed **2456**, Humana, New York, NY: 15–28
188. Korthout, T., Poramba-Liyanage, D. W., va Kruijsbergen, I., Verzijlbergen, K. F., van Gemert, F. P. A., van Welsse, T., et al. (2018) Decoding the chromatin proteome of a single genomic locus by DNA sequencing. *PLoS Biol.* **16**, 1–25
189. Zhang, F., and Lupski, J. R. (2015) Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110
190. Martinez-Corral, R., Friedrich, D., Frömel, R., Velten, L., Gunawardena, J., and DePace, A. H. (2025) Emergence of activation or repression in transcriptional control under a fixed molecular context. *PNAS* **122**, 1–12
191. Perez, A. A., Goronzy, I. N., Blanco, M. R., Yeh, B. T., Guo, J. K., Lopes, C. S., et al. (2024) ChIP-DIP maps binding of hundreds of proteins to DNA simultaneously and identifies diverse gene regulatory elements. *Nat. Genet.* **56**, 2827–2841
192. Peters-Clarke, T. M., Coon, J. J., and Riley, N. M. (2024) Instrumentation at the leading edge of proteomics. *Anal. Chem.* **96**, 7976–8010
193. Guo, T., Steen, J. A., and Mann, M. (2025) Mass-Spectrometry-Based proteomics: from single cells to clinical applications. *Nature* **638**, 901–911