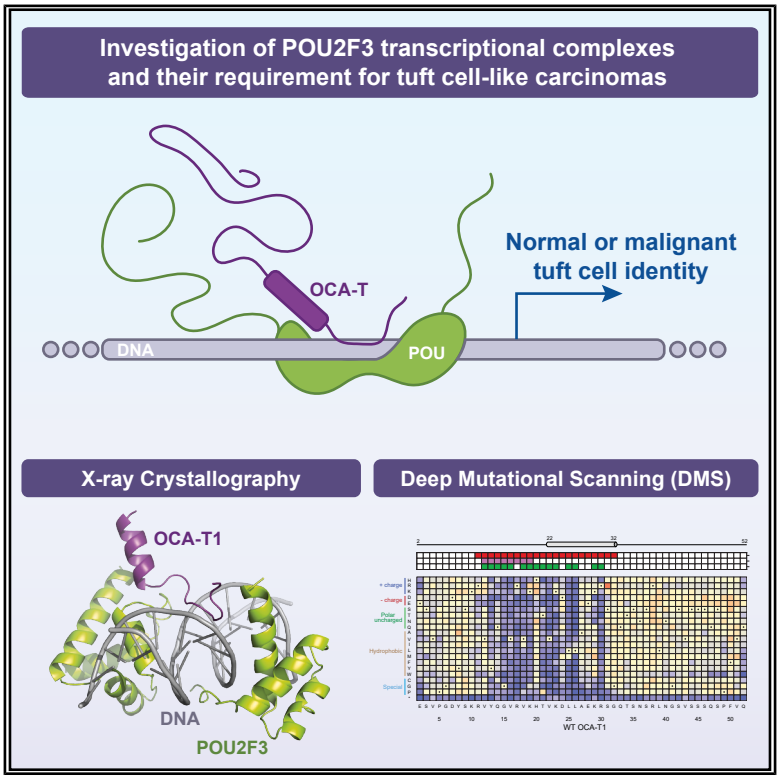
Cell Reports

Article

Structural basis of DNA-dependent coactivator recruitment by the tuft cell master regulator POU2F3

Graphical abstract



Highlights

- Validation of POU2F3 and OCA-T1 as tumor-maintenance dependencies
- Crystal structures of DNA-dependent POU2F3 complexes with OCA-T1 and OCA-T2
- Deep mutational scanning reveals functional hotspots of POU2F3 and OCA-T1
- A role for OCA-T1 in stabilizing POU2F3 chromatin occupancy

Authors

Aktan Alpsoy, Jonathan J. Ipsaro, Damianos Skopelitis, ..., Hyo-Eun C. Bhang, Leemor Joshua-Tor, Christopher R. Vakoc

Correspondence

leemor@cshl.edu (L.J.-T.), vakoc@cshl.edu (C.R.V.)

In brief

Crystal structures and deep mutational scanning experiments by Alpsoy et al. reveal how the transcriptional coactivators POU2F3 and OCA-T form a DNA-dependent complex required for tuft cell lineage specification. Protein features essential for the function of this complex are identified, uncovering a structural vulnerability in tuft cell-like lung cancers.

Alpsoy et al., 2025, Cell Reports *44*, 116572
November 25, 2025
https://doi.org/10.1016/j.celrep.2025.116572

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Structural basis of DNA-dependent coactivator recruitment by the tuft cell master regulator POU2F3

Aktan Alpsoy,[1,4] Jonathan J. Ipsaro,[1,2,4] Damianos Skopelitis,[1,4] Sujay Pal,[1] Frank S. Chung,[3] Shannon Carpenter,[3] John J. Desmarais,[1] Xiaoli S. Wu,[1] Kenneth Chang,[1] Matthew T. DiMare,[3] Erin Harten,[3] Staci Bergman,[3] Justin B. Kinney,[1] Jeffrey A. Engelman,[3] Hyo-Eun C. Bhang,[3] Leemor Joshua-Tor,[1,2,*] and Christopher R. Vakoc[1,5,*]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

[2]Howard Hughes Medical Institute, W.M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

[3]Treeline Biosciences, Watertown, MA 02472, USA

[4]These authors contributed equally

[5]Lead contact

*Correspondence: leemor@cshl.edu (L.J.-T.), vakoc@cshl.edu (C.R.V.)

https://doi.org/10.1016/j.celrep.2025.116572

SUMMARY

The transcription factor POU2F3 defines the identity of tuft cells and underlies a distinct molecular subtype of small cell lung cancer (SCLC). Although POU2F3 is considered undruggable, its activity critically depends on the coactivators OCA-T1 and OCA-T2. Here, we demonstrate that acute suppression of either POU2F3 or OCA-T1 induces regression of tuft cell-like SCLC xenografts *in vivo*. To explore the structural basis and druggability of this dependency, we determine crystal structures of POU2F3 bound to OCA-T1 or OCA-T2 in complex with DNA, revealing a tripartite, DNA-dependent interface. We further employ deep mutational scanning to assess the functional impact of 4,218 missense variants in POU2F3 and OCA-T1, uncovering both mutation-sensitive hotspots and structurally constrained regions critical for tumor cell fitness. These findings define a transcriptional complex that integrates DNA recognition with coactivator recruitment and nominate POU2F3-OCA-T as a structurally tractable vulnerability in tuft cell-like carcinomas.

INTRODUCTION

Most human cancers can be classified based on their resemblance to specific normal cell types, which is often reflected in the histological appearance or transcriptome of a tumor.[1] While cancer cell identity can be dynamic and evolve in response to selection pressures, it is often sufficiently stable to influence prognosis and guide therapeutic strategies. Importantly, many cancers remain dependent on lineage-defining transcription factors (TFs), such as those targeted by nuclear hormone receptor antagonists in hormone-driven tumors.[2] High-throughput genetic screening has validated that most cancer types exhibit essential dependencies on lineage-defining TFs, underscoring their broad potential as therapeutic targets.[3,4]

Studies from our group and others have described a subset of human carcinomas that resemble the tuft cell lineage.[5–7] Tuft cells are chemosensory epithelial cells found in mucosal tissues, where they contribute to microbial sensing and local immune activation.[8] POU2F3 (also known as Skn-1A or OCT11) is a master regulator of the tuft cell fate, as evidenced by the complete absence of tuft cells in *Pou2f3*-deficient mice.[9,10] Approximately 15% of small cell lung cancers (SCLCs) express POU2F3 and other tuft cell markers (termed "SCLC-P"), and POU2F3 is essential for the growth of these cells *in vitro* but dispensable in other SCLC subtypes.[5,11] High POU2F3 expression has also been observed in other lung cancer subtypes with histologic overlap to SCLC, including large cell neuroendocrine carcinomas and basaloid squamous cell carcinomas.[12] Moreover, POU2F3^{+} tuft cell-like cancers have been identified across a range of additional tissues—including the thymus, prostate, bladder, colon, pancreas, and pharynx—where they frequently co-express additional tuft lineage markers.[6,7,13] Because tuft cells play a specialized role in mucosal immunity against specific pathogens, POU2F3 is an attractive candidate for lineage-directed therapy in tuft cell-like carcinomas.

POU2F3 belongs to the Pit-Oct-Unc (POU) family of TFs, characterized by a bipartite DNA-binding domain (DBD) composed of a POU-specific (POU_S) and a POU-homeodomain (POU_H) subdomain, linked by an unstructured segment.[14–16] Recent work from our lab and others identified two tuft cell-specific coactivators—OCA-T1 (encoded by *POU2AF2*) and OCA-T2 (*POU2AF3*)—that interact with POU2F3 in a DNA sequence-dependent manner.[17–19] These proteins contain a conserved N-terminal region (also known as the OCA peptide) that is critical for binding to POU2F3, an interaction that requires a DNA octamer motif with an adenine at the fifth nucleotide position.[17] The OCA peptide of OCA-T1 and OCA-T2 is

homologous to a segment of OCA-B (encoded by *POU2AF1*), a known coactivator of POU2F1/OCT1 and POU2F2/OCT2.[20] While the OCA-B-POU2F1 interaction has been structurally characterized,[21] the structural basis for POU2F3 engagement with OCA-T1 and OCA-T2 has not been defined. Beyond their OCA peptide, OCA-T1 and OCA-T2 also harbor a poorly conserved C-terminal transactivation domain that likely recruits additional components of the transcriptional machinery.

In this study, we integrate structural and genetic approaches to elucidate the molecular basis of POU2F3 interactions with its coactivators and their requirement for tuft cell-like carcinoma. Using inducible knockdown models, we demonstrate that both POU2F3 and OCA-T1 are required for the maintenance of SCLC-P xenografts, reinforcing their relevance as therapeutic targets. To explore the tractability of this complex, we determined high-resolution X-ray crystal structures of POU2F3 bound to OCA-T1 or OCA-T2 and DNA. These structures validate that OCA-T1 and OCA-T2 are paralogs of OCA-B, which bind to POU TFs in a highly similar manner. We further demonstrate the biological significance of this binding interface by performing deep mutational scanning (DMS) of thousands of missense variants in POU2F3 and OCA-T1, uncovering numerous mutations that disrupt complex formation and SCLC-P tumor cell viability. This screen also revealed new alleles of POU2F3 that impair OCA-T1 binding, which we further characterized using epigenomic approaches. Together, these findings establish the structural and functional basis of POU2F3 coactivation and support the candidacy of this interaction as a drug target in tuft cell-like carcinomas.

RESULTS

Evaluation of POU2F3 and OCA-T1 dependency in established SCLC-P xenografts

CRISPR-Cas9 knockout of POU2F3 or OCA-T1 in SCLC-P cell lines prevents tumor formation when implanted in immunodeficient mice.[17] However, whether these factors are required to maintain established tumors—a critical attribute of a therapeutic target—has not been determined. To address this, we investigated the effects of inducible POU2F3 or OCA-T1 suppression in pre-established NCI-H211 xenografts (Figure 1A). We first identified potent short hairpin RNAs (shRNAs) and cloned them into lentiviral vectors engineered for doxycycline (dox)-inducible expression. *In vitro* validation confirmed efficient, dox-dependent knockdown of POU2F3 or OCA-T1 (Figures S1–S3). NCI-H211 cells harboring these constructs were implanted into immunodeficient mice, and tumors were allowed to grow to approximately 240 mm^3 before randomization to standard or dox-supplemented chow (Figure 1A). Serial tumor measurements revealed that inducible knockdown of either POU2F3 or OCA-T1 led to significant attenuation of tumor growth, with evidence of regression (Figures 1C–1F). To verify *in vivo* target suppression, we analyzed tumors from a parallel cohort of xenografted mice by RT-qPCR 4 days after dox administration. These analyses confirmed >80%–90% knockdown of POU2F3 or OCA-T1 relative to controls (Figures 1G and 1I). At the study endpoint (27 or 28 days post-induction), target gene expression was partially restored, suggesting selective pressure for shRNA escape *in vivo* (Figures 1H and 1J). Together, these results demonstrate that POU2F3 and OCA-T1 are not only required for tumor initiation but are also essential for the maintenance of established SCLC-P tumors, validating this TF-coactivator complex as a bona fide tumor maintenance dependency.

High-resolution crystal structures of POU2F3 bound to DNA and coactivators OCA-T1 or OCA-T2

To define the structural basis of POU2F3 function and evaluate its tractability as a therapeutic target, we determined high-resolution crystal structures of POU2F3 bound to the octamer DNA motif in complex with either OCA-T1 or OCA-T2. Based on prior mapping of minimal interaction domains (residues 1–52 of OCA-T1 and 1–49 of OCA-T2) and guided by the OCA-B:POU2F1 structure, we co-purified ternary complexes of POU2F3, DNA, and either coactivator (Figures 2A and 2B; S5A). We initially tested three distinct constructs of OCA-T2 (amino acids 1–38, 1–44, and 1–49) and two OCA-T1 constructs (amino acids 1–32 and 1–44) based on Chasmen et al. (1999). One OCA-T2 construct (1–38) formed clear, relatively large crystals that diffracted well. Subsequently, we designed three additional OCA-T1 constructs (6–43, 6–46, and 1–43) based on the homology and length of the OCA-T2 fragment (1–38). Large, uniformly shaped crystals grew for the OCA-T1 (6–46) construct in multiple crystallization conditions. Crystals diffracting to 1.7 Å (OCA-T1) and 2.1 Å (OCA-T2) resolution were ultimately obtained. Both crystals belong to space group $P2_1$ and contain either two (OCA-T1) or four (OCA-T2) complexes per asymmetric unit. Structures were refined to R_{work}/R_{free} values of 0.18/0.20 (OCA-T1) and 0.22/0.26 (OCA-T2; Table S1).

POU2F3 adopts the canonical bipartite architecture of POU family TFs, with structured POU_S and POU_H subdomains joined by a flexible, unresolved linker (Figures 2C and 2D). α-helices from both subdomains (α3 and α7) form base-specific hydrogen bonds along the DNA major groove, recognizing the octamer motif in a manner analogous to POU2F1.[22] The resolved regions of OCA-T1 and OCA-T2 consist of an N-terminal coil and a C-terminal α-helix, similar to OCA-B[21] (Figures 2C and 2D). The OCA-T1 and OCA-T2 α-helices dock into a hydrophobic cleft on the POU_S subdomain, while the N-terminal coil extends along the DNA backbone, forming multiple contacts with DNA bases and phosphate groups (Figures 2C and 2D). Surface area calculations revealed that OCA-T1 buries ∼1855 $Å^2$ and OCA-T2 buries ∼1515 $Å^2$ of solvent-accessible surface area across its interfaces with POU2F3 and DNA, consistent with a stable ternary complex.

Key residues in OCA-T1 and OCA-T2 mediate interactions with both POU2F3 and DNA (Figures 3A–3E; Figures 4A and 4B; S5). V22 and K23 in OCA-T1 (V17 and K18 in OCA-T2) form backbone hydrogen bonds with conserved leucines L235 and L237 in the turn connecting the α3 and α4 helices of POU_S (Figures 3A; S5A). Salt bridges are observed between POU2F3 E192 and OCA-T1 R30 (R25 in OCA-T2) and between POU2F3 E189 and OCA-T1 K29 (R24 in OCA-T2). L25 and L26 (L20 and L21 in OCA-T2) engage in hydrophobic interactions with POU2F3 residues L188 and L191 (Figures 3B; S5B). OCA-T1 and OCA-T2 also establish direct and water-mediated hydrogen bonds with DNA. The backbone amide of V16 (V11 in OCA-T2)

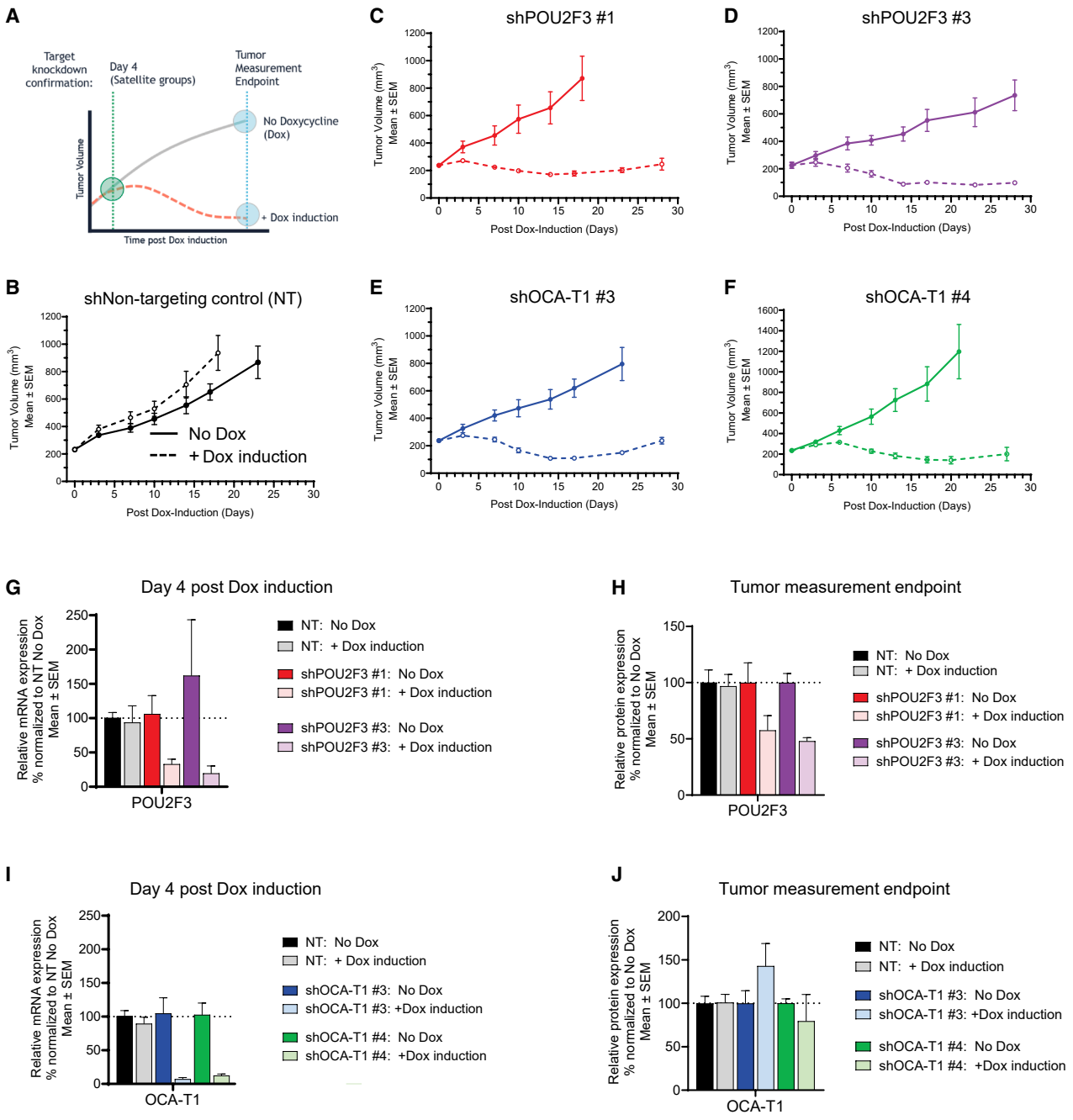


Figure 1. POU2F3 and OCA-T1 are essential to maintain established SCLC-P xenografts

(A) Experimental scheme of *in vivo* dependency validation. Satellite groups refer to auxiliary cohorts of mice to gather supporting information on shRNA knockdown efficiency, without compromising the integrity of the main study endpoints.

(B–F) Tumor volume measurement of NCI-H211 xenografts (n = 6) upon doxycycline induction. NCI-H211 was engineered to express non-targeting control (B), POU2F3 (C and D), or OCA-T1 (E and F) targeting shRNAs under the doxycycline-inducible promoter. Tumor volume was monitored up to day 27–28 (endpoint) post dox induction.

(G–J) Confirmation of POU2F3 or OCA-T1 knockdown at day 4 and days 27–28 (endpoint) post dox induction (n = 3 for each time point). RT-qPCR was performed using homogenized tumor lysates to measure the mRNA expression level of POU2F3 and OCA-T1. All error bars depict mean ± SEM.

hydrogen bonds with the position 5 adenine of the octamer motif, and its backbone carbonyl group has a water-mediated interaction with cytosine at position 4, while its side chain engages with thymidine (position 6′) through a non-bonded contact (Figures 3C; S5C). The V18 side chain further engages adjacent thymidines at positions 5′ and 6′ through non-bonded contacts

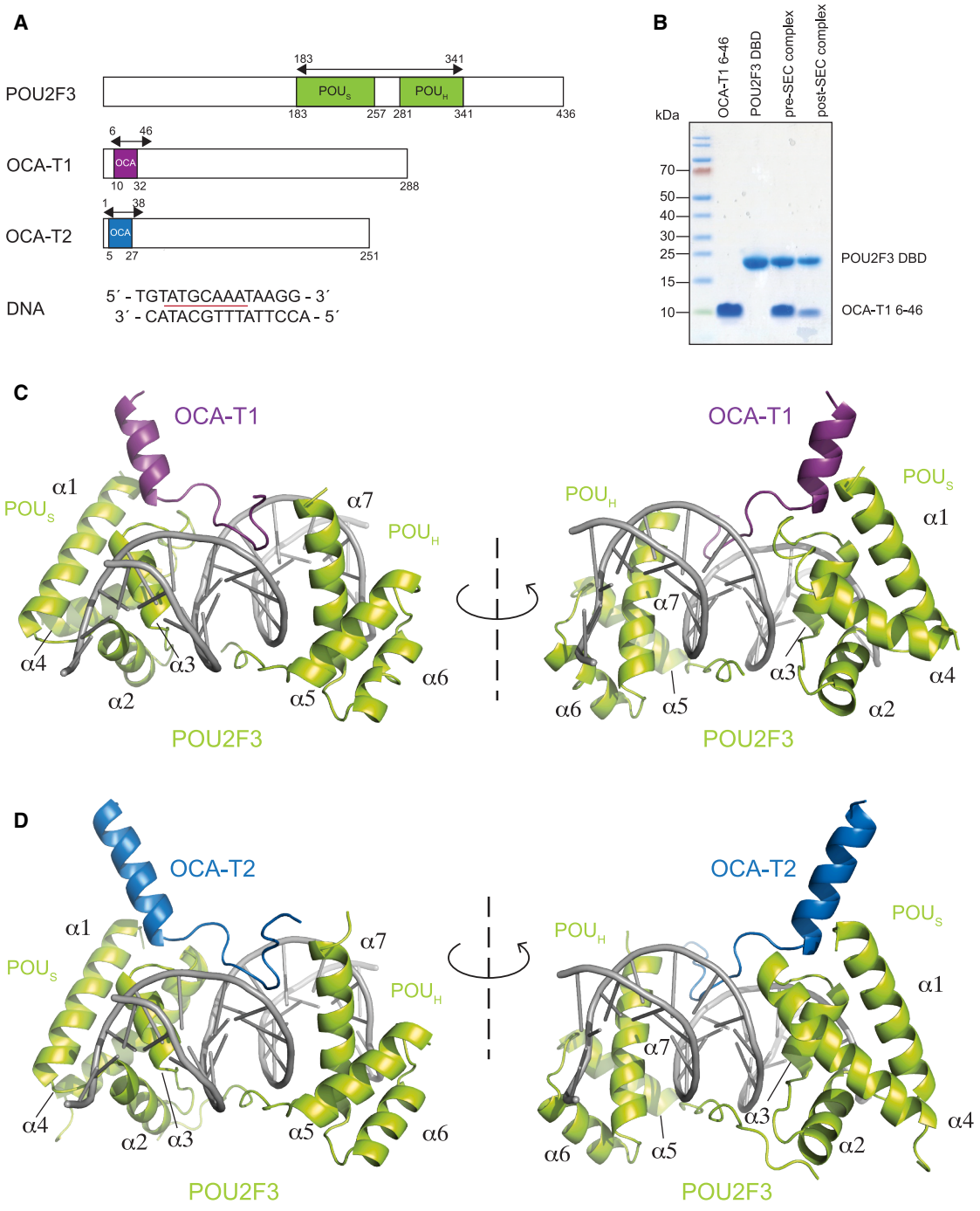


Figure 2. X-ray crystal structures of POU2F3 bound to coactivators OCA-T1 and OCA-T2 with DNA

(A) Domain schematic of POU2F3, OCA-T1, and OCA-T2. The recombinant fragments used for the structural studies are depicted with double-sided arrows; the numbers show the limits of the domains based on UniProt annotations. The double-stranded DNA with single-base overhangs included in the structure is also shown, with the core octamer sequence underlined.

(B) Representative SDS-PAGE image of purified fragments and reconstituted complexes before and after size-exclusion chromatography.

(C and D) Representation of OCA-T1 (C) and OCA-T2 (D) ternary complexes.

(Figures 3C; S5C). Notably, a π-cation interaction between Y13 and R17 appears to stabilize the register of the Q14-G15-V16 stretch, further reinforcing the DNA-binding interface (Figures 3D; S5D). POU2F3 residues also contribute to sequence-specific recognition: Q226 and T227 contact adenine (position 1), thymidine (position 2), and cytidine (position 3), while

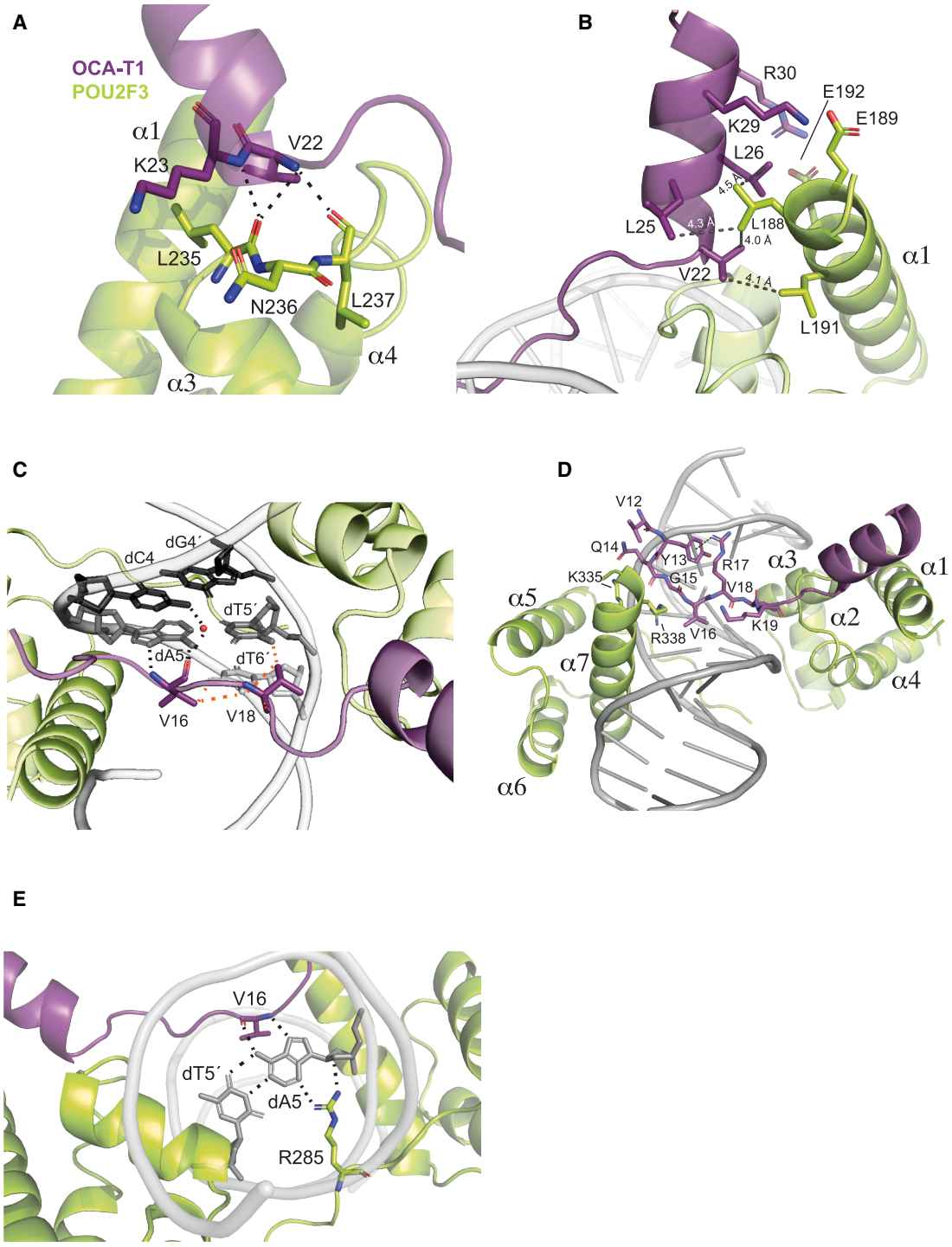


Figure 3. Critical interface residues of the POU2F3:OCA-T1:DNA ternary complex

(A) Critical backbone-backbone interactions between OCA-T1 and POU_S.

(B) The N-terminal residues of the POU_S domain engage with the α-helix of OCA-T1 through salt bridges or hydrophobic interactions.

(C) The sequence-specific interactions of critical valines at the N-terminal ordered region of OCA-T1.

(D) Stick model representation of the N-terminal ordered region of OCA-T1 along the major groove. A π-cation interaction between Y13 and R17 is displayed with a dashed line.

(E) The fifth adenine of the octamer sequence is hydrogen bonded (yellow dashed lines) with both the V16 backbone (OCA-T1) and guanidino group of R285 (POU2F3).

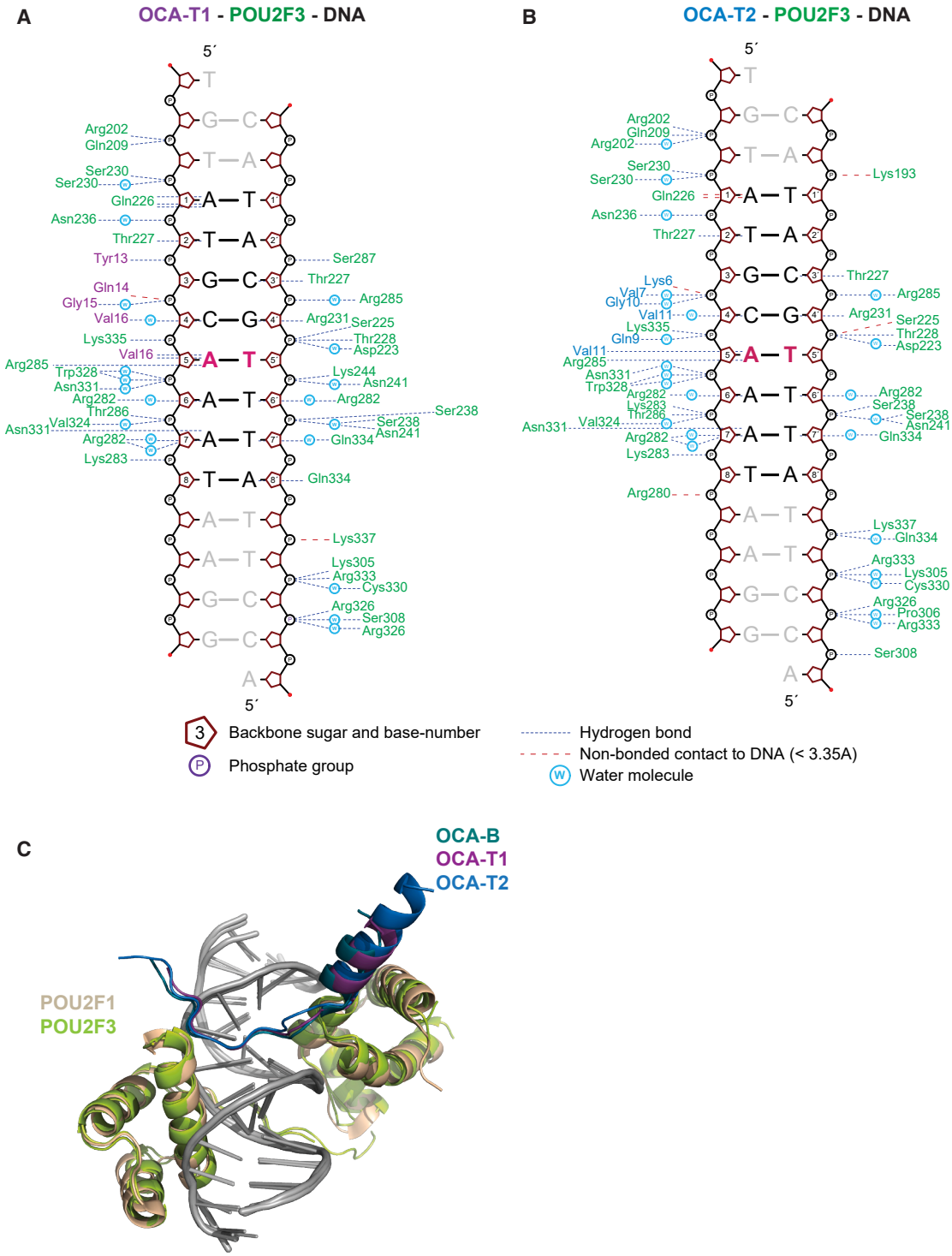


Figure 4. Summary of the POU2F3 and OCA-T1/T2 residues that directly or indirectly interact with DNA

(A and B) NUCPLOT diagram of (A) OCA-T1 and (B) OCA-T2 ternary complexes. Nucleobases within the core octamer motif are labeled in pink (adenine 5) or black (others) and are numbered according to their position within the octamer motif. Nucleobases outside the core octamer motif are labeled gray. OCA-T1 and OCA-T2 residues are indicated with purple or blue labels, respectively, while POU2F3 residues are shown in green. Water molecules that bridge protein-DNA interactions are shown as light blue circles. Hydrogen bonds are shown by blue dashes; non-bonded contacts are displayed as red dashes.

(C) Structural comparison of OCA-B-POU2F1-DNA (PDB: 1CQT) with the OCA-T1 and OCA-T2 ternary complexes.

R282 makes water-mediated contacts with adenine (position 6), thymidine (position 6′), and adenine (position 7). N331 makes a water-mediated contact with the backbone and a direct contact with adenine (position 7; Figures 4A and 4B). R285 in the POU_H subdomain contacts the same fifth-position adenine targeted by OCA-T1 V16 (OCA-T2 V11), anchoring the DNA through interactions in both the major and minor grooves (Figures 3E; 4A and 4B; S5E). Together, these observations reveal a highly coordinated binding interface in which OCA-T and POU2F3 cooperatively engage both DNA and each other to stabilize sequence-specific TF occupancy.

Structural alignment of the POU2F3:OCA-T1 and POU2F3:OCA-T2 complexes with the POU2F1:OCA-B structure yielded an RMSD of 0.8 Å across 138 aligned atoms, confirming strong conservation of the core architecture (Figure 4C). However, the higher resolution of our structures enabled more complete modeling of side chains and solvent interactions (Figures S4B–S4D). This refinement revealed water-mediated contacts, distinct α-helix geometries (three helical turns in OCA-T2 vs. two in OCA-T1 and OCA-B) and previously uncharacterized DNA contacts outside of the 8 base-pair octamer motif, such as those mediated by K337 (Figures 4A and 4B). These features clarify the atomic basis of coactivator engagement and offer a framework for structure-based inhibitor design targeting this transcriptional complex.

Finally, we used the CASTpFold algorithm to analyze potential cavities in the POU2F3:OCA-T1 structure (Figure S6).[23] Three solvent-accessible pockets were identified with shallow and/or extended interfaces, typical of DNA-TF complexes. The largest pocket (~1,022 $Å^3$) is surface-exposed, while a smaller pocket (~38 $Å^3$) lies within the major groove, bordered by the ordered OCA-T1 coil and the basic terminus of POU_H and OCA-T1 K19 (Figures S6A and S6B). These findings suggest limited, but potentially exploitable, pockets for ligand binding within this interface.

Functional dissection of the POU2F3-OCA-T1 interface by deep mutational scanning

To assess the functional relevance of our structural model in tuft cell-like carcinomas, we performed DMS of the POU2F3 DBD and the OCA peptide of OCA-T1. Using the SCLC-P cell line NCI-H211, we evaluated the fitness impact of all possible individual amino acid substitutions within these segments. For this purpose, we engineered a lentiviral vector that co-expresses an sgRNA targeting endogenous *POU2F3* or *OCA-T1* along with a CRISPR-resistant rescue cDNA (Figure 5A). Into this vector, we cloned mutagenized cDNA libraries encoding all missense and nonsense mutations across 51–70 amino acid windows spanning the POU2F3-OCA interface (Figure 5A). We generated four independent lentiviral libraries: three covering the POU2F3 DBD and one targeting the OCA-T1 OCA peptide, as well as 600–1,000 variants containing silent mutations within each window (Figure 5A). After transduction into $Cas9^+$ NCI-H211 cells, cDNA allele frequencies were quantified by deep sequencing of integrated cDNA, PCR-amplified from genomic DNA at early and late time points (Figure 5B). Loss-of-function variants were expected to decline over time due to their inability to rescue proliferation arrest induced by endogenous gene knockout (Figure 5B). Assay fidelity was supported by the consistent depletion of nonsense mutations and the neutrality of silent mutations, with high concordance between two independent biological replicates (Figures 5C and 5D).

Across the 4,218 missense mutations tested, we observed a broad spectrum of functional effects, revealing regional differences in mutational tolerance (Figures 5E and 6). Disordered regions absent from the crystal structures—such as the POU2F3 linker between POU_S and POU_H, and distal portions of OCA-T1—were generally permissive to substitution (Figures 5E and 6). In contrast, substitution-intolerant hotspots were pronounced within α3 and α7 helices of POU2F3 (which mediate DNA binding) and in structurally defined regions of OCA-T1 that engage DNA and POU2F3 (Figures 6A and 6B). These loss-of-function effects in the DMS screens showed a strong correlation with impaired POU2F3-OCA-T1 interaction detected using a NanoBRET assay (measured across 37 missense mutations), further validating the structure and reliability of the DMS data (Figures 5F; S7). Importantly, these findings provide strong evidence that the central segment of the OCA peptide responsible for binding POU2F3 is essential for the viability of SCLC-P tumor cells.

Within the OCA-T1 α-helix, residues on the solvent-facing side that do not directly contact POU2F3—such as D24, A27, and E28—were generally tolerant of substitution (Figure 6A). However, proline substitutions at these positions were consistently deleterious (Figure 6A), likely due to α-helix-disrupting effects, indicating that structural integrity, rather than specific side-chain interactions, is critical at these sites. One notable residue was H20, which exhibited broad mutational tolerance. This residue is not conserved among OCA paralogs, and neither its backbone nor side chain forms defined contacts in the crystal structure, consistent with its apparent dispensability. In contrast, K19 (a conserved residue in the OCA family) showed a strict requirement for a positively charged side chain, with only arginine tolerated as a substitute (Figure 6A). Although side-chain density for K19 is weak in the crystal structure (data not shown), its critical role inferred from DMS suggests that its charge contributes to a basic surface at the OCA-DNA interface, potentially important for recruiting additional nuclear factors. The absence of strong density may reflect a flexible conformation.

Surprisingly, K29, which is located within the OCA-T1 α-helix and predicted to form salt bridges with acidic residues on POU_S, was highly permissive to substitution. Several variants, including K29L, K29Y, and K29W, supported cell proliferation, and the K29M variant even outperformed wild-type OCA-T1 in promoting SCLC-P fitness and in interaction strength detected using NanoBRET assays (Figures 5E and 6A). These effects may arise from enhanced hydrophobic packing at the POU_S-OCA interface, which is rich in conserved leucines at positions 25 and 26 of OCA-T1. Notably, substitutions at L25 or L26 of OCA-T1 were uniformly deleterious, consistent with their tight packing against POU_S and intolerance to steric disruption (Figure 6A). Finally, R30—the most C-terminal residue of the OCA peptide and conserved across all OCA family members—was completely intolerant to substitution (Figure 6A). This residue forms a salt bridge in the crystal structure, and its strict conservation and essentiality in the DMS screen underscore its indispensable role in stabilizing the POU2F3-OCA-DNA complex.

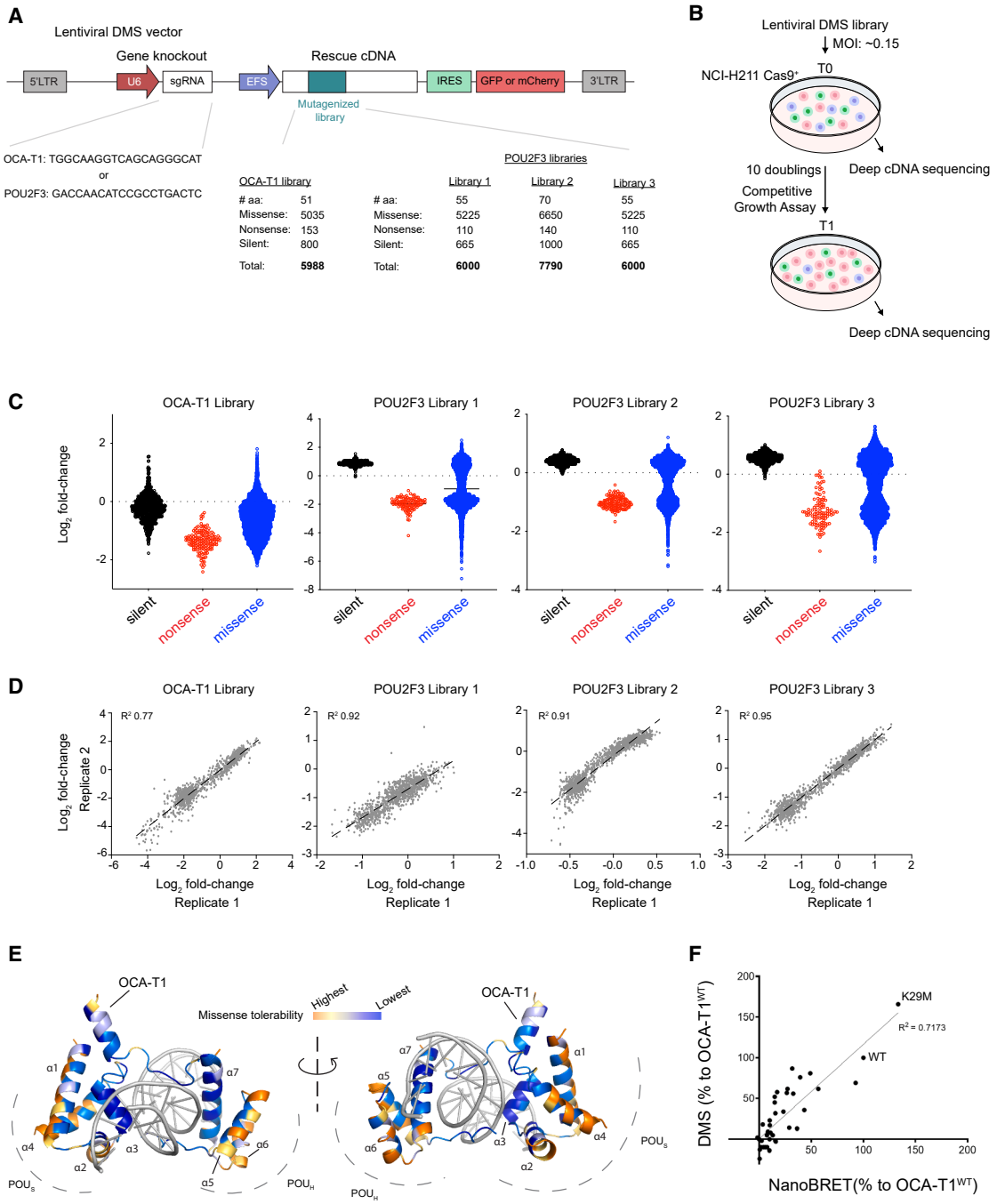


Figure 5. Deep mutational scanning strategy and controls

(A) Summary of deep mutagenesis strategy. The numbers of independent lentiviral constructs are indicated for each of the four libraries, which evaluate the effects of 4,218 missense mutations.

(B) Summary of the library transduction and drop-out screen strategy.

(C) The average dropout effects of silent, missense, and nonsense substitutions across different libraries.

(D) Correlation plots depicting the concordance between two independent screens for each library.

(E) Structural visualization of position-specific missense tolerability. The missense tolerability metric assignment is based on the number of substitutions that lead to growth defects compared to wild-type peptide. Positions are colored based on their permissiveness to substitutions (orange being more permissive; blue having restricted tolerance to substitutions).

(F) Concordance between NanoBRET assay and DMS for the OCA-T1 OCA peptide. Each dot represents a single substitution. The binding and viability scores are expressed as percent binding (NanoBRET) or percent rescue (DMS) relative to wild-type OCA-T1.

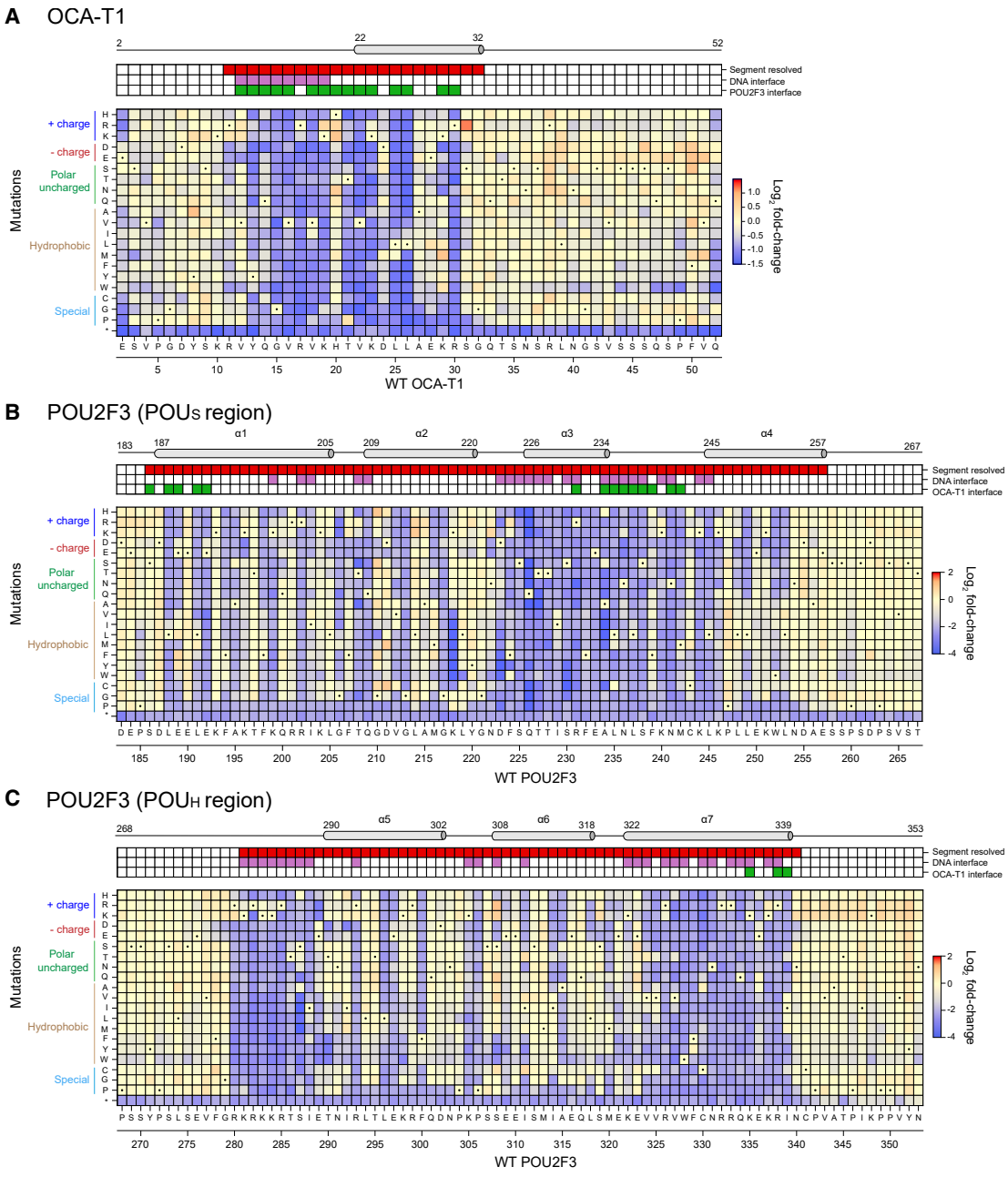


Figure 6. Deep mutational scanning results of OCA-T1 and POU2F3

(A–C) DMS results for (A) the OCA-T1 OCA peptide and flanking sequences, (B) the POU2F3 POU_S domain, and (C) the POU2F3 POU_H domain are shown. Secondary structure elements are displayed above each heatmap. The residues that are resolved in the crystal structure are indicated with red (top bar); residues at the DNA interface are indicated with lilac (second bar); residues at the protein-protein interface are indicated with green (bottom bar). The heatmaps summarize the position-specific growth effects of amino acid or stop codon substitutions relative to the wild-type protein. Substitutions are shown along the vertical axis; UniProt-annotated positions and wild-type sequences of the mutagenesis window are shown along the horizontal axis. Silent mutations are denoted by dots.

Select α-helix 1 mutations of POU2F3 retain DNA binding *in vitro* but reduce chromatin occupancy at specific *cis*-regulatory elements in SCLC-P cells

To further elucidate the role of OCA-T1 as a coactivator, we sought to identify separation-of-function mutations of POU2F3 that preserve DNA binding *in vitro* but reduce binding to OCA-T1. Genomic characterization of such an allele would provide a deeper mechanistic insight into how OCA-T1 supports POU2F3 function *in vivo*. Structural analysis identified the LEELE motif (residues 188–192) within the α1 helix of the POU_S domain as a key contact surface for OCA-T1, distinct from the DNA-binding interface (Figures 3A and 3B). DMS further

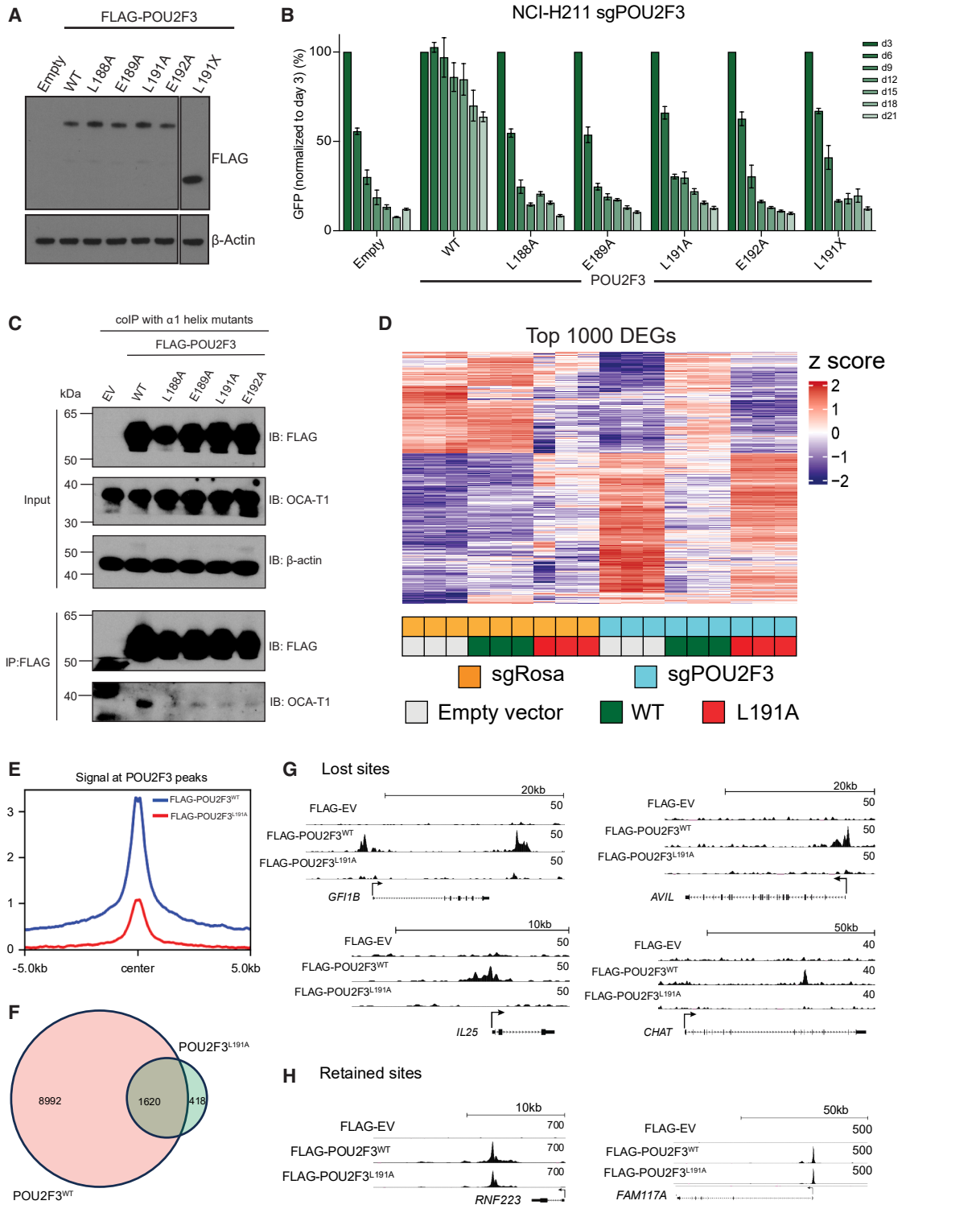


Figure 7. POU2F3 mutations retain DNA binding *in vitro* but reduce chromatin occupancy at specific *cis*-regulatory elements

(A) Western blot showing the expression of FLAG-tagged POU2F3 constructs used in the competition-based growth assay.

(B) Competition-based growth assay in *POU2F3* knockout NCI-H211 cells that exogenously express CRISPR-resistant mutant and wild-type POU2F3 (n = 3). Data is shown as mean ± SEM.

(C) Co-immunoprecipitation assay for the POU2F3-OCA-T1 interaction in cell lines that exogenously express mutant or wild-type POU2F3.

(legend continued on next page)

supported the functional importance of this motif: substitutions at L188, E189, L191, or E192 produced strong loss-of-function phenotypes comparable to the effect of nonsense mutations (Figure 6B). To validate these observations, we engineered FLAG-tagged, CRISPR-resistant POU2F3 cDNAs containing alanine substitutions at each of the candidate residues and assessed their ability to rescue the proliferation arrest induced by CRISPR-mediated knockout of endogenous POU2F3 in NCI-H211 cells. In competition-based proliferation assays, the L188A, E189A, L191A, and E192A mutants failed to support cell proliferation, despite robust protein expression confirmed by western blotting (Figures 7A and 7B). Co-immunoprecipitation assays revealed impaired binding to OCA-T1 for each mutant (Figure 7C). For simplicity, we focused our subsequent genomic analyses on the L191A mutant of POU2F3. *In vitro* assays using size exclusion chromatography and DNA pull-down assays demonstrated that DNA binding by $POU2F3^{L191A}$ was largely intact, consistent with the crystal structure placing L191 distal to the DNA-binding interface (Figure S8).

To assess the downstream effects of OCA-T1 in a cellular context, we compared the genomic function of $POU2F3^{L191A}$ with wild-type POU2F3. RNA-seq analysis in NCI-H211 cells revealed that the expression of $POU2F3^{L191A}$ cDNA suppressed endogenous POU2F3 target genes, suggesting a dominant-negative loss-of-function effect (Figures 7D; S9). In a gene-complementation RNA-seq assay, we expressed either wild-type or L191A mutant POU2F3 cDNA in NCI-H211 cells followed by CRISPR-mediated inactivation of the endogenous POU2F3 (Figures 7D; S9). While wild-type POU2F3 cDNA rescued most transcriptional changes induced by POU2F3 knockout, the L191A mutant failed to do so (Figures 7D; S9). CUT&RUN sequencing performed using an anti-FLAG antibody revealed that the L191A mutation markedly reduced genome-wide occupancy of POU2F3 (Figures 7E; S10). Peak calling analysis showed that ~85% of POU2F3 peaks were lost upon introduction of the L191A mutation (e.g., *GFI1B* and *AVIL*; Figures 7F and 7G; S10), whereas ~15% of POU2F3-binding sites were unaffected by this mutation (e.g., *RNF223* and *FAM117A*; Figures 7F and 7H; S10). These results suggest that OCA-T1 can promote chromatin association of POU2F3 at specific sites *in vivo*, even though it is dispensable for DNA binding *in vitro*.[17]

DISCUSSION

This study defines the structural and functional mechanism of POU2F3 coactivators in tuft cell-like SCLC through a combination of *in vivo* functional genetics, structural biology, DMS, and epigenomics. Using inducible RNA interference in xenograft models, we demonstrate that both POU2F3 and its coactivator OCA-T1 are essential for the maintenance of established tumors, validating this TF-coactivator complex as a bona fide dependency in SCLC-P. Structural analysis of POU2F3 bound to DNA and OCA-T1 or OCA-T2 revealed a conserved tripartite interface, stabilized by coordinated protein-protein and protein-DNA interactions. DMS confirmed the functional importance of residues at this interface for SCLC-P fitness, showing that mutational intolerance precisely maps to regions predicted to mediate DNA binding and coactivator engagement. These findings collectively establish the POU2F3-OCA-T interaction as an obligate, structurally defined complex that supports the growth and survival of tuft cell-like cancers.

Our structures of the POU2F3-OCA-T complexes reveal a DNA-dependent interaction in which both the TF and its coactivator simultaneously engage the same DNA element. This mode of binding is reminiscent of other DNA-templated TF-cofactor assemblies, such as the interaction between Hox proteins and Pbx cofactors[24] or between IRF4 and PU.1 on composite DNA elements,[25] where cooperativity is achieved through DNA-mediated spatial alignment of binding surfaces. However, the POU2F3-OCA-T interface exhibits architectural features not commonly seen in those paradigms. Unlike Hox-Pbx and IRF4-PU.1 complexes, where both partners contribute independent DBDs that recognize adjacent motifs, OCA-T lacks a canonical DBD and instead interacts with both the POU2F3 protein (via both backbone and side chain interactions) and the octamer motif DNA via a short, conserved OCA peptide. Compared to coactivator interactions mediated by TF activation domains (e.g., p53-CBP[26] or Myb-p300[27]), which typically rely on short linear motifs interacting with globular coactivator domains in a DNA-independent manner, the POU2F3-OCA-T interaction is exceptional in its strict reliance on DNA binding to stabilize the interface. This DNA dependence likely imparts additional specificity and context sensitivity to cofactor recruitment, ensuring that OCA-T engagement is tightly coupled to POU2F3 occupancy of its cognate octamer motif and correct positioning of its two domains upon DNA binding. The structure thus exemplifies a strategy by which transcriptional regulators integrate DNA recognition with cofactor recruitment to ensure precise gene activation, thereby expanding our understanding of how coactivator proteins can achieve high-affinity, structurally constrained binding through DNA-templated assembly.

DMS revealed a broad spectrum of functional consequences for amino acid substitutions in POU2F3 and OCA-T1, highlighting regions of both high- and low mutational tolerance. As expected, residues integral to the tripartite interface were generally intolerant to substitution, underscoring their essential role in maintaining complex stability and transcriptional output. However, we also identified amino acids outside the tripartite interface that are sensitive to specific mutations. Changes such as S3M and P49W of OCA-T1 suggest potential sites involved in yet-to-be-discovered protein-protein interactions, allosteric regulation, or maintenance of protein folding. Moreover, while most deleterious mutations identified by DMS reflect

(D) Heatmap depicting the relative expression changes of the top 1,000 differentially expressed genes in control (sgROSA) or *POU2F3* knockout (sgPOU2F3) cells that were transduced with empty vector, wild-type POU2F3 (WT), or $POU2F3^{L191A}$ expression constructs.
(E) Relative enrichment of wild-type or mutant ($POU2F3^{L191A}$) POU2F3 at high-confidence POU2F3 binding sites, as determined by CUT&RUN analysis.
(F–H) The overlap between POU2F3 peaks in $POU2F3^{WT}$ and $POU2F3^{L191A}$ cells, as determined from CUT&RUN analysis. Example regions where POU2F3 peaks are lost (G) or retained (H) by the L191A mutation.

loss of function, we also observed clear examples of gain-of-function and dominant-negative variants. For example, we validated that the K29M mutation of OCA-T1 enhances POU2F3 binding and may result in hyperactivation of target genes, particularly if negative feedback or chromatin constraints are bypassed. Future experiments could help distinguish between these mechanistic classes and define the full spectrum of mutational consequences observed in the screen. In addition, an investigation of pairwise mutational effects in this assay could reveal redundancies among side interactions within the binding interface.

Notably, our epigenomic characterization of POU2F3^{L191A} suggests that OCA-T1 can facilitate POU2F3 engagement with chromatinized DNA rather than with naked DNA for at least a subset of genomic binding sites. One possibility is that OCA-T1 promotes local chromatin accessibility by recruiting SWI/SNF chromatin remodeling complexes, which are known to support POU2F3 function in SCLC-P tumor models.[28,29] Alternatively, OCA-T1 might influence POU2F3 occupancy of partially occluded octamer motifs on the surface of the nucleosome via direct distortion of DNA or nucleosome structure (independently of other nuclear proteins).[30] In either of these models, the loss of OCA-T1 binding caused by the L191A mutation would impair the ability of POU2F3 to engage nucleosome-bound DNA *in vivo*, even if its intrinsic DNA-binding capacity remains largely intact. These findings underscore the importance of considering chromatin context when interpreting TF function and suggest that coactivators like OCA-T1 can act as gatekeepers for TF access to genomic DNA.

From a therapeutic standpoint, the structural and functional data presented here highlight several potential vulnerabilities in the POU2F3-OCA-T complex that may be amenable to pharmacologic disruption. Although TFs and their coactivators are classically considered ''undruggable,'' our structural studies reveal defined surfaces and residues that mediate high-affinity cooperative binding. Our findings suggest that disruption of any one connection—TF-coactivator, TF-DNA, or coactivator-DNA—would destabilize the essential output of POU2F3 in SCLC-P tumors. The DMS data further pinpointed hotspot residues within both POU2F3 and OCA-T1 that are exquisitely sensitive to single amino-acid substitutions, suggesting that these positions may correspond to energetically constrained contacts that are particularly susceptible to steric interference or conformational disruption. Our structure reveals a shallow but highly conserved groove on POU2F3 that cradles the OCA peptide, representing a potential ligandable surface. Given the limited depth of this pocket, macrocyclic compounds or peptidomimetics may offer viable strategies to disrupt OCA-T binding. More broadly, this work provides a mechanistic foundation for targeting lineage-specific transcriptional dependencies in cancer and illustrates how integrative structural and genetic analysis might render such targets more tractable to therapeutic development.

Limitations of the study

In this project, we used crystallography and DMS to define the biochemical and structural basis of tuft lineage-specific transcriptional complexes and their functions in cancer. A limitation of our study is that we focused only on the structurally resolvable segments of the ternary complex, as both POU2F3 and OCA-T1/2 contain predicted intrinsically disordered regions outside these resolved portions. In addition, the coverage and depth required for DMS prevented us from extending the scanning window across the entire polypeptides, which may contain additional biologically important segments. Our analysis also centered on perturbations within the ternary complex itself. We did not test whether amino acid substitutions alter interactions with other cofactors, such as chromatin remodelers, despite recent studies suggesting that OCA-T1 and POU2F3 cooperate with such factors.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Christopher R. Vakoc (vakoc@cshl.edu).

Materials availability

All unique reagents generated in this study will be available from the lead contact upon request.

Data and code availability

- Structures and data for the OCA-T2 (residues 1–38) + POU2F3 (DBD) + DNA [low-resolution] complex, the OCA-T2 (residues 1–38) + POU2F3 (DBD) + DNA [high-resolution] complex, and the OCA-T1 (residues 6–46) + POU2F3 (DBD) + DNA complex were deposited in the Protein Data Bank with accession codes PDB: 9PFN, PDB: 9PFO, and PDB: 9PFP, respectively.
- RNA-seq data have been deposited at NCBI GEO with accession code GEO: GSE302002. CUT&RUN-seq data have been deposited at NCBI GEO with accession code GEO: GSE302003. Deep mutational scanning (DMS) data have been deposited at NCBI GEO with accession code GEO: GSE302004.
- The code required to produce DMS library sequences and to analyze library results can be found at https://github.com/jackdesmarais/DMS_Designer, and a version of record is saved at https://zenodo.org/records/15800486 [DOI: https://doi.org/10.5281/zenodo.15800486].
- Uncropped western blot images were deposited to Mendeley Data with DOI number 10.17632/4fdmy5cx6s.1

ACKNOWLEDGMENTS

We thank the Vakoc and Leemor Joshua-Tor lab members for discussions and suggestions throughout the course of this study. This work was supported by the Cold Spring Harbor Laboratory NCI Cancer Center support grant CA045508. Additional funding was provided to C.R.V. by the Pershing Square Sohn Cancer Research Alliance, Treeline Biosciences, National Institutes of Health grants CA013106 and CA242919, Department of Defense grant W81XWH1910317, and the Cold Spring Harbor Laboratory and Northwell Health Affiliation. D.S. is supported by a National Institutes of Health R50 grant CA305054. L.J. is a Howard Hughes Medical Institute Investigator. Additional funding was provided to J.B.K. by the National Institutes of Health grants GM133777 and HG011787.

AUTHOR CONTRIBUTIONS

A.A., J.J.I., D.S., L.J., and C.R.V. conceived the project. A.A. prepared and crystallized the complexes, collected and analyzed the crystallography data, and wrote the first draft. J.J.I. supervised the complex crystallization, collection, and analysis of crystallography data. D.S. designed the deep mutational scanning approach with input from J.J.D., X.S.W., J.B.K., and K.C.; executed and analyzed the screen with help from J.J.D.; and

performed and analyzed RNA-seq. S.P. performed the co-IP and binding assays with input from A.A. and executed and analyzed CUT&RUN experiments. F.S.C. and S.C. performed the tumor xenograft experiments. J.J.D. helped design the deep mutational scanning libraries and provided support for the analysis of the screen. M.T.D., E.H., and S.B. performed the NanoBRET assays. C.R.V., L.J., H.-E.C.B., J.A.E., and J.B.K. supervised the project and acquired funding for the work. C.R.V. prepared the final draft with input from all authors.

DECLARATION OF INTERESTS

C.R.V. has received consulting fees from Flare Therapeutics, Roivant Sciences, and C4 Therapeutics; has served on the advisory boards of KSQ Therapeutics, Syros Pharmaceuticals, and Treeline Biosciences; has received research funding from Boehringer Ingelheim and Treeline Biosciences; and owns stock in Treeline Biosciences.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT to revise and proofread the draft of this manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116572.

Received: July 10, 2025
Revised: September 6, 2025
Accepted: October 29, 2025
Published: November 18, 2025

REFERENCES

1. Garraway, L.A., and Sellers, W.R. (2006). Lineage dependency and lineage-survival oncogenes in human cancer. Nat. Rev. Cancer *6*, 593–602. https://doi.org/10.1038/nrc1947.
2. Le Magnen, C., Shen, M.M., and Abate-Shen, C. (2018). Lineage Plasticity in Cancer Progression and Treatment. Annu. Rev. Cancer Biol. *2*, 271–289. https://doi.org/10.1146/annurev-cancerbio-030617-050224.
3. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010.
4. Bhagwat, A.S., and Vakoc, C.R. (2015). Targeting Transcription Factors in Cancer. Trends Cancer *1*, 53–65. https://doi.org/10.1016/j.trecan.2015.07.001.
5. Huang, Y.H., Klingbeil, O., He, X.Y., Wu, X.S., Arun, G., Lu, B., Somerville, T.D.D., Milazzo, J.P., Wilkinson, J.E., Demerdash, O.E., et al. (2018). POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. Genes Dev. *32*, 915–928. https://doi.org/10.1101/gad.314815.118.
6. Yamada, Y., Simon-Keller, K., Belharazem-Vitacolonnna, D., Bohnenberger, H., Kriegsmann, M., Kriegsmann, K., Hamilton, G., Graeter, T., Preissler, G., Ott, G., et al. (2021). A Tuft Cell-Like Signature Is Highly Prevalent in Thymic Squamous Cell Carcinoma and Delineates New Molecular Subsets Among the Major Lung Cancer Histotypes. J. Thorac. Oncol. *16*, 1003–1016. https://doi.org/10.1016/j.jtho.2021.02.008.
7. Koh, J., Kim, H., Moon, K.C., Lee, C., Lee, K., Ryu, H.S., Jung, K.C., and Jeon, Y.K. (2023). Molecular Classification of Extrapulmonary Neuroendocrine Carcinomas With Emphasis on POU2F3-positive Tuft Cell Carcinoma. Am. J. Surg. Pathol. *47*, 183–193. https://doi.org/10.1097/PAS.0000000000001977.
8. Schneider, C., O'Leary, C.E., and Locksley, R.M. (2019). Regulation of immune responses by tuft cells. Nat. Rev. Immunol. *19*, 584–593. https://doi.org/10.1038/s41577-019-0176-x.
9. Gerbe, F., Sidot, E., Smyth, D.J., Ohmoto, M., Matsumoto, I., Dardalhon, V., Cesses, P., Garnier, L., Pouzolles, M., Brulin, B., et al. (2016). Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. Nature *529*, 226–230. https://doi.org/10.1038/nature16527.
10. Yamashita, J., Ohmoto, M., Yamaguchi, T., Matsumoto, I., and Hirota, J. (2017). Skn-1a/Pou2f3 functions as a master regulator to generate Trpm5-expressing chemosensory cells in mice. PLoS One *12*, e0189340. https://doi.org/10.1371/journal.pone.0189340.
11. Rudin, C.M., Poirier, J.T., Byers, L.A., Dive, C., Dowlati, A., George, J., Heymach, J.V., Johnson, J.E., Lehman, J.M., MacPherson, D., et al. (2019). Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat. Rev. Cancer *19*, 289–297. https://doi.org/10.1038/s41568-019-0133-9.
12. Baine, M.K., Febres-Aldana, C.A., Chang, J.C., Jungbluth, A.A., Sethi, S., Antonescu, C.R., Travis, W.D., Hsieh, M.S., Roh, M.S., Homer, R.J., et al. (2022). POU2F3 in SCLC: Clinicopathologic and Genomic Analysis With a Focus on Its Diagnostic Utility in Neuroendocrine-Low SCLC. J. Thorac. Oncol. *17*, 1109–1121. https://doi.org/10.1016/j.jtho.2022.06.004.

13. Brady, N.J., Bagadion, A.M., Singh, R., Conteduca, V., Van Emmenis, L., Arceci, E., Pakula, H., Carelli, R., Khani, F., Bakht, M., et al. (2021). Temporal evolution of cellular heterogeneity during the progression to advanced AR-negative prostate cancer. Nat. Commun. *12*, 3372. https://doi.org/10.1038/s41467-021-23780-y.

14. Ryan, A.K., and Rosenfeld, M.G. (1997). POU domain family values: Flexibility, partnerships, and developmental codes. Genes Dev. *11*, 1207–1225. https://doi.org/10.1101/gad.11.10.1207.

15. Sturm, R.A., and Herr, W. (1988). The Pou Domain Is a Bipartite DNA-Binding Structure. Nature *336*, 601–604. https://doi.org/10.1038/336601a0.

16. Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G., et al. (1988). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. Genes Dev. *2*, 1513–1516. https://doi.org/10.1101/gad.2.12a.1513.

17. Wu, X.S., He, X.Y., Ipsaro, J.J., Huang, Y.H., Preall, J.B., Ng, D., Shue, Y.T., Sage, J., Egeblad, M., Joshua-Tor, L., and Vakoc, C.R. (2022). OCA-T1 and OCA-T2 are coactivators of POU2F3 in the tuft cell lineage. Nature *607*, 169–175. https://doi.org/10.1038/s41586-022-04842-7.

18. Szczepanski, A.P., Tsuboyama, N., Watanabe, J., Hashizume, R., Zhao, Z., and Wang, L. (2022). POU2AF2/C11orf53 functions as a coactivator of POU2F3 by maintaining chromatin accessibility and enhancer activity. Sci. Adv. *8*, eabq2403. https://doi.org/10.1126/sciadv.abq2403.

19. Zhou, C., Huang, H., Wang, Y., Sendinc, E., and Shi, Y. (2022). Selective regulation of tuft cell-like small cell lung cancer by novel transcriptional co-activators C11orf53 and COLCA2. Cell Discov. *8*, 112. https://doi.org/10.1038/s41421-022-00470-7.

20. Luo, Y., Fujii, H., Gerster, T., and Roeder, R.G. (1992). A novel B cell-derived coactivator potentiates the activation of immunoglobulin promoters by octamer-binding transcription factors. Cell *71*, 231–241. https://doi.org/10.1016/0092-8674(92)90352-d.

21. Chasman, D., Cepek, K., Sharp, P.A., and Pabo, C.O. (1999). Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: specific recognition of a protein-DNA interface. Genes Dev. *13*, 2650–2657. https://doi.org/10.1101/gad.13.20.2650.

22. Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., and Pabo, C.O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell *77*, 21–32. https://doi.org/10.1016/0092-8674(94)90231-3.

23. Ye, B., Tian, W., Wang, B., and Liang, J. (2024). CASTpFold: Computed Atlas of Surface Topography of the universe of protein Folds. Nucleic Acids Res. *52*, W194–W199. https://doi.org/10.1093/nar/gkae415.

24. LaRonde-LeBlanc, N.A., and Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes Dev. *17*, 2060–2072. https://doi.org/10.1101/gad.1103303.

25. Escalante, C.R., Brass, A.L., Pongubala, J.M.R., Shatova, E., Shen, L., Singh, H., and Aggarwal, A.K. (2002). Crystal structure of PU.1/IRF-4/DNA ternary complex. Mol. Cell *10*, 1097–1105. https://doi.org/10.1016/s1097-2765(02)00703-7.

26. Lee, C.W., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2010). Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. Biochemistry *49*, 9964–9971. https://doi.org/10.1021/bi1012996.

27. De Guzman, R.N., Goto, N.K., Dyson, H.J., and Wright, P.E. (2006). Structural basis for cooperative transcription factor binding to the CBP coactivator. J. Mol. Biol. *355*, 1005–1013. https://doi.org/10.1016/j.jmb.2005.09.059.

28. Duplaquet, L., So, K., Ying, A.W., Pal Choudhuri, S., Li, X., Xu, G.D., Li, Y., Qiu, X., Li, R., Singh, S., et al. (2024). Mammalian SWI/SNF complex activity regulates POU2F3 and constitutes a targetable dependency in small cell lung cancer. Cancer Cell *42*, 1352–1369.e13. https://doi.org/10.1016/j.ccell.2024.06.012.

29. He, T., Xiao, L., Qiao, Y., Klingbeil, O., Young, E., Wu, X.S., Mannan, R., Mahapatra, S., Redin, E., Cho, H., et al. (2024). Targeting the mSWI/SNF complex in POU2F-POU2AF transcription factor-driven malignancies. Cancer Cell *42*, 1336–1351.e9. https://doi.org/10.1016/j.ccell.2024.06.006.

30. Michael, A.K., Grand, R.S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., Bunker, R.D., Schenk, A.D., Graff-Meyer, A., Pathare, G.R., et al. (2020). Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science (New York, N.Y.) *368*, 1460–1465. https://doi.org/10.1126/science.abb0074.

31. Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr. D Biol. Crystallogr. *66*, 133–144. https://doi.org/10.1107/S0907444909047374.

32. Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. *62*, 72–82. https://doi.org/10.1107/S0907444905036693.

33. Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. *69*, 1204–1214. https://doi.org/10.1107/S0907444913000061.

34. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674. https://doi.org/10.1107/S0021889807021206.

35. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132. https://doi.org/10.1107/S0907444904019158.

36. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. *75*, 861–877. https://doi.org/10.1107/S2059798319011471.

37. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315. https://doi.org/10.1002/pro.3330.

38. Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (1997). NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. Nucleic Acids Res. *25*, 4940–4945. https://doi.org/10.1093/nar/25.24.4940.

39. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

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

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

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

43. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

44. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-GFP-HRP	Abcam	Cat# ab6663; RRID:AB_305636
Anti-FLAG antibody-HRP (Mouse)	Sigma Aldrich	Cat# A8592; RRID:AB_439702
OCA-T1 (E4A8S) Rabbit mAb	Cell Signaling Technology	Cat# 89883S
POU2F3	Cell Signaling Technology	Cat# 36135S; RRID:AB_2924784
anti-FLAG	Epicypher	Cat# 13-2031
IgG control	Epicypher	Cat# 13–0042; RRID:AB_2923178
OCA-T1 (Western Blot)	In-house (CSHL)	N/A
POU2F3 (Western Blot)	Cell Signaling Technology	Cat #92579S; RRID:AB_3096173
Vinculin	Sigma	Cat# V9264; RRID:AB_10603627
Goat anti-mouse secondary antibody	LI-COR Biotech	Cat# 926-68070
Goat anti-rabbit secondary antibody	LI-COR Biotech	Cat# 926-32211
Bacterial and virus strains		
BL21-CodonPlus (DE3)-RIPL Competent Cells	Agilent	Cat# 230280
BL21(DE3)	Agilent	Cat# 200131
Chemicals, peptides, and recombinant proteins		
Puromycin	Gibco	Cat# A1113803
TransIT Lenti Transfection Reagent	Mirus	Cat# #MIR6600
OPTI-MEM	Gibco	Cat# #11058021
Polybrene	Millipore	Cat# #TR-1003-G
Matrigel	Corning	Cat# 354234
RIPA Buffer	Thermo Scientific	Cat# 89901
Halt protease/phosphatase cocktail	Thermo Scientific	Cat# 1861281
Antibody diluent	LI-COR Biotech	Cat# 927-65001
T-PER Tissue Protein Extraction Reagent	Thermo Scientific	Cat# 78510
Ni-NTA Agarose	Qiagen	Cat# 30230
Glutathione agarose resin	GoldBio	Cat# G-250-50
Lipofectamine 2000	Thermo Scientific	Cat# 11668019
HaloTag Mammalian Pull-Down System	Promega	Cat# G6501
Dynabeads MyOne Streptavidin T1 Beads	Thermo Scientific	Cat#65001
Anti-FLAG® M2 Magnetic Beads	Sigma	Cat# M8823
AMPure XP beads	Beckman Coulter	Cat# A63881
Critical commercial assays		
RNeasy Plus Mini Kit	Qiagen	Cat# 74136
iScriptTM Reverse Transcription Supermix kit	Bio-Rad	Cat# 1708841
TaqMan™ Fast Advanced Master Mix	Thermo Scientific	Cat# 4444557
CellTiter-Glo Cell Proliferation Assay	Promega	Cat# G9241
Rapid Gold BCA Protein Assay Kit	Thermo Scientific	Cat# A55860
BCA Protein Assay Kit	Thermo Scientific	Cat# 23227
In-Fusion® Snap Assembly Master Mix	Takara	Cat# 638947
Micellula DNA Emulsion & Purification Kit	EurX	Cat# E3600
NanoBRET Nano-Glo Detection System	Promega	Cat# N1663
CUTANA™ ChIC/CUT&RUN Kit	Epicypher	Cat# 14-1048
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina	NEB	Cat# E7645

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Structure of POU2F3 POU domains bound to coactivator OCA-T2 and DNA (2.8 angstrom resolution)	This paper	PDB ID: 9PFN
Structure of POU2F3 POU domains bound to coactivator OCA-T2 and DNA (2.1 angstrom resolution)	This paper	PDB ID: 9PFO
Structure of POU2F3 POU domains bound to coactivator OCA-T1 and DNA	This paper	PDB ID: 9PFP
Uncropped western blot images	This paper	Mendeley Data: https://doi.org/10.17632/4fdmy5cx6s.1
RNA-seq	This paper	NCBI GEO: GSE302002
CUT&RUN-seq	This paper	NCBI GEO: GSE302003
Deep Mutational Scanning (DMS)	This paper	NCBI GEO: GSE302004
Experimental models: Cell lines		
NCI-H526	ATCC	Cat# CRL-5811
NCI-H211	ATCC	Cat# HTB-182
HEK293T	ATCC	Cat# #CRL-3216
Experimental models: Organisms/strains		
Female athymic nude mice	Envigo	N/A
Oligonucleotides		
shRNA sequences, see Table S2	This paper	N/A
Recombinant DNA		
pRSITPGP-U6Tet-sh-PGK-TetRep-2A-TagGFP2-2A-Puro	Cellecta, Inc	N/A
Lentiviral packaging mix	Cellecta, Inc	Cat# CPCP-K2A
pGEX-4T1	Cytiva	Cat# GE28-9545-49
pHIS-Parallel	Wu et al.[17]	N/A
Software and algorithms		
XDS	Kabsch, W. (2010).	N/A
Combat	N/A	N/A
POINTLESS	Evans, P. (2006)	N/A
AIMLESS	Evans, P.R., and Murshudov, G.N. (2013)	N/A
PHASER 2.8.3	McCoy, A.J. et al. (2007)	N/A
Coot 0.9.6	Emsley, P., and Cowtan, K. (2004).	N/A
PHENIX (versions 1.19.2, 1.20.1)	Liebschner, D. et al. (2019)	N/A
PyMol (versions 2.5.4 and 3.1.6.1)	N/A	https://pymol.org/2/
Bowtie2 (v2.3.4.2)	Langmead, B., and Salzberg, S.L. (2012)	N/A
SAMtools (v1.9)	Li, H. et al. (2009)	N/A
MACS2 (v2.2.x)	Zhang, Y. et al. (2008)	N/A
deepTools (v3.5.2)	Ramirez, F. et al. (2016)	N/A
The code for producing deep mutation scanning library sequences and analyses of screen results	This study	https://github.com/jackdesmarais/DMS_Designer
Other		
Doxycycline- supplemented diet	Teklad Global diet	Cat# TD.01306
Lysing Matrix D tubes	MP Bio	Cat# 6913500
Econo-Pac® 10DG Desalting Prepacked Gravity Flow Columns	Bio-Rad	Cat# 7322010
Swissci 3 Well Midi Crystallization Plate in UVP	Hampton Research	Cat# HR3-125

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Mice

All mouse experiments were performed with approval from and under the guidance of the Institutional Animal Care and Use Committee and all procedures were in accordance with the approved guidelines. Six-to eight-week-old female athymic nude mice were obtained from Envigo and housed in individually ventilated cages with food and water provided *ad libitum*. Animal rooms were maintained under a 12-h light-dark cycle, with temperature and humidity ranges maintained within the recommended guidelines. Mice were allowed to acclimatize for at least 72 h prior to study initiation.

Mammalian cell culture

NCI-H526 (ATCC #CRL-5811), NCI-H211(ATCC #HTB-182), and 293T (ATCC #CRL-3216) were purchased from ATCC. NCI-H526 and NCI-H211 were cultured in RPMI-1640 media (Gibco #61870036) and supplemented with 10% fetal bovine serum (FBS; Gibco #A5669801). 293T was cultured in DMEM (Gibco #10566016) supplemented with 10% FBS. All cell lines were maintained using the manufacturer's recommendations. shRNA-engineered cell lines were maintained in RPMI-160 media, supplemented with 10% FBS and 2 μg/mL puromycin (Gibco #A1113803). Puromycin-containing media was replaced every 2 to 3 days. NCI-H526 cell line was isolated from the lung carcinoma of a 55-year-old White male. NCI-H211 cell line was isolated from the lung carcinoma of a 50-year-old White female. 293T cells were derived from a female human embryonic kidney cells.

METHOD DETAILS

shRNA constructs for POU2F3 and OCA-T1 knockdown

shRNA sequences targeting POU2F3, OCA-T1, and controls (non-targeting and pan-lethal PSMA1) were cloned into pRSITPGP-U6Tet-sh-PGK-TetRep-2A-TagGFP2-2A-Puro, an inducible shRNA expression lentiviral vector (Cellecta, Inc.) Additional shRNA sequences for POU2F3 were curated from Project DRIVE and inserted into the same vector backbone. Individual shRNA sequences are listed in Table S2.

Lentivirus production of POU2F3 and OCA-T1 shRNA

For all vectors purchased from Cellecta, Inc, lentivirus was generated and titered by following Cellecta's standard lentivirus production protocol and a p24 ELISA assay kit. 293T cells were plated at a density of 2.5 million cells with 10 mL of media in 10 cm dishes 24 h prior to transfection. Per dish on the day of transfection, 2.4 μg of shRNA plasmid, 6 μg of packaging mix (Cellecta, Inc. #CPCP-K2A; contains plasmids that separately encode Gag/Pol and VSV-G) and 30 μL of TransIT Lenti Transfection Reagent (Mirus #MIR6600) were diluted in 500 μL OPTI-MEM (Gibco #11058021) and incubated at room temperature for 20 min before being added drop-wise to the dish. The media was replaced 20 h post transfection with 10 mL of fresh media. 48 h later, the media was harvested and clarified of cellular debris by centrifugation and 0.22 μm filtration (Millipore #SE1M179M6). Clarified virus was then aliquoted and stored at −80°C.

Stable cell line generation

Lentiviral infection of cell lines was performed by plating 500,000 cells per well in a 6-well plate and infecting at an MOI of 1 (1 viral particle per cell) in the presence of 8 μg/mL polybrene (Millipore #TR-1003-G). Cells were spinfected for 30 min at 1750 rpm at room temperature. Media was replaced 24 h later, and stable pools of cells were selected and maintained in 2 μg/mL puromycin.

Quantitative PCR

Total RNA was harvested using the RNeasy Plus Mini Kit (Qiagen #74136) and quantified using a Nanodrop. Per condition, 500 ng of total RNA was converted to cDNA using iScript Reverse Transcription Supermix kit (Bio-Rad #1708841). Gene expression was measured using the standard protocol for Thermo's TaqManTM Fast Advanced Master Mix (Thermo #4444557) and TaqMan gene expression assays (OCA-T1 #Hs01057842_m1, POU2F3 #Hs00205009_m1, GAPDH #Hs03929097_g1). GAPDH was used as an endogenous control. Prepped plates were measured on a Quant Studio.

Proliferation assays

CellTiter-Glo (CTG) proliferation assays (Promega #G9241) were performed in 96-well white, clear-bottom plates at a final concentration of 5000 cells per well in a final volume of 100 μL. To induce shRNA-mediated knockdown, half of the wells were spiked with doxycycline at a final concentration of 200 ng/mL. Fresh doxycycline was added every 2 to 3 days. One plate was seeded per time point. To measure a time point, 50 μL of CTG was added to each well and measured on an Envision instrument.

Mouse xenograft

Mice (6–8 weeks old) were inoculated subcutaneously in the right dorsal axillary region with 5 million NCI-H211 cells (in 50% Matrigel; Corning #354234) stably expressing one of the following Dox-inducible shRNAs: non-targeting shRNA (NT), POU2F-targeting shRNA (#1 or #3) or OCA-T1 targeting shRNA (#3 or #4). Tumor volumes and body weights were measured twice per week. Tumor volume

was determined by calipering in two dimensions and calculated as (width^2 × length × π/6). When the tumor volumes reached approximately 200 mm^3, mice with each cell line were randomly assigned to receive a doxycycline-supplemented diet (Teklad Global diet TD.01306, 625 ppm doxycycline) or the standard chow for up to 28 days. Four days after randomization, tumor samples were collected and flash-frozen from 3 mice in each group to assess early target knockdown. Samples were also collected from select groups 14–17 days after randomization and from all groups at the end of the efficacy study.

Western blotting using cultured cell and tumor lysates

Whole-cell lysates were prepared using RIPA buffer (Thermo #89901) and Halt protease/phosphatase cocktail (Thermo #1861281). Lysates were quantified using the Rapid Gold BCA Protein Assay Kit (Thermo #A55860). Western blots were carried out using standard methods. Antibodies for OCA-T1 (custom reagent from Cold Spring Harbor Laboratory) and POU2F3 (CST #92579S) were diluted 1:1000 and vinculin (Sigma #V9264) was diluted 1:2000. Secondary antibodies (LI-COR Biotech, goat anti-mouse #926–68070, goat anti-rabbit #926–32211) were diluted 1:25000. Primary antibodies were diluted in antibody diluent (LI-COR Biotech #927–65001) and incubated overnight at 4°C. Secondary antibodies were diluted in the same buffer and incubated on a benchtop rocker for 1 h. Blots were imaged using an LI-COR Odyssey Imaging System.

To prepare tumor lysates, tumor samples were homogenized in Lysing Matrix D tubes (MP Bio #6913500) in T-PER Tissue Protein Extraction Reagent (Thermo #78510) with Halt protease/phosphatase inhibitor cocktail (Thermo #1861281) using a Precellys Evolution homogenizer (Bertin Technologies). Protein content of tumor lysates was analyzed using a BCA Protein Assay Kit (Thermo #23227). Protein separation and immunodetection were performed using a Jess instrument (ProteinSimple by Bio-Techne). Primary antibodies were diluted as follows: OCA-T1 (Custom) 1:100, POU2F3 (CST #92579S) 1:100. Sample loading was normalized between all samples detecting the total protein amount in each sample using the Protein Simple kit #DM-TP01.

Expression constructs and protein purification

DNA encoding OCA-T1 (UniProt: Q8IXP5; residues 6–46), OCA-T2 (UniProt: A8K830-5; residues 1–38) and POU2F3 (UniProt: Q9UKI9-1; residues 183–341) were amplified using the Takara GXL polymerase kit. OCA-T1 and OCA-T2 were then cloned into the vector pGEX-4T1 with an N-terminal GST and TEV protease recognition site. The POU2F3 DNA binding domain was cloned downstream of a His_6-eGFP cassette and TEV protease site using the In-fusion cloning kit (Takara). Following sequence validation, *E. coli* BL21-CodonPlus (DE3)-RIPL Competent Cells (Agilent, Cat# 230280) were transformed with each plasmid separately. Transformants were selected and cultured in LB medium (with appropriate antibiotics) at 37°C with shaking until the desired culture density was reached. Cultures for OCA-T1 and OCA-T2 expression were induced at $OD_{\lambda=600}$ 1.0 with 1 mM IPTG at 30°C for 3 h. POU2F3 expression was induced at $OD_{\lambda=600}$ 1.0 with 0.5 mM IPTG at 16°C overnight. Bacterial cultures were harvested by centrifuging at 4600*g* for 15 min at 4°C and pellets were stored at −20°C until lysis.

Frozen pellets were thawed and lysed in extraction buffers (for OCA-T1 and OCA-T2: 50 mM HEPES, pH 8.0, 500 mM KCl, 100 mM potassium acetate, 1 mM EDTA, 20% glycerol, 1% NP-40, 1 mM DTT, 0.2 mM PMSF, and protease inhibitor cocktail; for POU2F3, 50 mM sodium phosphate, pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM DTT, 0.2 mM PMSF, and protease inhibitor cocktail). Resuspended lysates were supplemented with 100 μg/mL lysozyme and sonicated. The crude lysate was clarified by centrifugation using Beckman Type 45-Ti rotor at 40 krpm 4°C for 1 h.

The OCA-T1 and OCA-T2 proteins were affinity purified using glutathione agarose beads (Goldbio) whereas POU2F3 was purified using Ni-NTA beads (Qiagen). Bound beads were washed followed by protein elution using 20 mM glutathione in 50 mM Tris buffer pH 8.0 (OCA-T1 and OCA-T2) or 50 mM sodium phosphate, 100 mM NaCl, 250 mM imidazole pH 8.0 (POU2F3). The affinity-purified proteins were desalted using Econo-Pac 10DG Desalting Prepacked Gravity Flow Columns into PBS, pH 7.2, (for OCA-T1/2) or 50 mM sodium phosphate, pH 8.0, with 200 mM NaCl (POU2F3) then incubated with TEV protease (1:50 mass ratio) overnight at 4°C.

The cleaved tag and any remaining undigested material were removed through an additional application to affinity resin, this time collecting the flowthrough. The protein solutions were then concentrated using Amicon centrifugal filters (3 kDa cutoff for OCA-T1 and OCA-T2; 10 kDa cutoff for POU2F3). After concentration, the proteins were further purified by size exclusion chromatography using a Superdex 75 Increase 10/300 GL column with PBS as the running buffer. Purity of the fractions was assessed by SDS-PAGE, and peak fractions were pooled and concentrated.

Preparation of DNA duplexes

A 15-mer oligonucleotide bearing the octamer site was ordered from IDT with standard desalting and sodium salt exchange. Each strand was resuspended in 10 mM Tris, pH 8.0, 50 mM NaCl, then the two strands mixed in an equimolar ratio. Duplexes were formed by heating the mixture at 95°C for 5 min then gradually cooling to 20°C at a rate of 5°C/min using a thermocycler.

Reconstitution of the ternary complex

Purified proteins and freshly prepared DNA were mixed at 1:1.2:1.5 molar ratio (POU2F3:DNA:OCA-T1/2) in binding buffer (5 mM HEPES, pH 7.5, 25 mM ammonium acetate, 1 mM DTT). The reconstituted complex was fractionated by gel filtration using a Superdex 200 Increase 10/300 GL column. Peak fractions were pooled and concentrated (10 kDa cutoff) to a final concentration of 200–400 μM.

Crystallization conditions

Crystallization screens were set up using 96-well sitting drop trays [Swissci 3 Well Midi Crystallization Plate in UVP (Hampton Research #HR3-125)]. 200 nL of ternary complex at 200–400 μM were mixed with an equal volume of reservoir solution. Trays were incubated at 18°C and routinely monitored. Rod-shaped or planar crystals appeared within one week in several conditions. Crystals from conditions with 0.1 M MIB pH 4 (2:3:3 molar ratio of malonate: imidazole: boric acid) plus 25% PEG-1500 for OCA-T1; or 0.1 M MMT (1:2:2 molar ratio of DL-malic acid:MES:Tris base) pH 6 plus 25% PEG-1500 for OCA-T2 structure were harvested.

Cryoprotection conditions

For the OCA-T1 structure and the low-resolution OCA-T2 structures, crystals were harvested and sequentially soaked in cryoprotectant solutions consisting of 0.1 M MIB (2:3:3 molar ratio of malonate:imidazole:boric acid), pH 4.0, 30% PEG-1500 and then 0.1 M MIB (2:3:3 molar ratio of malonate:imidazole:boric acid), pH 4.0, 35% PEG-1500 for 15 s each. Crystals were flash-frozen in liquid nitrogen.

For the OCA-T2 structure, crystals were harvested and sequentially incubated in cryoprotectant solutions consisting of 0.1 M MIB (2:3:3 molar ratio of malonate:imidazole:boric acid), pH 6.0, 30% PEG-1500 and then 0.1 M MIB (2:3:3 molar ratio of malonate:imidazole:boric acid), pH 6.0, 35% PEG-1500 for 15 s each. Crystals were flash-frozen liquid nitrogen.

Data collection and processing

X-ray diffraction data were collected at AMX beamline (17-ID-1) at the National Synchrotron Light Source II (NSLS-II) at Brookhaven National Laboratory (BNL), Upton, NY or the NE-CAT beamline (24-ID-E) at the Advanced Photon Source (APS) at Argonne National Laboratory (ANL), Lemont, IL. Indexing and integration were performed with XDS.[31] Additional processing was performed with programs from the CCP4 suite including COMBAT, POINTLESS,[32] and AIMLESS.[33]

Phasing, model building, refinement, and validation

All structures were determined by molecular replacement using the program Phaser (version 2.8.3).[34] The OCA-T2 (residues 1–38) + POU2F3 (DBD) + DNA [low resolution] structure used the OCA-B/DNA/OCT-1 complex (PDB: 1CQT) as a search model. OCA-T2 (residues 1–38) + POU2F3 (DBD) + DNA [high resolution] used the OCA-T2 (residues 1–38) + POU2F3 (DBD) + DNA [low resolution] structure as the search model. Finally, the OCA-T1 (residues 6–46) + POU2F3 (DBD) + DNA used the OCA-T2 (residues 1–38) + POU2F3 (DBD) + DNA [high resolution] structure as the search model.

Iterative model building was performed with Coot 0.9.6.[35] Refinement and validation were conducted using PHENIX.[36] Structures were validated using Molprobity (http://molprobity.biochem.duke.edu/).[37] Molecular graphics were generated using The PyMOL Molecular Graphics System (versions 2.5.4 and 3.1.6.1) (https://pymol.org/2/). Protein–DNA interaction diagrams were generated using NUCPLOT v.1.1.4 (https://www.ebi.ac.uk/thornton-srv/software/NUCPLOT/).[38] Data collection and refinement statistics are listed in Table S1.

Data deposition

Structures and data for the OCA-T2 (residues 1–38) + POU2F3 (DBD) + DNA [low resolution] complex, the OCA-T2 (residues 1–38) + POU2F3 (DBD) + DNA [high resolution] complex, and the OCA-T1 (residues 6–46) + POU2F3 (DBD) + DNA complex were deposited in the Protein Data Bank with accession codes PDB: 9PFN, PDB: 9PFO, and PDB: 9PFP, respectively.

Deep mutational scanning library design and cloning

Deep mutational scanning (DMS) libraries were designed to cover all single amino acid changes and stop codons within targeted regions of the OCA-T1 and POU2F3 coding sequences. For OCA-T1, the mutagenized region spanned amino acid positions 2–52, while for POU2F3, positions 183–353. The POU2F3 region was covered by three partially overlapping libraries: Library 1 (amino acids 183–237), Library 2 (amino acids 233–302), and Library 3 (amino acids 299–353), with 4–5 amino acids overlappoing between adjacent libraries. Five replicate library members were created for each single amino acid change, two library members were created for each stop codon, and between 665 and 1000 library members were created for the wild-type amino acid sequence for each library. A nucleotide sequence was generated for each library member. Each amino acid position was converted into a missense codon by randomly selecting one of the two most frequently used codons for that amino acid. Duplicate nucleotide sequences were avoided by repeating the generation process for any sequence that matched an already existing sequence.

In summary, OCA-T1 library includes 5988 cDNAs (800 WT cDNAs, 153 stop codons, and 5035 missense cDNA variant members), POU2F3 Library 1 include 6000 cDNAs (665 WT cDNAs, 110 stop codons, and 5225 missense cDNA variant members), POU2F3 Library 2 include 7790 cDNAs (1000 WT cDNAs, 140 stop codons, and 6650 missense cDNA variant members) and POU2F3 Library 3 include 6000 cDNAs (665 WT cDNAs, 110 stop codons, and 5225 missense cDNA variant members).

Deep mutational scanning library amplification

All four DMS libraries were synthesized as cDNA oligo pools (TWIST Bioscience) and amplified by emulsion-based PCR using the Micellula DNA Emulsion & Purification Kit (EurX Cat. No. E3600) according to the manufacturer's instructions. Emulsion conditions were optimized to define single-molecule amplification in individual compartments (micelle). Based on optimal conditions, emulsions

were prepared at ~10^{10} compartments per milliliter, along with ~10^9 oligos (~1 ng) from each oligo pool and multiple individual 100 μL PCRs were performed for each library.

Amplified libraries were cloned into lentiviral backbone expression vectors using Gibson Assembly (NEB Gibson Assembly Master Mix) to generate CRISPR-resistant OCA-T1 or POU2F3 cDNA libraries. Each construct encoded a 3×Flag-tagged version of OCA-T1 or POU2F3, under the EFS promoter. In the OCA-T1 vector, the cDNA was followed by an IRES-GFP reporter, while in the POU2F3 vector, by an IRES-mCherry reporter.

Each vector also co-expressed an sgRNA targeting the corresponding endogenous gene, under the human U6 promoter. The OCA-T1 vector encoded the sgOCA-T1 (TGGCAAGGTCAGCAGGGCAT), and the POU2F3 vector included sgPOU2F3 (GACCAACATCCGCCTGACTC). These sgRNAs enabled simultaneous disruption of the endogenous gene and functional interrogation of CRISPR-resistant cDNA variants within the DMS libraries.

Cloned libraries were amplified to maintain ~300-fold representation and used to transfect HEK 293T cells for lentiviral library production.

Deep mutational scanning library screening

NCI-H211 cells were first transduced with LentiV_Cas9_puro vector (Addgene #108100) to stably express Cas9. Cas9$^+$ cells were subsequently transduced with either the OCA-T1 or POU2F3 DMS libraries, targeting a representation of 2,000 cells per variant. Transductions were performed at a low multiplicity of infection (MOI ~0.1–0.15) to ensure single-copy cDNA integration per cell.

Briefly, NCI-H211 Cas9$^+$ cells were transduced by seed-infection, and transduction efficiency was assessed on day 3 post-infection by measuring GFP$^+$ (OCA-T1) or mCherry$^+$ (POU2F3) expression. Two biological replicates were performed for each library. Cells were collected on day 3 (initial time point) and then passaged every 3 days, maintaining the target library representation for 10 population doublings. At the final time point, cells were harvested for genomic DNA extraction.

Deep mutational scanning genomic DNA preparation

Cells harvested at the initial and final time points of each DMS screen were washed with PBS and lysed in DNA extraction buffer at a density of 10^7 cells/mL. The lysis buffer consisted of 10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM EDTA, 0.1% SDS, and 0.2 mg/mL Proteinase K. Samples were incubated at 56°C for 48 h. Following lysis, genomic DNA was purified using Tris-saturated phenol (Thermo Fisher Scientific) and precipitated by adding three volumes of 100% ethanol and sodium acetate, pH 5.2, to a final concentration of 75 mM. Precipitation was carried out overnight at −20°C. Samples were centrifuged at 20,000*g* for 30 min at 4°C, washed with 70% ethanol, and air-dried. DNA pellets were resuspended in ultrapure water.

Deep mutational scanning illumina-compatible library preparation and sequencing

Illumina-compatible sequencing libraries were prepared from genomic DNA samples using a two-step PCR amplification strategy. In the first PCR step, the pooled cDNA variant population was amplified using primers flanking the mutagenized window of each DMS library. These primers also carried universal adapter sequences at their 5′ ends. In the second PCR step, amplified products were barcoded using custom Illumina-compatible P5/P7 stacked barcoded primers (Integrated DNA Technologies). These primers annealed to the universal adapter sequences introduced during the first PCR step.

All PCR reactions for both amplification steps were performed using AmpliTaq Gold DNA polymerase (Thermo Fisher Scientific), following the manufacturer's instructions. For the first-round PCR, approximately 800 individual reactions were set up per sample to ensure sufficient representation of the cDNA variant population. Each reaction was performed using 1 μg of genomic DNA. Cycling conditions were as follows: 10 min at 95°C, followed by 20 cycles of 30 s at 95°C, 45 s at 62°C, and 30 s at 72°C, with a final extension for 7 min at 72°C.

The first-round PCR primers used for each Deep Mutational Scanning (DMS) library were:

OCA-T1 DMS Library.

Forward: GTGGAAAGGACGAAACACCCGCGGGATCCGCCACCATG.

Reverse: AGTTCTTGAGAGGCCATGTTTATCGGTGACCGGCGAACCTGGCAT.

POU2F3 DMS Library 1.

Forward: GTGGAAAGGACGAAACACCCCAGCTCTGGAGGGGCC.

Reverse: AGTTCTTGAGAGGCCATGTTTATCGTGCACATGTTCTTGAAGCT.

POU2F3 DMS Library 2.

Forward: GTGGAAAGGACGAAACACCGACCACCATCTCACGATTT.

Reverse: AGTTCTTGAGAGGCCATGTTTATCGGAGCTGGGTTTTGGGTT.

POU2F3 DMS Library 3.

Forward: GTGGAAAGGACGAAACACCCGCCTCACCCTCGAGAAG.

Reverse: AGTTCTTGAGAGGCCATGTTTATCGGGAGATACCAGCCGGGA.

Each forward primer included a 5′ universal adapter sequence (GTGGAAAGGACGAAACACC), and each reverse primer included the 5′ universal adapter sequence (AGTTCTTGAGAGGCCATGTTTATCG).

A single PCR fragment (200–300 bp) was generated from each sample, gel purified from a 2% agarose gel and used as input for the second-round PCR. For all libraries, 40 parallel PCR reactions were performed per sample using 10 ng of purified first-round PCR

product per reaction. The cycling conditions were: 10 min at 95°C, followed by 6 cycles of 30 s at 95°C, 45 s at 62°C, and 30 s at 72°C, with a final extension for 7 min at 72°C. The following barcoded P5/P7 primers were used for the second-round PCR:

P5 Forward primer:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNHNNNN(Barcode)GTGGAAAGGACGAAACACC.

P7 Reverse primer:

CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGTTCTTGAGAGGCCATGTTTATCG.

The final libraries (~400 bp) were gel-purified and quantified using the Qubit dsDNA BR Assay (Thermo Fisher Scientific, Q32850). Libraries were diluted, pooled and sequenced on Illumina NextSeq 2000 platforms. The OCA-T1 pooled libraries were sequenced using a P3 flow cell in paired-end 150 bp (PE150) mode. The POU2F3 libraries were sequenced using the P2 flow cell in single-end 300 bp (SE300) mode.

Deep mutational scanning data analysis and visualization

An enrichment score was calculated for each member in each DMS library. First, the number of reads associated with each library member was converted to an abundance by dividing library member counts by the total number of reads for each time point of the screen. Then, the fold change in abundance during selection was determined by dividing the abundance at the last timepoint of DMS screen by the abundance at the initial time point. Finally, a $\log_2$ transformation was applied to convert fold changes to relative fitness. The Deep Mutational Scanning data can be found in Table S3.

Correspondence between library replicates was assessed by averaging fitness across library members corresponding to the same variant to produce a variant effect score. Replicate library variant effect scores were compared to assess both correlation and precision of scores across replicates. As the replicates from the second library had scores that were highly correlated but were at different scales, a linear correction was applied to bring the scores to the same scale. This correction was produced by fitting a linear regression of one replicate's variant effect score against the other replicate's variant effect score. The regression equation was then used to scale one replicate's relative fitnesses to match the scale of the other replicate's relative fitnesses. Combined variant effect scores were then computed by averaging relative fitnesses across library replicates before averaging library members that corresponded to the same variant.

Partially overlapping POU2F3 libraries were merged for joint analysis by matching the variant effect scales in the libraries. For merging the second and third libraries, variants from the overlapping section shared by the libraries were used for scale-matching. A linear regression was fit for one library's combined variant effect score against the other library's combined variant effect score. The regression equation was then used to scale one library's relative fitnesses to match the scale of the other library's relative fitnesses. This method could only be applied to merge library three into library two however, because the overlapping section shared by library one and library two was intolerant to mutations and so did not provide enough examples of variants at the higher end of the variant effect distribution to effectively learn the scale conversion. Instead, library one and library two were merged by fitting a linear regression to their positive and negative controls. These controls included synonymous mutations as positive controls and stop codons as negative controls. The regression equation was then used to scale one library's relative fitnesses to match the scale of the other library's relative fitnesses so that the libraries could be directly compared. Combined variant effect scores were then re-computed by averaging abundances across library replicates before averaging library members that corresponded to the same variant.

Variants were classified as tolerated if they had combined variant effect scores greater than a threshold (−0.5 for POU2F3 and OCA-T1) or deleterious if their combined variant effect scores were equal to or below the threshold. The mutation tolerability score for each position was computed as the fraction of variants at each position classified as tolerated.

The code required to produce library sequences and to analyze library results can be found at [https://github.com/jackdesmarais/DMS_Designer] and a version of record is saved at [https://zenodo.org/records/15800486 DOI: https://doi.org/10.5281/zenodo.15800486].

NanoBRET assay

Point mutation library plasmids were generated in Promega pFC14K (C-terminal HaloTag) for POU2F3 and in Promega pFN31K (N-terminal NanoLuc) for OCA-T1. Transfection complexes were formed at the acceptor/donor (A/D) ratio of 10:1 in OptiMEM: per 100,000 cells, 25 ng of POU2F3-HaloTag, 2.5 ng of NanoLuc-OCA-T1, and 122.5 ng of pGEM-3Z carrier DNA were (150 ng of DNA in total). Lipofectamine 2000 (Thermo #11668019) was added at 3:1 (0.45 μL) before incubating 20 min at room temperature. HEK293 cells were harvested in trypsin and resuspended at 500,000 cells/mL in OptiMEM +10% FBS. HaloTag-618 ligand was added from 1000-fold concentrated stock to final concentration 100 nM. The DNA:lipofectamine complex was mixed with cells and the resultant mixture was plated in white, solid-bottom, 96-well plates to final volume 100 μL/well with 50,000 cells/well. After 24 h, the NanoBRET readout was measured with a NanoBRET Nano-Glo detection system (Promega #N1663) according to the manufacturer's protocol. Donor and acceptor emission were measured and the BRET ratio reported in mBRET. mBRET ratios were normalized to POU2F3:OCA-T1 WT:WT pair. NanoBRET data can be found in Table S4.

HaloTag pull-down

Reverse transfection of HEK293T cells was performed by complexing 1.25 μg of POU2F3-HaloTag (in pFC14K) and 1.25 μg NanoLuc-OCA-T1 (in pFN31K) with 7.5 μL Lipofectamine 2000 in OptiMEM for 20 min at room temperature. One million HEK293T cells were mixed with the transfection complex and plated in 10 cm culture dish in a final volume of 10 mL OptiMEM. After 24 h, cells were collected by scraping, washed once with PBS, and flash frozen. Samples were processed via the HaloTag Mammalian Pull-Down System (Promega #G6501) according to the manufacturer's protocol. Prior to pull-down, the salt concentration of the lysates was adjusted to 60 mM. The captured complex was eluted in SDS elution buffer (provided) and analyzed via western blot using standard methods. Antibodies for OCA-T1 (developed at Cold Spring Harbor) and POU2F3 (CST #92579S) were diluted 1:1000.

POU2F3 cDNA rescue assay

POU2F3 wild-type (WT) and various mutant cDNA constructs, including missense mutations (L188A, E189A, L191A, E192A) and a nonsense mutation at L191, were cloned into a lentiviral FLAG-tagged expression vector containing a GFP fluorescent reporter and a blasticidin resistance cassette, separated by a P2A self-cleaving peptide sequence. All *POU2F3* cDNA constructs were engineered to be CRISPR-resistant versions of the endogenous gene through the introduction of silent mutations within a target site recognized by an sgRNA (sgPOU2F3). Lentiviral supernatants were generated from each construct, including an empty vector control**,** and used to transduce H211 cells stably expressing Cas9**.** Transduced cells were selected with blasticidin for 7 days, and expression of each construct was validated by western blotting using an anti-FLAG antibody and by fluorescence microscopy to confirm GFP expression. Validated cDNA-expressing H211 cells were subsequently transduced with lentiviral sgRNAs targeting either *POU2F3* (sgPOU2F3) or the *Rosa26* safe harbor locus (sgROSA control), achieving an initial GFP-positivity rate of ∼30–40% at day 3 post-infection. The percentage of GFP-positive cells was measured every three days over a 21-day period using a Guava easyCyte flow cytometer. To evaluate the functional rescue efficiency for each construct, the percentage of GFP-positive cells at each time point was normalized to the day 3 value for each construct. These normalized values were then compared to those of the WT *POU2F3* cDNA across all timepoints.

Purification of the POU2F3 L191A mutant protein

E. coli BL21(DE3) cells (Agilent #200131) were transformed with plasmids encoding the respective proteins (His_6-GFP, His_6-GFP-$POU2F3^{WT}$, His_6-GFP-$POU2F3^{L191A}$). Transformants were selected using ampicillin and cultured in LB medium. When the cultures reached an $OD_{\lambda=600}$ of 0.8–1.0, protein expression was induced with 1 mM IPTG. The cultures were then incubated for 16–18 h at 18°C. Bacterial cells were harvested by centrifugation at 6000*g* for 10 min at 4°C, and the pellets were stored at −20°C until lysis.

Frozen cell pellets were thawed and resuspended in lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 20% glycerol, 10 mM imidazole, pH 8), supplemented with a protease inhibitor cocktail, 0.7 μL/mL β-mercaptoethanol, and 0.2 mM PMSF. Cells were lysed by sonication then clarified by centrifugation at 16,000*g* for 40 min at 4°C. The soluble supernatant was incubated with Ni-NTA resin (pre-equilibrated with lysis buffer) for 2 h at 4°C. The beads were then washed three times with wash buffer (50 mM Na_2HPO_4, 300 mM NaCl, 20% glycerol, 20 mM imidazole, pH 8.0) supplemented with protease inhibitor cocktail, 0.7 μL/mL β-mercaptoethanol, and 0.2 mM PMSF. Proteins were eluted by incubating the washed beads in elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 20% glycerol, 250 mM imidazole, pH 8.0) supplemented with protease inhibitor cocktail and 0.2 mM PMSF, for 30 min at 4°C. The eluted proteins were supplemented with 1 mM dithiothreitol and 1 mM EDTA, then concentrated using Amicon Ultra centrifugal filter units with appropriate molecular weight cutoffs. The concentrated proteins were further purified by gel filtration using a Superdex 200 Increase 10/300 GL column (Cytiva/GE Healthcare Life Sciences). The purity of the peak fractions was assessed by SDS-PAGE. Fractions with highly purified proteins were pooled, concentrated, and stored for binding assays.

DNA binding assay (size-exclusion chromatography)

Purified proteins (2 μM; His_6-GFP, His_6-GFP-$POU2F3^{WT}$, His_6-GFP-$POU2F3^{L191A}$) were mixed with octameric DNA duplexes at a 1.2:1 molar ratio and incubated at room temperature for 5 min in 250 μL of binding buffer (10 mM HEPES, pH 7.5, 25 mM ammonium acetate, 1 mM DTT). The binding mixture was injected into a Superdex 200 increase 10/300 GL column (Cytiva/GE Healthcare Life Sciences) and separated using size exclusion buffer at a flow rate of 0.6 mL/min. Absorbance at 260 nm was monitored to assess DNA binding of wild-type and mutant POU2F3 proteins.

DNA pull-down assay

Single-stranded DNA oligos were synthesized by Integrated DNA Technologies (biotinylated forward strand and non-biotinylated reverse strand). Double-stranded DNA probes were generated by mixing complementary oligonucleotides in a 1:1 ratio, followed by heating at 95°C for 5 min and then annealing by cooling to 25°C at a rate of 5°C/min in a thermocycler. For each reaction, 20 nM double-stranded DNA probes were incubated in the presence of 20 nM respective proteins.

Twenty-five microliters of Dynabeads MyOne Streptavidin T1 (Thermo Fisher Scientific #65001) were washed twice with wash buffer (20 mM Tris, pH 7.5, 1 M NaCl, 1 mM EDTA), followed by a PBS wash. The beads were then washed once with NETN 150 buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, and protease inhibitor cocktail). The beads were incubated with DNA oligos in NETN 150 buffer for 1 h at 4°C.

The DNA-immobilized beads were washed once with wash buffer, followed by two washes with NETN 150 buffer. The beads were then incubated with respective proteins (His_6-GFP, His_6-GFP-POU2F3^{WT}, His_6-GFP-POU2F3^{L191A}) in NETN 150 buffer for 2 h at 4°C. After incubation, the beads were washed three times with NETN 150 buffer, followed by a final PBS wash. Finally, the bound proteins were eluted by resuspending the beads in 50 μL Laemmli sample buffer (Bio-Rad #1610737) and heating at 95°C for 8–10 min.

Co-immunoprecipitation assay

NCI-H526 cells stably expressing wild-type or α1-helix mutants of POU2F3 (3×FLAG-tagged) were grown in 15 cm dishes. Approximately 15 million cells were harvested and washed with ice-cold PBS. Cells were lysed by incubating them in 1 mL of buffer A (10 mM HEPES, pH 7.9, 5 mM magnesium acetate, 1 mM $MgCl_2$, 0.3 M sucrose, 0.1% NP-40, protease inhibitor cocktail, and 0.2 mM PMSF) on ice for 15 min.

Nuclei were pelleted by centrifugation at 9800*g* for 10 min at 4°C. The supernatant was discarded, and the nuclear pellets were resuspended in 1 mL of co-IP buffer (50 mM Tris, pH 7.5, 80 mM NaCl, 0.5% NP-40, 10% glycerol, 1.5 mM $MgCl_2$, and protease inhibitors). The suspension was incubated at 4°C for 1 h. Extracts were then centrifuged at 16000*g* for 30 min at 4°C. The resulting supernatant (nuclear extract) was transferred to a fresh tube. Five percent of the total extract was saved as an input sample.

In parallel, 25 μL of Anti-FLAG M2 Magnetic Beads (Sigma #M8823) were washed with Co-IP buffer. The washed beads were then added to the nuclear extracts and incubated at 4°C for 12–14 h. After incubation, the beads were washed three times with Co-IP buffer and resuspended in 50 μL of Laemmli sample buffer (Bio-Rad #1610737). Samples were heated at 95°C for 8–10 min and analyzed by western blotting.

RNA extraction and library construction

Gene complementation RNA-seq experiments were conducted in NCI-H211 cells stably expressing Cas9. CRISPR-resistant cDNAs encoding either wild-type POU2F3 (WT) or a leucine-to-alanine missense mutation at position 191 (L191A) were cloned into an FLAG-tagged lentiviral expression vector. Both POU2F3 cDNAs contained silent mutations within the sgRNA recognition site targeted by sgPOU2F3. The lentiviral vector also co-expressed a GFP fluorescent reporter and a blasticidin resistance gene, separated by a P2A self-cleaving peptide. Stable Cas9$^+$ NCI-H211 cell lines were generated expressing either the WT, L191A, or an empty vector control.

Each stable line was subsequently infected in triplicate with a lentivirus expressing either a non-targeting control sgRNA (sgROSA) or sgPOU2F3. These sgRNAs were cloned into a modified LRG2.1 vector (Addgene #125594) in which GFP was replaced with an mCherry fluorescent reporter and a neomycin resistance cassette, separated by a P2A sequence. All infections with sgRNA vectors resulted in >95% mCherry-positive cells. Samples were collected five days post-infection for RNA-seq analysis. Three independent biological replicates were processed for each condition.

Total RNA was extracted using TRIzol reagent (Thermo Fisher Scientific) following the manufacturer's instructions. Cells were lysed in 1 mL of TRIzol, and RNA was resuspended in RNase-free water. RNA-seq libraries were prepared using 2 μg of total RNA using the TruSeq Sample Prep Kit v2 (Illumina), according to the standard protocol. Briefly, RNA samples underwent poly-A enrichment, fragmentation, and cDNA synthesis, followed by end repair, A-tailing, adapter ligation with single-end indices, and library amplification. Barcoded libraries were pooled and sequenced using single-end 76 bp reads on an Illumina NextSeq platform.

RNA-seq data analysis

Single-end 76 bp raw sequencing reads were pseudoaligned to the human transcriptome (hg38) using Kallisto with 100 bootstrap iterations. Genes with low expression (maximum read count below 20 across all conditions) were filtered out. Structural RNAs and non-coding RNAs were excluded from downstream analyses.

Counts from all experimental groups (POU2F3 WT**,** POU2F3 L191A, and empty vector control) were normalized and used to calculate log_2 fold-change (LFC) values. These LFC values were further transformed and subjected to *Z* score calculation. All conditions included three independent biological replicates. Heatmaps were generated using the z-scores of normalized LFC values to visualize the top 1000 differentially expressed genes, employing the ComplexHeatmap R package.

CUT&RUN analysis

NCI-H211 cells stably expressing wild-type POU2F3 or α1-helix mutant POU2F3^{L191A} (3×FLAG-tagged) were cultured in 15 cm dishes. CUT&RUN was performed using the CUTANA ChIC/CUT&RUN Kit (EpiCypher #14–1048) according to the manufacturer's instructions. Briefly, 500,000 cells per reaction were harvested, washed, and bound to Concanavalin A (ConA)-coated magnetic beads that had been pre-activated with bead activation buffer. Following bead binding, cells were permeabilized using digitonin-containing antibody buffer and incubated overnight (12–14 h) at 4°C with 0.1 μg of primary antibody per reaction—either anti-FLAG (EpiCypher #13–2031) or IgG control (EpiCypher #13–0042).

The next day, beads were washed and incubated with 2.5 μL pAG-MNase per reaction for 10 min at room temperature. Targeted chromatin digestion was initiated by the addition of 1 μL 100 mM $CaCl_2$ and carried out for 2 h at 4°C. Chromatin fragments were released by incubating samples at 37°C for 10 min following the addition of stop buffer supplemented with 1 μL *E. coli* spike-in DNA. DNA was purified using SPRI beads, ethanol-washed, and eluted in 0.1× TE buffer. The purified DNA was stored at −20°C until library preparation.

CUT&RUN DNA libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs #E7645) following the manufacturer's instructions. Briefly, the CUT&RUN DNA was subjected to end repair and A-tailing, followed by adaptor ligation using NEBNext adaptors and ligation reagents. USER Enzyme was added to complete the ligation reaction. Adaptor-ligated DNA was purified using AMPure XP beads (0.9× bead ratio), washed twice with 80% ethanol, air-dried, and eluted in 0.1× TE buffer. Library amplification was performed by PCR using NEBNext Ultra II Q5 Master Mix, a universal i5 primer, and sample-specific i7 index primers (NEBNext Oligos for Illumina). Libraries were amplified using 14 PCR cycles, followed by purification with a 0.9× AMPure XP bead cleanup and elution in 0.1× TE buffer. Final library concentration was measured using a Qubit fluorometer, and fragment size distribution was assessed using an Agilent Bioanalyzer with the High Sensitivity DNA kit (Agilent #5067-4626). Libraries were paired-end sequenced (2 × 75 bp) using the Illumina NextSeq platform with a Mid Output kit.

Sequencing reads were first aligned to the *E. coli* K-12 reference genome using Bowtie2 (v2.3.4.2) to remove spike-in control reads.[39] Unaligned read pairs were then mapped to the human genome (hg38) using Bowtie2 with default parameters. Aligned SAM files were converted to BAM format, sorted, and indexed using SAMtools (v1.9) for downstream analysis.[40] Peak calling was performed with MACS2 (v2.2.x) using the callpeak command on paired-end, sorted BAM files.[41] For each sample (FLAG-EV, FLAG-POU2F3^{WT}, and FLAG-POU2F3^{L191A}), the corresponding IgG control was used as background (-c), and peaks were called with a q-value threshold of 0.01. Peak files were filtered to remove mitochondrial reads (chrM), sorted by genomic coordinates, and merged using bedtools merge to collapse overlapping peaks while retaining the maximum peak score.[42] Merged peak sets were further filtered to exclude ENCODE blacklist regions and non-standard chromosomes to improve peak quality. Overlap analysis between POU2F3^{WT} and POU2F3^{L191A} peak sets was conducted using bedtools intersect to identify shared, WT-specific, and L191A-specific peaks.

To visualize genomic signal, bigWig files were generated from sorted and indexed BAM files using the bamCoverage function of deepTools (v3.5.2),[43] with blacklisted regions excluded to avoid signal artifacts. Average chromatin occupancy metaplots were generated using the computeMatrix and plotProfile functions of deepTools. Signal matrices were computed in reference-point mode, centered on peak summits and extended ±5 kbp, using bigWig and BED files corresponding to POU2F3 peak regions. The resulting matrices were visualized as line plots to compare CUT&RUN signal enrichment between POU2F3^{WT} and POU2F3^{L191A} samples.

To compare FLAG-POU2F3 CUT&RUN data to endogenous POU2F3 occupancy, ChIP-seq data were obtained from GEO accession number GEO: GSE186614 (samples GSM5657746: NCIH211_POU2F3_ChIP; GSM5657747: NCIH211_POU2F3_Input; corresponding raw data SRR16578472 and SRR16578473). Motif enrichment analysis was performed using the HOMER suite on peaks derived from FLAG-POU2F3^{WT} CUT&RUN and endogenous POU2F3 ChIP-seq experiments after excluding ENCODE blacklist regions.[44] Motifs were identified using the findMotifsGenome.pl command with a 300 bp window centered on each peak summit and the hg38 reference genome with repeat masking enabled. Enrichment was calculated relative to background sequences automatically generated by HOMER and the most significantly enriched motifs were ranked by their *p*-value.

Access to DNA sequencing data

RNA-seq data has been deposited at NCBI GEO with accession GSE302002. CUT & RUN data have been deposited at NCBI GEO with accession code GEO: GSE302003. Deep mutational scan data have been deposited at NCBI GEO with accession code GEO: GSE302004.

QUANTIFICATION AND STATISTICAL ANALYSIS

In Figure 1, quantification details for tumor size measurements were explained in respective STAR Methods section. Tumor volume data was displayed as the mean ± SEM and the number of animals in each treatment group are displayed in each figure legend. Quantification details for the expression analysis by RT-qPCR and Western blotting were explained in the respective STAR Methods section. Data collection and refinement statistics for the crystallography data (represented in Figures 2, 3, and 4, S4 and S5) are listed in Table S1. Quantifications associated with the DMS-related figures (Figures 5 and 6) are explained in respective STAR Methods section and raw data are provided in Table S3. Bioinformatics-associated analyses in Figures 7, S9 and S10 were performed using the tools, algorithms, and software indicated in the respective STAR Methods section. Threshold values and other statistical parameters were indicated in the respective STAR Methods sections.