Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DACH1 expression is inversely associated with promoter methylation of SLIT2 and ESR1 in TCGA cervical squamous carcinoma

Jose Bird

jbird@birdaisolutions.com

Independent Researcher https://orcid.org/0009-0006-2744-0606

Research Article

Keywords: DACH1, DNA methylation, cervical cancer, master regulator analysis, ARACNe, CADM1, ESR1, SLIT2

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10386608/v3

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare potential competing interests as follows: The author is the founder of Bird AI Solutions. This work was conducted independently and received no funding.

DACH1 expression is inversely associated with promoter methylation of SLIT2 and ESR1 in TCGA cervical squamous carcinoma

Short Communication

Jose A. Bird, PhD
Independent Researcher
ORCID: 0009-0006-2744-0606 · Corresponding author: jbird@birdaisolutions.com

Abstract

Background: Promoter methylation silences tumor-suppressor and adhesion genes in squamous cancers, but the regulators that oppose this in cervical cancer are poorly characterized. We looked for candidate regulators in cervical squamous cell carcinoma, using the TCGA cervical cancer cohort (CESC) restricted to squamous samples.

Methods: A 24-gene panel of methylation-silenced genes was tested for regulator enrichment in a TCGA CESC ARACNe network (616 regulators; none significant after Benjamini-Hochberg correction, so enrichment only ranked candidates). The screen was repeated with a 67-gene panel as a post hoc sensitivity analysis (pass rule fixed beforehand). DACH1, the top candidate by panel coverage, was chosen for follow-up after the ranking was seen. We correlated DACH1 expression with promoter methylation (HM450) of its predicted targets in squamous samples (n = 253). We then tested robustness: methylation burden, tumor purity, probe-level recomputation, each target's own expression, regulators matched for regulon size and expression, and an independent cohort (GSE68339, n = 121).

Results: DACH1 has no PubMed records in cervical cancer (six in head and neck cancer). It had the highest panel coverage among the ten top-ranked drivers in the 24-gene screen (4 of 23 genes; BH-adjusted p = 0.17) but ranked 158th of 1,521 regulators in the 67-gene panel (BH-adjusted p = 0.43), so the nomination was not robust to panel composition. DACH1 expression correlated inversely with promoter methylation of CADM1 (ρ = −0.249), ESR1 (−0.248) and SLIT2 (−0.263); BH-adjusted p = 8.8e-5 for each. Adjusting for burden and purity left these unchanged. SLIT2 survived probe-level recomputation; ESR1 weakened modestly and CADM1 weakened. Adjusting for each target's own expression removed most of the association, and DACH1 was not distinguishable from 77 matched regulators (one-sided p = 0.10, 0.10, 0.13). ESR1 replicated in GSE68339 (ρ = −0.358, BH-adjusted p = 1.7e-4); CADM1 and SLIT2 trended the same way.

Conclusions: DACH1 is a candidate regulator associated with lower methylation at SLIT2 and ESR1 (CADM1 suggestive: significant in TCGA but weaker with position-based probes and not replicated independently) in cervical squamous carcinoma. The associations are correlative. Most of the association disappears when each target's own expression is accounted for, and DACH1 did not stand out from similar regulators, so specificity to DACH1 is not shown and a role for DACH1 in cervical cancer is not established. ChIP-seq and bisulfite sequencing after DACH1 perturbation are the next step.

Keywords: DACH1, DNA methylation, cervical cancer, master regulator analysis, ARACNe, CADM1, ESR1, SLIT2

Introduction

Aberrant promoter hypermethylation is a well-established mechanism of tumor-suppressor and adhesion-gene silencing across epithelial cancers, and cervical squamous cell carcinoma is no exception: genes including CADM1 and FHIT [1] are recurrently silenced during cervical carcinogenesis, whereas RASSF1A and APC methylation is more frequent in adenocarcinoma than in squamous carcinoma [1]. Far less is known about which transcription factors causally oppose or permit this silencing process — that is, which regulators sit upstream of the methylation-silenced gene network rather than being silenced by it.

DACH1 (Dachshund homolog 1) has been reported as a tumor suppressor in several cancers, with loss, frequently via promoter hypermethylation, documented in breast, gastric, and esophageal carcinoma, although its role is context-dependent (Discussion). In breast cancer, DACH1 is down-regulated and its promoter is methylated in about a third of tumors, more often in triple-negative and metastatic disease [2]; DACH1 has also been reported to bind and inhibit estrogen receptor-α (ESR1) function in ER-α–positive breast cancer [3]. In gastric cancer, DACH1 silencing activates TGF-β signaling and promotes invasion [4]; in esophageal squamous carcinoma, DACH1 methylation is frequent and increases through dysplasia to cancer [5]; and restoring DACH1 expression suppresses tumor growth in xenograft models of both [4,5]. This cross-cancer pattern supports DACH1 as a biologically plausible epigenetically regulated tumor suppressor with an existing, if mechanistically distinct, link to ESR1 — but to our knowledge it has not been examined in cervical cancer, nor as a regulator of methylation-silenced gene programs in cervical or head/neck squamous carcinoma. In head and neck cancer, DACH1 expression is reduced in laryngeal, tongue and hypopharyngeal carcinoma, DACH1 is a target of microRNA-200c in nasopharyngeal carcinoma, and its promoter methylation has been reported in oral and oropharyngeal cancer [6–11].

Here we used network-based master regulator inference against a tumor-specific regulatory network from the TCGA cervical cancer cohort (CESC; squamous cell carcinoma and endocervical adenocarcinoma) to screen a curated panel of methylation-silenced genes for upstream transcriptional drivers, cross-referenced novelty against the PubMed literature for cervical cancer and for head/neck squamous cell carcinoma, and tested the top candidate, DACH1, for an expression-methylation signature in methylation data from the squamous samples of the same TCGA CESC cohort.

Results

Master regulator enrichment ranks DACH1 as the top-coverage driver in the 24-gene panel, but not in a larger panel

A 24-gene panel of methylation-silenced genes recurrently reported in cervical and/or head/neck squamous carcinoma (Methods) was tested for TF driver enrichment against the TCGA CESC ARACNe regulatory network. Twenty-three of 24 panel genes resolved in the network (RASSF1A was not represented under this symbol). A sensitivity rerun using the official symbol RASSF1 (24 of 24 genes resolved; 649 regulators tested) returned the same ten drivers in the same order with the same overlap genes, RASSF1 overlapping none of them; DACH1 enrichment was 8.07 (Fisher p = 1.37e-3; BH-adjusted p = 0.195). Across 616 regulators tested, DACH1 showed the highest signature coverage among the ten top-ranked drivers (4/23 genes: CADM1, EDNRB, ESR1, SLIT2; enrichment = 8.42, Fisher p = 1.16e-3, BH-adjusted p = 0.17). No regulator reached significance after Benjamini-Hochberg correction across all 616 regulators

tested (adjusted p = 0.109–0.17 for the ten drivers), so we use the enrichment to rank and nominate candidates; the evidence for DACH1 rests on the methylation analyses below.

The screen was then repeated with a larger panel of 67 genes (the original 24 plus 43 further genes reported as methylation-silenced in cervical and/or head and neck squamous carcinoma; 66 resolved in the network; 1,521 regulators tested). This was a post hoc sensitivity analysis, not an independent validation, because the added genes were chosen with knowledge of the original panel and DACH1's targets (Methods). The criterion was fixed before the run: DACH1 would count as holding up if it stayed in the top ten by raw Fisher p with a BH-adjusted p no worse than 0.17. It did not. DACH1 still overlapped four panel genes, but its Fisher p was 0.047, it ranked 158th, and its BH-adjusted p was 0.43. The top-ranked regulators in the larger panel were AEBP1 (BH-adjusted p = 0.001), S1PR1 (0.006) and PSD2 (0.037). These three regulators were not examined further: DACH1 had been nominated from the 24-gene panel, and the larger panel was run afterwards only to test whether that nomination held. Re-running the original 24-gene panel through the same analysis reproduced the original result exactly, so the change in rank is due to the panel, not the method. DACH1's prominence in the 24-gene screen therefore depended on the panel and should not be taken as evidence that DACH1 is a driver of these genes; the methylation analyses below are the only support.

In the 24-gene screen, AR showed the strongest statistical enrichment overall (enrichment = 27.20, Fisher p = 1.77e-4, BH-adjusted p = 0.109). Its coverage was narrower: 3 of 23 genes (CADM1, SFRP1, SLIT2). Its regulon is much smaller (95 targets, versus 409 for DACH1), so fewer overlapping genes still give a lower p. AR also has an extensive prior literature in both cervical (124 records) and head/neck (493 records) cancer; these counts may be inflated by symbol ambiguity. This indicates an established rather than novel regulatory relationship. SOX8 matched AR's coverage (3 of 23 genes: DCC, FHIT, TFPI2; regulon of 152 targets) at a comparable Fisher p (7.0e-4; BH-adjusted p = 0.17) and was likewise novel in cervical cancer (1 record) with a small head/neck literature (5 records). We prioritized DACH1 over SOX8 for downstream validation because it covered one additional panel gene despite a similar enrichment p-value, and because DACH1 — unlike SOX8 — has a reported cross-cancer history as a methylation-silenced tumor suppressor (Introduction), giving its candidacy independent biological support beyond the network statistics alone. SOX8 remains a reasonable candidate for follow-up and is reported in full in Supplementary Table S1.

TF Driver	Coverage	Enrichment	p-adj (BH, all 616)	CESC records	HNSC records	Classification
DACH1	4/23 genes	8.42	0.17	0	6	Novel (CESC); emerging (HNSC)
AR	3/23 genes	27.20	0.109	124	493	Established, both contexts
SOX8	3/23 genes	17.00	0.17	1	5	Novel (CESC); emerging (HNSC)

Table 1. Top-ranked TF drivers by signature coverage and statistical enrichment against the TCGA CESC ARACNe network, with PubMed record counts specific to cervical (CESC) and head/neck squamous cell carcinoma (HNSC) (query definitions in Methods; novel, < 5 records; emerging, 5–20; established, > 20). p-adj is Benjamini-Hochberg over all 616 regulators tested; none reached 0.05, so drivers are ranked by panel coverage. Shown are DACH1, AR and SOX8 from the ten top-ranked drivers; the full driver list, including SALL2 (also 3 of 23 genes), provisional (≤ 2-gene overlap) candidates and per-source CASCADE evidence, is provided in Supplementary Table S1.

The literature gap for DACH1 is specific to cervical cancer (no records) and is small in head and neck cancer (six records: hypopharyngeal, laryngeal, two tongue, one nasopharyngeal, and one oral/oropharyngeal methylation study [6–11]), and it is not gene-wide: DACH1 is reported as a methylation-silenced tumor suppressor in breast, gastric, and esophageal cancer (Introduction). The DACH1–ESR1 finding is a new facet of a partially existing relationship — prior literature documents DACH1 binding and inhibiting ER-α in breast cancer, whereas this study identifies an inverse correlation between DACH1 expression and ESR1 promoter methylation in CESC. CADM1 and SLIT2 appear to be entirely unreported DACH1 targets in any cancer context.

DACH1 lacks corroborating support from the four CASCADE corroboration sources (DoRothEA, LINCS, DepMap, super-enhancer annotation) in the initial master regulator analysis, in contrast to AR and SOX8 which both carried DoRothEA curated-regulon validation; the TCGA methylation-correlation analysis (a different data type, from the same cohort as the network) therefore provides the primary supporting evidence rather than supplementing CASCADE cross-validation. No LINCS knockdown signatures or DepMap essentiality data were returned for DACH1 in this context, consistent with DACH1 being understudied in perturbation screens relative to oncogenes; absence from these databases most likely reflects a profiling gap and is not by itself evidence against functional relevance. DACH1 is also absent from IntOGen's Compendium of Mutational Cancer Driver Genes [12] in both CESC and HNSC (PIK3CA confirmed present in both as a positive control); as with the CASCADE evidence sources above, this reflects the absence of a positive-selection signal specifically in this mutation-based compendium, not confirmation that DACH1 lacks driver activity through mechanisms — such as promoter methylation — that this compendium does not capture.

A subsequent causal chain analysis (Pipeline 3), run when the tool supported only the GREmLN population-average baseline network (epithelial cells), predicted 154 DACH1 targets, including CDH2 (N-cadherin, an established EMT and invasion marker) and FAT1 (a tumor suppressor recurrently mutated in squamous cancers including CESC and HNSC), which relate to DACH1's reported role in suppressing invasion in other epithelial cancers. This is a population-average result: it did not carry over to the TCGA CESC tumor network (409 targets), where CDH2 and FAT1 are not among those targets, so these genes should not be read as cervical tumor-specific findings. This analysis does not add evidence for DACH1 as a regulator in cervical tumors.

DACH1 expression correlates inversely with target promoter methylation in TCGA CESC

To test whether the DACH1–target relationships inferred from network topology appear in a different data type (methylation, from the same cohort), we computed Spearman correlation between DACH1 RNA-seq expression and HM450 promoter methylation beta values. We did this for each of its four predicted regulon targets. Only squamous cell carcinoma samples within the TCGA CESC cohort were used (n = 253, after excluding adenocarcinoma and

adenosquamous subtypes and dropping three squamous samples lacking matched RNA-seq data).

Target gene	n	Spearman ρ	p-value	p-adj (BH, 4 targets)	Direction
CADM1	253	−0.249	6.38e-5	8.8e-5	Inverse — supports hypothesis
ESR1	253	−0.248	6.60e-5	8.8e-5	Inverse — supports hypothesis
SLIT2	253	−0.263	2.35e-5	8.8e-5	Inverse — supports hypothesis
EDNRB	253	−0.157	1.23e-2	1.23e-2	No clear trend — does not support (|ρ| < 0.2)

Table 2. Spearman correlation between DACH1 expression and promoter methylation at predicted target loci, squamous-restricted TCGA CESC cohort (n = 253; cBioPortal cesc_tcga_rna_seq_v2_mrna and cesc_tcga_methylation_hm450 profiles). Methylation is the cBioPortal gene-level value, which cBioPortal takes from the probe most anticorrelated with the gene's expression; a position-based recomputation is in Robustness checks. Negative ρ indicates higher DACH1 expression associated with lower target methylation. p-values are raw; p-adj is Benjamini-Hochberg across the four targets.

Three of four predicted targets — CADM1, ESR1, and SLIT2 — showed a consistent inverse correlation between DACH1 expression and promoter methylation in the squamous-restricted cohort (|ρ| = 0.25–0.26, all p < 7e-5; Benjamini-Hochberg-adjusted p = 8.8e-5 for each across the four targets tested), exceeding both the directional (|ρ| > 0.2) and significance thresholds pre-specified for this analysis. For comparison, the unfiltered mixed-histology cohort (n ≈ 306, ~83% squamous) produced modestly inflated effect sizes (|ρ| = 0.34–0.37), consistent with residual histotype-driven covariation; the squamous-only results are the primary reported values throughout this paper. The direction (higher DACH1, lower methylation) is what would be expected if DACH1 opposed silencing at these loci; the robustness checks below test how far the data support that. EDNRB, despite its inclusion in the ARACNe-derived regulon overlap, showed a nominally significant but below-threshold association in the squamous-only cohort (ρ = −0.157, p = 1.23e-2; |ρ| < 0.2 effect-size threshold), indicating that the network-inferred edge for this target does not meet the pre-specified criteria for a supporting regulatory relationship (Figure 1).

An independent replication cohort (GEO GSE68339, cohort 2, cervical squamous carcinoma, n = 121, using a different expression platform [Illumina HumanHT-12 V4] and a defined methylation-probe-selection method; Supplementary Methods) [13] was tested for the same DACH1–target relationship (Supplementary Methods): ESR1 replicated (ρ = −0.358, p = 5.6e-5, BH-adjusted p = 1.7e-4; partial ρ = −0.356 after burden-index correction), while CADM1 and SLIT2 trended in the same direction but did not reach significance after correction at this smaller sample size (BH-adjusted p ≈ 0.061–0.062), and EDNRB was not significant (ρ = −0.045, p = 0.62). Full cohort description, probe-selection methodology, and the complete comparison table are provided in Supplementary Table S2.

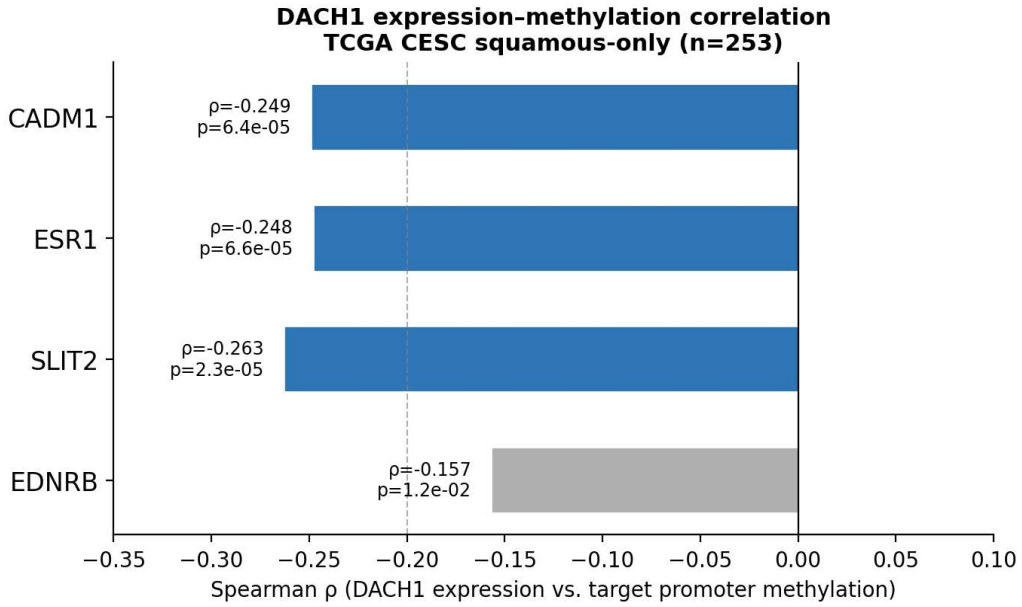


Figure 1. Spearman correlation coefficients between DACH1 expression and promoter methylation for each predicted target gene, squamous-restricted TCGA CESC (n = 253). Blue bars denote targets meeting both significance ($p < 0.05$) and directional-effect ($|\rho| > 0.2$) thresholds; gray denotes targets that do not (EDNRB: nominally significant at $p = 0.012$ but $|\rho| = 0.157$, below the pre-specified effect threshold). Values use cBioPortal gene-level methylation; position-based values (CADM1 −0.13, ESR1 −0.19, SLIT2 −0.26) are in Robustness checks. Values shown are summary correlation statistics; per-sample data available via cBioPortal (cesc_tcga study).

Associations are not explained by a global methylation burden index

DACH1 is itself reported as methylation-silenced in several other cancer types (Introduction), raising the possibility that the CADM1/ESR1/SLIT2 correlations reflect a shared genome-wide hypermethylation phenotype (a CIMP-like pattern) rather than a locus-specific regulatory relationship. To test this, we computed partial Spearman correlations between DACH1 expression and each target's methylation, controlling for a methylation burden index (mean HM450 beta across up to eight established CESC methylation markers: CDKN2A, RARB, DAPK1, CDH1, MGMT, MAL, SOCS1, PAX1; PAX1 methylation data were unavailable in the squamous subset and the index was computed from the remaining seven genes for this analysis).

Target gene	n	Naive ρ	Naive p	Partial ρ	Partial p
CADM1	253	−0.249	6.38e-5	−0.245	8.48e-5
ESR1	253	−0.248	6.60e-5	−0.242	1.04e-4
SLIT2	253	−0.263	2.35e-5	−0.258	3.22e-5

Table 3. Naive vs. partial Spearman correlations between DACH1 expression and target promoter methylation, squamous-restricted TCGA CESC (n = 253). Partial correlations control for a methylation burden index (mean HM450 beta across CDKN2A, RARB, DAPK1, CDH1, MGMT, MAL, SOCS1; PAX1 excluded due to missing squamous-subset methylation data). EDNRB is not shown; its burden-adjusted values are in Supplementary Table S3.

All three partial correlations were essentially unchanged from the naive squamous-only correlations (CADM1: ρ = −0.249 → −0.245; ESR1: −0.248 → −0.242; SLIT2: −0.263 → −0.258), and remained significant after adjustment (all p < 1.1e-4). This indicates that the DACH1-target associations are not an artifact of a shared genome-wide hypermethylation burden as captured by this index.

Robustness checks: tumor purity, probe-level methylation, target expression and matched regulators

Tumor purity. ABSOLUTE purity estimates [14] (TCGA PanCanAtlas, via the GDC, distributed with [15]) were available for 241 of the 253 squamous samples. DACH1 expression was not correlated with purity (ρ = −0.04). Partial correlations controlling for purity were essentially unchanged (CADM1 −0.243, ESR1 −0.263, SLIT2 −0.269; unadjusted −0.249, −0.248, −0.263), as were those controlling for purity and the burden index (overall methylation across seven CESC marker genes; Table 3) together (−0.232, −0.247, −0.261).

Probe-level methylation. Methylation is measured at individual CpG sites (probes) on the HM450 array. cBioPortal's value for each gene is the probe most anticorrelated with that gene's expression, which could inflate the correlations. We instead rebuilt each gene's value from probes at a fixed position near the transcription start site (TSS), chosen without reference to expression: the mean of TSS200 probes (0–200 bases upstream of the TSS), or TSS1500 probes (200–1,500 bases) if none. SLIT2 was unchanged (ρ = −0.264 vs. cBioPortal −0.263), ESR1 weakened modestly (−0.190 vs. −0.248) and CADM1 weakened (−0.131 vs. −0.249). EDNRB was null (−0.093 vs. −0.157).

Adjustment for target expression. Methylation at these loci is strongly anticorrelated with the target's own expression (ρ = −0.47 to −0.58), and DACH1 expression correlates positively with target expression (ρ = 0.29–0.47). Given the target's own expression, the partial correlation between DACH1 and methylation was near zero for CADM1 (−0.024) and SLIT2 (−0.053) and −0.129 for ESR1 (p = 0.04) with cBioPortal values; with probe-level values it was +0.04, −0.08 and −0.08. The associations therefore cannot be separated from the DACH1–target expression relationship with these data.

Matched-regulator null. To ask whether the pattern is specific to DACH1, we repeated the analysis for other regulators. For each of the 5,999 network regulators with expression data, we correlated its expression with promoter methylation of SLIT2, ESR1 and CADM1 in the same 253 samples. We then compared DACH1 with the 77 regulators most similar to it: regulon size within 0.5–2 times DACH1's 409 targets, and mean expression within ±1 log2 unit (rule fixed before the analysis). DACH1's correlations here were −0.243 (SLIT2), −0.215 (ESR1) and −0.114 (CADM1); these differ slightly from Table 2 because a different expression source was used. The matched regulators were also mildly negative (median ρ = −0.08, −0.08 and −0.02), so a weak negative correlation is common. The fraction of matched regulators at least as negative as DACH1 was 0.10 (SLIT2), 0.10 (ESR1) and 0.13 (CADM1), so about 1 in 10 does as well or better and DACH1 does not stand out. The pool is small (the smallest possible fraction is 1 in 78, about 0.013), so the test has limited resolution, and specificity to DACH1 is not shown.

Exploratory comparison with SOX8 and SALL2

As a post hoc comparison, we ran the same correlation for two other regulators from the screen, SOX8 and SALL2, against their own overlap genes (SOX8: DCC, FHIT, TFPI2; SALL2: CADM1, SCGB3A1, WIF1) in the same 253 samples. With cBioPortal values none reached |ρ| > 0.2:

DCC had no data, SOX8–FHIT was positive, SOX8–TFPI2 was −0.187 (p = 0.003), and the strongest SALL2 value was SCGB3A1 at −0.169 (p = 0.007). This does not show that DACH1 is special: SOX8 and SALL2 ranked similarly in the screen but were tested against different genes, DCC had no cBioPortal data, and SOX8–TFPI2 and SOX8–DCC (probe-level ρ = −0.293) and SALL2–SCGB3A1 (probe-level ρ = −0.235) were as strong as the DACH1 associations. Only SOX8–FHIT clearly differed from all three DACH1 associations. SOX8 and SALL2 are therefore exploratory comparators. Full values are in Supplementary Tables S1 (Note B) and S3.

Discussion

This study nominates DACH1 as a candidate regulator associated with reduced promoter methylation at SLIT2 and ESR1, and more tentatively CADM1, in cervical squamous carcinoma. The candidate came from combining ARACNe-based enrichment of a methylation-silenced gene panel against a tumor-specific network, used only to rank candidates and not robust to enlarging the panel to 67 genes (Results), with expression–methylation correlation in squamous TCGA samples (n = 253). The effect sizes (|ρ| ≈ 0.25) and significance levels (all p < 7e-5) make the SLIT2 association unlikely to be assay or analytic noise alone, and argue against it less strongly for ESR1, which weakened at probe level (−0.190); both survived adjustment for methylation burden and tumor purity. Specificity is not shown, however. Part of the association is expected without any regulatory role: the targets were inferred from expression in this same cohort, so they are co-expressed with DACH1 (ρ = 0.29–0.47), and each target's expression is itself inversely related to its own promoter methylation (ρ = −0.47 to −0.58), so an inverse DACH1–methylation correlation follows from co-expression alone. Adjusting for each target's own expression removed most of the association, so the data are as consistent with DACH1 simply varying together with these targets, without a regulatory role in cervical cancer, as with a methylation-specific effect. The SOX8/SALL2 comparison was inconclusive, and DACH1 was not distinguishable from regulators matched for regulon size and expression (Results). The association is therefore best read as a candidate relationship whose specificity to DACH1 remains untested. ESR1 already had a documented relationship to DACH1 (Introduction); CADM1 and SLIT2 are new candidates. SLIT2 was the most robust to probe-level recomputation, ESR1 replicated in an independent cohort, and CADM1 weakened with independent probes.

DACH1 has no prior literature in cervical cancer and a small head and neck cancer literature (six papers, one of which reports DACH1 promoter methylation in oral and oropharyngeal cancer), even though it is reported as a recurrently methylation-silenced tumor suppressor in breast, gastric, and esophageal cancer; in gastric cancer its loss activates TGF-β signaling and promotes invasion and proliferation. Its function is, however, reported to be context-dependent: a colorectal cancer study reports that high DACH1 expression is associated with poor prognosis and that its knockdown reduced tumor cell growth in vitro and in vivo [16], so a tumor-suppressive role in cervical squamous carcinoma is an inference from other tissues rather than an established finding. The cervical literature gap may reflect under-investigation, or it may reflect that DACH1 is not relevant in cervical cancer; our data cannot tell which.

We note an asymmetry: the inverse direction (higher DACH1, lower methylation) is opposite to a model in which DACH1 itself drives silencing of its targets. Several readings remain open. DACH1 may actively maintain an unmethylated state at CADM1, ESR1 and SLIT2, consistent with its reported tumor-suppressive rather than silencing role in other cancers. It may be co-regulated with a broader demethylation or chromatin-accessibility program. Or it may have no regulatory role and simply vary together with these targets, as noted above. This study cannot

distinguish among them; the first is testable by DACH1 perturbation followed by bisulfite sequencing.

The EDNRB result is reported alongside the three positive findings rather than omitted. Its lack of correlation despite ARACNe regulon membership illustrates a general limitation of network-inference-based target prediction: edges supported by mutual-information-based network reconstruction do not always correspond to a detectable methylation-expression relationship in independent bulk data, and should not be assumed to generalize uniformly across all genes in a regulon. With probe-level methylation values EDNRB was null ($\rho = -0.093$), consistent with this reading.

Limitations. The network enrichment that nominated DACH1 did not reach significance after correction across all 616 regulators tested, so it serves as a candidate-prioritization step. The panel was small (24 genes, 23 resolved in the network), so the enrichment test had few overlapping genes per regulator and limited power; when the panel was enlarged to 67 genes DACH1 ranked 158th of 1,521 regulators (BH-adjusted $p = 0.43$; Results), so the enrichment nomination is not robust to panel composition and offers no support for DACH1 beyond having prompted the methylation analyses. The 43 genes added to make the larger panel were chosen by the author with a language model's assistance, from recalled literature, and not individually verified (Methods), so that panel is itself an imperfect test.

All findings are correlative and derived from bulk tumor profiling; we cannot establish causality or resolve cell-type or subclonal heterogeneity. Correlations were computed in squamous samples only ($n = 253$), excluding adenocarcinoma and adenosquamous subtypes (~17% of the cohort); the mixed-histology cohort gave inflated effect sizes ($|\rho| \approx 0.35$ vs. ≈ 0.25), so we report the squamous-only results and do not use the pooled estimate. DACH1 expression is right-skewed with a large low-expression cluster (94 of 253 samples, 37%, <10 RSEM), so the correlations may partly reflect discrete expression states rather than a continuous relationship. Purity was controlled and did not change the associations; copy-number variation at the CADM1, ESR1 and SLIT2 loci remains uncontrolled. cBioPortal's gene-level methylation values select the probe most anticorrelated with expression, which can inflate correlations; with probe-level data the CADM1 and ESR1 associations weakened (Results). Adjusting for each target's own expression removed most of the association, so a methylation-specific effect cannot be separated from the expression relationship, or from DACH1 having no regulatory role in cervical cancer. DACH1 was not distinguishable from matched regulators on any single target ($n = 77$; one of several reasonable matching rules), and the enrichment nomination did not hold in the larger panel, so specificity to DACH1 is not shown.

In an independent cohort (GSE68339, $n = 121$), only the ESR1 association replicated after multiple-testing correction; the CADM1 and SLIT2 associations therefore rest primarily on the TCGA CESC cohort and warrant further independent validation. We tested global tumor hypermethylation burden as a confound via partial correlation (Results); the associations were essentially unchanged. We did not test whether DACH1's own promoter methylation, or HPV type or TCGA methylation subtype, accounts for the association. DACH1 returned no support from any of the four CASCADE evidence sources (LINCS knockdown, DepMap essentiality, super-enhancer status, DoRothEA TF annotation) used for corroboration in the Orchestra paper [17]. That paper calibrated its at-least-two-of-four rule against IntOGen mutational drivers, a ground truth that does not capture genes acting through epigenetic mechanisms, and its DepMap source pools cell lines across lineages; the absence of corroboration here therefore provides no independent support for DACH1. The remaining evidence is the methylation–expression correlation in TCGA CESC and, for ESR1, the independent-cohort replication; as noted above, these do not establish that the relationship is specific to DACH1. ChIP-seq or

CUT&RUN evidence for DACH1 occupancy at the CADM1, ESR1, and SLIT2 promoters, combined with bisulfite sequencing following DACH1 knockdown or overexpression in cervical cancer cell lines, would be the logical next step to test a direct regulatory relationship.

Methods

Gene panel curation

A panel of 24 genes recurrently reported as silenced by promoter methylation in cervical and/or head/neck squamous cell carcinoma was assembled from the literature with the assistance of a language model (CDKN2A, CADM1, MAL, DAPK1, RASSF1A, CDH1, TIMP3, SOCS1, FHIT, PAX1, WIF1, SFRP1, TFPI2, DLEC1, SCGB3A1, RARB, APC, MGMT, ESR1, GSTP1, EDNRB, SLIT2, HOXA9, DCC). The larger-panel enrichment test (Results) used these 24 genes plus 43 further genes: CCNA1, CDH13, CDKN2B, DKK3, EPB41L3, HIC1, JAM3, MLH1, PAX5, RUNX3, SFRP2, SOX1, TWIST1, HS3ST2, ZNF582, SOX17, LMX1A, NKX6-1, GATA4, GATA5, RARRES1, TMEFF2, CCND2, RASSF2, CXCL12, SFRP4, SFRP5, THBS1, KIF1A, IRX1, NID2, TCF21, SYNE1, SHISA3, GRIA4, PCDH10, ID4, CHFR, SYK, TP73, PTEN, BRCA1, STK11.

Master regulator enrichment

The enrichment analysis and all tumor-specific claims use the TCGA CESC tumor network; the GREmLN population-average baseline network was used only for the post hoc causal chain analysis described below. The gene panel was tested against a TCGA CESC tumor-specific ARACNe-reconstructed regulatory network (6,054 regulators with target sets, of which 616 had at least one panel-gene target and were tested) [18–20] using Fisher's exact test for regulon overlap enrichment, with Benjamini-Hochberg correction applied across all regulators tested. Because the correction covers all 616 regulators, no driver reached an adjusted p of 0.05 (0.109–0.17); enrichment was therefore used to rank and nominate candidates, and the supporting evidence is the methylation analyses. Regulators with no panel-gene target (5,438 of 6,054) were not tested. Including them with Fisher p = 1 would enlarge the correction family about tenfold and could only raise the adjusted p-values, so the conclusion that none reached significance does not change.

The enrichment screen was repeated with a larger panel: the original 24 genes plus 43 further genes (67 genes, 66 resolved). The author chose the 43 genes with the assistance of a language model. They came from recalled literature on methylation silencing in cervical and/or head and neck squamous carcinoma, not from a systematic search or database, and were not individually verified against the source publications. The genes were chosen before the run, without consulting any result of this analysis, but with knowledge of the original panel and its DACH1 targets. The run used Orchestra's analyze_gene_signature workflow on the TCGA CESC network. Versions (per git describe) were Orchestra v1.3.11-7-g4c4c0dd, RegNetAgents v1.3.0-2-g7dabb72 and CASCADE v1.5.0-3-g50c02d4; the CASCADE working copy differed from that commit only by a gene-identifier cache file. BH correction was applied over all 1,521 regulators tested. An independent calculation of the same Fisher test reproduced the tool's top ten, overlaps and adjusted p-values, and the original 24-gene result, exactly. The pass rule was fixed before running: DACH1 in the top ten and BH-adjusted p no worse than 0.17. The panel was not changed after the result was seen. The result is reported in Results.

Enrichment analysis and causal chain analysis were performed using Orchestra (orchestra_mcp_server.py; analysis commit 8c048cb, included in Orchestra v1.0.0, Bird, 2026;

doi:10.5281/zenodo.21192969, concept DOI 10.5281/zenodo.21192968) [17]. Orchestra composes two MCP servers: RegNetAgents [21] (ARACNe regulatory network data) and CASCADE [22] (downstream perturbation prediction and functional evidence sources). It acts as the MCP server for the calling client and as an MCP client to the two independently running servers. A LangGraph-based workflow coordinates the calls between them (orchestra_langgraph_workflow.py). The revision checks and PubMed re-counts reported in this version used Orchestra v1.3.11-3-gd090671, RegNetAgents v1.3.0-1-g54b1cf8 and CASCADE v1.5.0-2-g281bb03 (git describe); a rerun of the original 23-gene panel under these versions reproduced the Table 1 drivers, ranks, overlaps, enrichments and Fisher p-values exactly; adjusted p-values were then recomputed over all regulators tested with Orchestra commit 4c4c0dd (Supplementary Table S1).

The causal chain analysis (Pipeline 3) queried the GREmLN population-average baseline network for epithelial cells [23], which is derived from CZ CELLxGENE Discover Census data [24]; at the time of the original analysis this tool supported only the population-average baseline network, and the comparison with the TCGA CESC tumor network is reported in Results. Because RegNetAgents' gene-identifier mapping was corrected after the original run, the analysis was re-run (Orchestra v1.3.11-3-gd090671, RegNetAgents v1.3.0-2-g7dabb72, CASCADE v1.5.0-3-g50c02d4); it again returned 154 predicted DACH1 targets, including CDH2 and FAT1. Top-ranked drivers were checked, where available, against the four orthogonal CASCADE evidence sources used in Orchestra's corroboration criterion [17]: DoRothEA curated regulons [25], LINCS knockdown signatures, DepMap essentiality data, and super-enhancer annotation. cBioPortal TCGA expression [26,27] was examined separately. For DACH1, none of the four sources returned supporting evidence. The pipeline flagged DACH1 as topologically central by network PageRank but lacking independent experimental or curated-regulon evidence. Pan-cancer cBioPortal expression was near baseline ($z = 0.01$). By contrast, AR and SOX8 both carried DoRothEA curated-regulon validation. Drivers with regulon overlap ≤ 2 genes were flagged as provisional and excluded from primary interpretation, because their enrichment is not robust to small changes in the gene list.

Cross-context novelty was assessed via PubMed record counts for each top-ranked driver in cervical cancer and in head/neck squamous cell carcinoma. Orchestra's built-in head/neck query uses a short fixed phrase list and misses papers that name only a subsite (laryngeal, oral, tongue, hypopharyngeal). The counts in Table 1 were therefore obtained on 2026-10-18 with the following PubMed queries: GENE[tiab] AND ((cervical[tiab] OR cervix[tiab]) AND (cancer*[tiab] OR carcinoma*[tiab] OR neoplasm*[tiab] OR tumor*[tiab] OR tumour*[tiab] OR CESC[tiab])) for cervical cancer, and GENE[tiab] AND (head and neck[tiab] OR HNSC*[tiab] OR laryn*[tiab] OR oral[tiab] OR oropharyn*[tiab] OR hypopharyn*[tiab] OR tongue[tiab] OR pharyn*[tiab]) AND (cancer*[tiab] OR carcinoma*[tiab] OR neoplasm*[tiab] OR tumor*[tiab] OR tumour*[tiab] OR SCC[tiab] OR HNSCC[tiab]) for head/neck squamous carcinoma. Records were classified as novel (< 5), emerging (5–20) or established (> 20). The head/neck query returned seven records for DACH1; six are counted, including one nasopharyngeal study, and one was excluded on reading (a transcriptomic study of viral infection in neurons and astrocytes). The DACH1 records were read individually; the counts for the other drivers in Table 1 are unread query totals.

Enrichment statistics and p-values were computed against the 23 network-resolved genes (RASSF1A was excluded from testing as it did not resolve in the network; a sensitivity rerun with RASSF1 is reported in Results); the coverage percentage in Supplementary Table S1 has been corrected to use 23 as the denominator for consistency, as the source pipeline's summary output had reported coverage against the full 24-gene input panel. Driver-gene status for DACH1 was additionally checked against IntOGen's Compendium of Mutational Cancer Driver

Genes [12] (release 2024.09.20, 633 genes) for CESC and HNSC, with PIK3CA as a positive control; full method detail is provided in Supplementary Methods.

Methylation-expression correlation

For DACH1 and its four ARACNe-predicted regulon targets (CADM1, EDNRB, ESR1, SLIT2), Spearman rank correlation was computed between DACH1 RNA-seq (RSEM, cesc_tcga_rna_seq_v2_mrna) expression and target gene promoter methylation (Illumina HM450 [28], cesc_tcga_methylation_hm450 beta values) restricted to squamous cell carcinoma samples in the TCGA CESC cohort (n = 253 after histotype filtering and exclusion of three squamous samples lacking matched RNA-seq data) via cBioPortal [26,27]. Targets with negative ρ and $|\rho| > 0.2$ at $p < 0.05$ were considered to support an inverse expression-methylation relationship; targets with fewer than 5 matched samples would have been flagged as uninformative. The four p-values were additionally adjusted by the Benjamini-Hochberg method across the four targets; the criterion applied to the targets was unchanged.

To test whether these associations were attributable to global tumor hypermethylation rather than locus-specific regulation, a methylation burden index was constructed as the mean HM450 beta value across up to eight established CESC methylation markers (CDKN2A, RARB, DAPK1, CDH1, MGMT, MAL, SOCS1, PAX1; none of these is a DACH1 target); PAX1 methylation data were unavailable for squamous samples and the index was computed from the remaining seven genes. Partial Spearman correlations were computed using a residuals-on-ranks approach (see Supplementary Methods).

As a post hoc exploratory comparison, the same naive Spearman correlation procedure (without partial adjustment) was applied to SOX8 and SALL2 and their predicted targets; partial correlation was not performed for either as no informative target met the $|\rho| > 0.2$ threshold at the naive stage.

An independent-cohort replication of the ESR1, CADM1, and SLIT2 correlations was additionally performed using GEO GSE68339 (cohort 2, n = 121) [13]; cohort description, promoter-probe-selection methodology (TSS200-region averaging per gene, with TSS1500 fallback for genes lacking TSS200 probes), and the full comparison are provided in Supplementary Methods and Supplementary Table S2.

Robustness checks: ABSOLUTE purity estimates [14] (TCGA PanCanAtlas, via the GDC, distributed with [15]; 241 of 253 squamous samples) were used in partial Spearman correlations (residuals-on-ranks, as above). Promoter methylation was recomputed from probe-level HM450 beta values (UCSC Xena GDC hub, TCGA-CESC; Illumina HM450 manifest, GEO platform GPL13534), using probes chosen by position relative to the transcription start site, not by correlation with expression: the mean of TSS200 probes, falling back to TSS1500 probes when a gene has none. Each target's own RNA-seq expression was added as a covariate in the same partial-correlation framework. For the SOX8 comparison, 95% confidence intervals were obtained by Fisher z transformation and by bootstrap, and differences between correlations by paired bootstrap (5,000 resamples); values are in Supplementary Table S3.

To ask whether the association is specific to DACH1, the same correlation was computed for every other network regulator using TCGA-CESC STAR TPM expression (log2[TPM+1]) for the same 253 samples from the UCSC Xena GDC hub [29] (TCGA-CESC.star_tpm.tsv.gz; Ensembl identifiers mapped to symbols with GENCODE v36), against promoter methylation (TSS200 probes, TSS1500 fallback, as above) of SLIT2, ESR1 and CADM1. The comparison set was the regulators with regulon size between 0.5 and 2 times DACH1's (409 targets) and mean expression within ±1 log2 unit of DACH1's; this matching rule and the one-sided $p \leq 0.05$

criterion were fixed before the analysis was run, and the empirical p was (number of matched regulators with ρ at or below DACH1's + 1)/(pool size + 1). Full results are in Supplementary Table S3. Statistics were computed in Python 3.12.8 with SciPy 1.16.3, NumPy 2.3.5 and pandas 2.3.3.

AI Usage Disclosure

Claude Code (Anthropic) was used to assist with software implementation, statistical analysis, and paper drafting. As the client calling Orchestra's tools, the language model chose which tools to run and in what order, summarized their outputs, and assisted the author in compiling the 24-gene panel and in choosing the 43 additional genes of the 67-gene panel, drawing on recalled literature and not a systematic search (details under Master regulator enrichment in Methods). The statistical results reported in this paper — Fisher's exact test enrichment values, Benjamini-Hochberg adjusted p-values, and Spearman/partial Spearman correlation coefficients — were computed by deterministic code, either Orchestra's tools or the standalone scripts listed under Data and code availability, and were not generated by the language model. The author reviewed and approved all analyses and text and takes full responsibility for the methods and conclusions.

References

1. Yang HJ. Aberrant DNA methylation in cervical carcinogenesis. Chin J Cancer. 2013;32(1):42-48. doi:10.5732/cjc.012.10033. PMID: 22943599.

2. Nasirpour MH, Salimi M, Majidi F, Minuchehr Z, Mozdarani H. Study of DACH1 expression and its epigenetic regulators as possible breast cancer-related biomarkers. Avicenna J Med Biotechnol. 2023;15(2):108–117. PMID: 37034893.

3. Popov VM, Zhou J, Shirley LA, Quong J, Yeow WS, Wright JA, Wu K, Rui H, Vadlamudi RK, Jiang J, Kumar R, Wang C, Pestell RG. The cell fate determination factor DACH1 is expressed in estrogen receptor-alpha-positive breast cancer and represses estrogen receptor-alpha signaling. Cancer Res. 2009;69(14):5752–5760. doi:10.1158/0008-5472.CAN-08-3992. PMID: 19605405.

4. Yan W, Wu K, Herman JG, Brock MV, Zhou Y, Lu Y, Zhang Z, Yang Y, Guo M. Epigenetic silencing of DACH1 induces the invasion and metastasis of gastric cancer by activating TGF-β signalling. J Cell Mol Med. 2014;18(12):2499–2511. doi:10.1111/jcmm.12325. PMID: 24912879.

5. Wu L, Herman JG, Brock MV, Wu K, Mao G, Yan W, Nie Y, Liang H, Zhan Q, Li W, Guo M. Silencing DACH1 promotes esophageal cancer growth by inhibiting TGF-β signaling. PLoS One. 2014;9(4):e95509. doi:10.1371/journal.pone.0095509. PMID: 24743895.

6. Li W, Xu L, Cao J, Ge J, Liu X, Liu P, et al. DACH1 regulates macrophage activation and tumour progression in hypopharyngeal squamous cell carcinoma. Immunology. 2023;170(2):253–269. doi:10.1111/imm.13667. PMID: 37243970.

7. Zhang J, Ren X, Wang B, Cao J, Tian L, Liu M. Effect of DACH1 on proliferation and invasion of laryngeal squamous cell carcinoma. Head Face Med. 2018;14(1):20. doi:10.1186/s13005-018-0177-1. PMID: 30261897.

8. Zhang L, Xiang FG, Wang CQ. Expression and significance of human Dachshund homolog 1 in tongue squamous cell carcinoma and tongue atypical hyperplasia. Shanghai Kou Qiang Yi Xue. 2017;26(1):64–68 (in Chinese). PMID: 28474069.

9. Zhang L, Wang CQ, Liu F, Dong ZQ, Zhao P, Dong XN, et al. Effects of human Dachshund homolog 1 on the proliferation, migration, and adhesion of squamous cell carcinoma of the tongue. Oral Surg Oral Med Oral Pathol Oral Radiol. 2016;121(1):58–66. doi:10.1016/j.oooo.2015.08.021. PMID: 26577501.

10. Paluszczak J, Sarbak J, Kostrzewska-Poczekaj M, Kiwerska K, Jarmuż-Szymczak M, Grenman R, et al. The negative regulators of Wnt pathway-DACH1, DKK1, and WIF1 are methylated in oral and

oropharyngeal cancer and WIF1 methylation predicts shorter survival. Tumour Biol. 2015;36(4):2855–2861. doi:10.1007/s13277-014-2913-x. PMID: 25487617.
11. Cao W, Sun J. MicroRNA-200c promotes tumor cell proliferation and migration by directly targeting dachshund family transcription factor 1 by the Wnt/β-catenin signaling pathway in nasopharyngeal carcinoma. Anticancer Drugs. 2019;30(3):218–224. doi:10.1097/CAD.0000000000000713. PMID: 30431444.
12. Martínez-Jiménez F, Muiños F, Sentís I, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer. 2020;20(10):555–572. doi:10.1038/s41568-020-0290-x.
13. Lando M, Fjeldbo CS, Wilting SM, Snoek BC, Aarnes EK, Forsberg MF, Kristensen GB, Steenbergen RDM, Lyng H. Interplay between promoter methylation and chromosomal loss in gene silencing at 3p11-p14 in cervical cancer. Epigenetics. 2015;10(10):970–980. doi:10.1080/15592294.2015.1085140.
14. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhim R, Pellman D, Levine DA, Lander ES, Meyerson M, Getz G. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012;30(5):413–421. doi:10.1038/nbt.2203. PMID: 22544022.
15. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J, Lazar AJ, Cancer Genome Atlas Research Network, Cherniack AD, Beroukhim R, Meyerson M. Genomic and functional approaches to understanding cancer aneuploidy. Cancer Cell. 2018;33(4):676–689.e3. doi:10.1016/j.ccell.2018.03.007. PMID: 29622463.
16. Hu X, Zhang L, Li Y, et al. Organoid modelling identifies that DACH1 functions as a tumour promoter in colorectal cancer by modulating BMP signalling. EBioMedicine. 2020;56:102800. doi:10.1016/j.ebiom.2020.102800.
17. Bird JA. Orchestra: Corroboration-Based Regulatory Candidate Discovery via Composed Bioinformatics MCP Agents. 2026. arXiv:2609.05496v4.
18. Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. Nature. 2017;543(7645):378–384. doi:10.1038/nature21386.
19. Margolin AA, Nemenman I, Basso K, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006;7(Suppl 1):S7. doi:10.1186/1471-2105-7-S1-S7. PMID: 16723010.
20. Giorgi FM, Alvarez MJ. ARACNe-AP gene regulatory networks from TCGA tumor datasets [Data set]. Zenodo. 2026. doi:10.5281/zenodo.22918956.
21. Bird JA. RegNetAgents: A Multi-Agent Framework for Cross-Network Regulatory Driver Identification in Cancer Genomics. 2026. arXiv:2607.14097v3.
22. Bird JA. CASCADE: An Agentic Regulatory Network Framework for Patient-Data-Validated Downstream Perturbation Prediction. 2026. arXiv:2608.05359v3.
23. Zhang M, Swamy V, Cassius R, Dupire L, Kanatsoulis C, Paull E, AlQuraishi M, Karaletsos T, Califano A. GREmLN: a cellular graph structure aware transcriptomics foundation model. ICLR 2026. https://openreview.net/forum?id=HdvI8bkdDG.
24. CZI Cell Science Program, Abdulla S, Aevermann B, et al. CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. Nucleic Acids Res. 2025;53(D1):D886–D900. doi:10.1093/nar/gkae1142.
25. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019;29(8):1363–1375. doi:10.1101/gr.240663.118. PMID: 31340985.
26. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401–404. doi:10.1158/2159-8290.CD-12-0095.
27. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1. doi:10.1126/scisignal.2004088. PMID: 23550210.

28. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288–295. doi:10.1016/j.ygeno.2011.07.007. PMID: 21839163.
29. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675–678. doi:10.1038/s41587-020-0546-8.

Supplementary Methods

Partial Spearman correlations were computed using the residuals-on-ranks method. Briefly, the ranks of DACH1 expression and target methylation were each regressed separately on the ranks of the covariate (methylation burden index, tumor purity, or target expression) using ordinary least squares. The Pearson correlation of the two residual vectors was then taken as the partial Spearman coefficient. With more than one covariate, the ranks of all covariates were entered together in the same least-squares fit. This approach approximates the partial Spearman correlation; p-values come from the Pearson test on the residuals.

Independent-cohort replication (GSE68339). GEO series GSE68339 cohort 2 (n = 121) provided matched Illumina HumanHT-12 V4 expression and Illumina HM450 methylation data for cervical squamous carcinoma samples [13]. All 121 samples were annotated as squamous histology, and no additional filtering was applied.
DACH1 expression was measured using probe ILMN_1755741. It was the only one of three HT-12 V4 probes mapping to DACH1 with signal above background (mean ~157, versus ~101–106 for the other two probes).
For each target and burden-index gene, promoter methylation was the mean HM450 beta value across all distinct manifest-annotated TSS200 probes for that gene (Illumina HumanMethylation450_15017482_v.1.1 manifest, GEO platform GPL13534). TSS1500 probes were used only when a gene had no annotated TSS200 probes. This applied to CDKN2A and MAL only, both burden-index genes.
This probe selection differs from cBioPortal's internal gene-level methylation profile used for the primary TCGA analysis. The TCGA values shown in Supplementary Table S2 are the primary-analysis values (Table 2). The two cohorts therefore differ in methylation-probe summarization, expression platform and population.
Benjamini-Hochberg correction was applied within the GSE68339 cohort across the three targets selected from the TCGA analysis (CADM1, ESR1, SLIT2). EDNRB, which was weak in the TCGA analysis, was excluded from that correction. Partial correlation for ESR1 used the same residuals-on-ranks method described above. The burden index was built from the seven GSE68339-available burden genes (RARB, DAPK1, CDH1, MGMT, SOCS1, CDKN2A, MAL), all resolved for all 121 patients.
Driver-gene annotation check. DACH1's cancer-driver status was checked against IntOGen's Compendium of Mutational Cancer Driver Genes [12] (release 2024.09.20) for the cesc and hnsc cancer types via RegNetAgents' driver-annotation module (regnetagents.driver_gene_client.annotate_genes). PIK3CA was used as a positive control and confirmed present as a driver in both CESC and HNSC. DACH1 was absent (IntOGen is_driver: False in both cancer types).

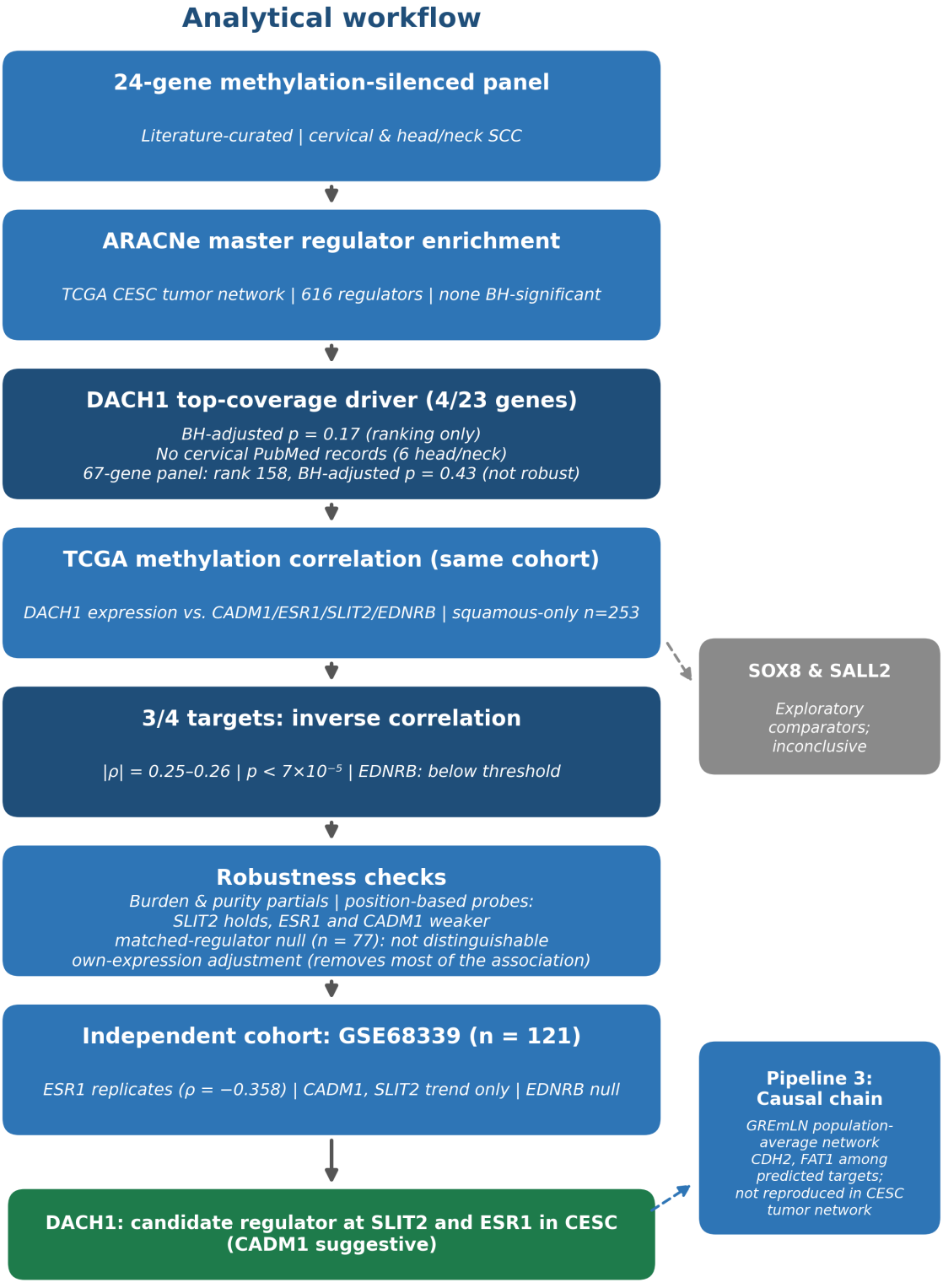

Analytical workflow
24-gene methylation-silenced panel
Literature-curated | cervical & head/neck SCC
ARACNe master regulator enrichment
TCGA CESC tumor network | 616 regulators | none BH-significant
DACH1 top-coverage driver (4/23 genes)
BH-adjusted p = 0.17 (ranking only)
No cervical PubMed records (6 head/neck)
67-gene panel: rank 158, BH-adjusted p = 0.43 (not robust)
TCGA methylation correlation (same cohort)
DACH1 expression vs. CADM1/ESR1/SLIT2/EDNRB | squamous-only n=253
SOX8 & SALL2
Exploratory comparators; inconclusive
3/4 targets: inverse correlation
|ρ| = 0.25–0.26 | p < 7×10⁻⁵ | EDNRB: below threshold
Robustness checks
Burden & purity partials | position-based probes:
SLIT2 holds, ESR1 and CADM1 weaker
matched-regulator null (n = 77): not distinguishable
own-expression adjustment (removes most of the association)
Independent cohort: GSE68339 (n = 121)
ESR1 replicates (ρ = −0.358) | CADM1, SLIT2 trend only | EDNRB null
Pipeline 3: Causal chain
GREmLN population-average network
CDH2, FAT1 among predicted targets;
not reproduced in CESC tumor network
DACH1: candidate regulator at SLIT2 and ESR1 in CESC (CADM1 suggestive)

Figure 2. Analytical workflow. Steps proceed from literature-curated gene panel through network-based master regulator enrichment (Orchestra workflow layer composing RegNetAgents [21] ARACNe + CASCADE [22]), TCGA methylation-expression correlation (same cohort, squamous-only, n=253), robustness checks (methylation-burden and tumor-purity partial correlations, probe-level methylation, a matched-regulator null, and adjustment for each target's own expression), and independent-cohort replication (GSE68339, n=121), to the final candidate finding. A larger 67-gene panel was run afterwards as a post hoc sensitivity analysis of the enrichment step; DACH1 did not hold up in it (noted in the DACH1 box; Results). SOX8 and SALL2 (gray, dashed) were tested as post hoc exploratory comparators outside the main pipeline, applied after DACH1's primary result was established. Pipeline 3 causal chain analysis (blue, dashed) was run as a post hoc exploratory analysis on the GREmLN population-average baseline network, predicting CDH2 and FAT1 among DACH1's downstream targets (not reproduced on the TCGA CESC tumor network). Darker blue boxes denote analytical outputs/results; lighter blue boxes denote analytical steps; the green box is the final candidate finding.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest: The author is the founder of Bird AI Solutions. This work was conducted independently and received no funding.

Ethics statement: This study used publicly available, de-identified data from The Cancer Genome Atlas (TCGA); no new human subjects were recruited, and institutional ethics approval was not required.

Author contributions: J.A.B. designed the study, directed and reviewed all analyses, and wrote the manuscript, with the AI assistance described in the AI Usage Disclosure.

Data and code availability: TCGA CESC RNA-seq and methylation data are publicly available via cBioPortal (https://www.cbioportal.org, study cesc_tcga). The analysis script used to generate the primary correlation results (Tables 2 and 3; dach1_partial_correlation.py) is available at https://github.com/jab57/Orchestra (commit 8c048cb); per-sample expression and methylation data are included in the same repository (dach1_partial_corr_cesc.csv). Network inference outputs and enrichment statistics are provided in Supplementary Table S1. The independent-cohort replication script (gse68339_replication.py) and its outputs (per-sample data, probe-selection record, results comparison, and burden-index partial-correlation result) are available in the same repository (commit 5fbd05a); per-sample GSE68339 data are in gse68339_cohort2_per_sample.tsv. Scripts and outputs for the robustness checks added in this version (purity, probe-level methylation, expression adjustment, RASSF1 rerun, recomputed adjusted p-values) are available in the revision_checks folder of the same repository (commit 180cccc); the aggregated results are in Supplementary Table S3. Expression values for the matched-regulator comparison were taken from the UCSC Xena GDC hub (TCGA-CESC.star_tpm.tsv.gz, SHA-256 977137fde408f880c24ec866801e6ebc0b65aa62c0400857c98adc1c46446cae; Ensembl-to-symbol mapping from GENCODE v36) and are not redistributed here. The scripts (gate3_matched_tf_null_dach1.py, gate4_larger_panel.py, which also contains the 67-gene panel; gate4_run_orchestra_tool.py runs Orchestra's own workflow on that panel, and its output is included with them) and their outputs are available in the screening folder of the same repository (commit 3575833); the matched-regulator results are summarized in Supplementary Table S3, and the 67-gene results are given in Results (full output: gate4_larger_panel_results.json). The TCGA ARACNe regulatory networks used here were generated by Giorgi and Alvarez [20] (CC BY-NC-ND 4.0); they are not redistributed with this work and were installed through RegNetAgents' scripts/setup_tcga_networks.py --accept-license. The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. Correspondence: Jose A. Bird, jbird@birdaisolutions.com.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1fulldriverlist.csv
- SupplementaryTableS2GSE68339replication.csv
- SupplementaryTableS3revisionchecks.csv