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.

NR5A1 consecutive-exon skipping in human adrenal gland, ovary, pituitary and testis: SRSF6 and CLK4 track it in all but testis

Kenji Ohe

oekenji@gmail.com

Fukuoka University https://orcid.org/0000-0002-1958-7572

Research Article

Keywords: NR5A1, SF-1, alternative splicing, exon skipping, GTEx, SRSF6, CLK4, adrenal gland, ovary, steroidogenesis

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11699570/v1

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

Additional Declarations: The authors declare no competing interests.

NR5A1 consecutive-exon skipping in human adrenal gland, ovary, pituitary and testis: SRSF6 and CLK4 track it in all but testis

Kenji Ohe

Department of Clinical Pharmacotherapy, Faculty of Pharmaceutical Sciences, Fukuoka University, 8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan

Correspondence: Kenji Ohe (ohekenji@fukuoka-u.ac.jp), ORCID 0000-0002-1958-7572

Abstract

The NR5A1 pre-mRNA can be spliced so that an upstream exon joins directly to exon 7, removing two or more internal exons and abolishing the full-length SF-1 receptor. The abundance of this splicing differs more than tenfold between human steroidogenic tissues. We asked what sets that level, using exon-exon junction counts for 1,054 GTEx samples of adrenal gland, ovary, pituitary and testis, with duplicated sample entries removed.

The difference is not part of a genome-wide tendency to skip exons. A global skipping index built from the annotated junctions of 48 housekeeping genes was 1.16-fold higher in ovary than in adrenal ($P = 6.3 \times 10^{-13}$), against 11.3-fold for the NR5A1 index; expressed per sample relative to its own global index, ovary stood 9.9-fold above adrenal ($P = 1.5 \times 10^{-57}$).

Within every tissue the index rose as the mRNA abundance of CLK2, CLK4 and SRSF6 fell. Measured against a ten-gene housekeeping total and adjusted for junction depth, RNA integrity, ischaemic time and exonic mapping rate, the rank partial correlations ran from −0.18 to −0.34, all twelve with $P < 0.002$. That normaliser is itself related to the index in ovary ($r = +0.29$) and testis (+0.25) but not in adrenal gland (+0.06) or pituitary (+0.05), so the twelve tests were repeated against the 48-gene panel, total reads, mapped reads and the exon 6→7 junction, with and without the normaliser in the model. The direction held throughout, ten to twelve of twelve cells negative under every scheme. In adrenal gland and pituitary all three factors held under every scheme (−0.15 to −0.33, $P \le 0.02$); in ovary, where every candidate normaliser including total reads was related to the index, SRSF6 and CLK4 held (−0.20 to −0.31) and CLK2 did not; in testis, with library size as the normaliser, none of the three reached significance. CSNK2A1 was not associated in adrenal gland or ovary.

The three factors did not track the global skipping index in a consistent direction, and their association with the NR5A1 index survived adjustment for it, so the effect is specific to NR5A1 rather than genome-wide.

Because a skip index of s leaves a full-length share of $1/(1 + s)$, loss of functional SF-1 alone cannot lower a target gene by more than about 1% per percentage point of index. Associations with steroidogenic genes in adrenal exceeded that bound, ran in the wrong direction, and disappeared when CLK2, CLK4 and SRSF6 were added to the model; in ovary no association exceeded the bound. We therefore find no evidence that this splicing fine-tunes steroidogenesis through SF-1 dosage in normal tissue.

In an unrelated public dataset, 64 RNA-seq runs of the adrenocortical carcinoma line NCI-H295R carried a skip index of 3.62%, against 0.90% for depth-matched GTEx adrenal tissue ($P = 6.5 \times 10^{-8}$), an independent confirmation that the junctions are present in this cell line.

Introduction

SF-1 (NR5A1) sets the transcriptional capacity of the adrenal cortex and the gonads [1], and its dosage matters: increasing gene dosage drives adrenocortical proliferation and tumour formation [2], while heterozygous loss-of-function variants cause 46,XY differences of sex development and primary ovarian insufficiency [3].

How much functional SF-1 a cell has depends not only on transcription but on which transcript is made. The NR5A1 pre-mRNA can be spliced so that an upstream exon joins directly to exon 7, bypassing two or more internal exons. Three such junctions, exon 2→7, exon 3→7 and exon 4→7, have been recovered from adrenocortical carcinoma cells treated with the CK2/CLK inhibitor CX-4945 [4,5], and none of the resulting transcripts encodes a full-length receptor.

In a companion analysis of GTEx and TCGA [6] we defined a skip index, the summed coverage of those three junctions divided by the coverage of the constitutive exon 6→7 junction, and found that it differs greatly between NR5A1-expressing tissues: highest in ovary (median 4.27%) and lowest in adrenal gland (0.38%), which carries the largest denominator of any tissue examined. The index was higher in adrenocortical carcinoma than in normal adrenal, and was not explained by read depth, splicing background or tumour purity. That work established the tissue pattern; it did not address what produces it.

This paper asks what sets the level of NR5A1 consecutive-exon skipping in normal human tissue. We ask three questions in turn. Is the tissue difference part of a genome-wide tendency to skip exons, or specific to NR5A1? Does the abundance of any splicing regulator track the index within tissues, in the direction that the pharmacology predicts? And does the amount of skipping present in normal tissue bear on the steroidogenic programme at a size that loss of full-length SF-1 could account for?

All four tissues analysed here — adrenal gland, ovary, pituitary and testis — are separate sets of donors, so an association that recurs in all four has been seen four times independently.

The tissue difference is not part of a genome-wide tendency to skip exons

Ovary carries eleven times the NR5A1 skip index of adrenal but only 1.16 times its genome-wide exon skipping, so the difference is specific to the gene.

We built a global skipping index from the annotated junctions of 48 housekeeping genes. Junctions sharing a donor or an acceptor form a group; the junction with the most reads is the constitutive reference and the longer ones skip at least one exon. Per sample the index is the summed skipping reads over all reads in groups carrying at least 20 reads, for samples with at least 500 such reads.

Tissue	n	Global skipping index	Ratio to adrenal (P)	NR5A1 skip index	Ratio to adrenal (P)
Adrenal gland	270	1.08%	—	0.38%	—
Ovary	180	1.25%	1.16 (6.3×10^{-13})	4.27%	11.3 (4.3×10^{-58})
Pituitary	224	1.47%	1.36 (6.4×10^{-58})	0.00%	—
Testis	380	2.54%	2.35 (4.1×10^{-103})	2.86%	7.5 (8.8×10^{-23})

Medians; Mann-Whitney against adrenal. The pituitary median NR5A1 index of zero reflects a small denominator rather than a low level, and pituitary is not compared on that measure.

Dividing each sample's NR5A1 index by its own global index removes whatever the two share. Ovary then stood 9.9-fold above adrenal ($P = 1.5 \times 10^{-57}$) and testis 3.3-fold ($P = 2.4 \times 10^{-15}$). Testis is therefore the tissue whose NR5A1 value is most nearly accounted for by its general splicing behaviour, and ovary the least. The companion analysis [6] normalised to the rate of unannotated junctions in fifteen control genes; the global skipping index used here counts annotated exon-skipping junctions in 48 housekeeping genes, so the two fold values are not comparable.

Within tissues the two indices did not move together in a consistent direction. Adjusted for the NR5A1 denominator, the reads behind the global index and the three quality variables, the rank partial correlation was −0.02 in adrenal ($P = 0.76$), +0.11 in ovary ($P = 0.13$), +0.14 in pituitary ($P = 0.036$) and −0.16 in testis ($P = 0.0019$). Whatever sets the NR5A1 index within a tissue is not the same thing that sets the genome-wide rate.

SRSF6 and CLK4 abundance tracks the index in adrenal gland, pituitary and ovary

Of sixteen splicing regulators tested, three — CLK2, CLK4 and SRSF6 — were associated with the skip index in all four tissues under the primary normalisation, all in the same direction: less of the factor, more skipping. How much of that survives a change of normalisation differs by tissue, and is set out below.

We measured the mRNA abundance of nine kinases, five SR proteins and two hnRNP proteins relative to the summed coverage of ten housekeeping genes, and correlated each with the skip index within tissue. The panel was chosen for the kinases that phosphorylate SR proteins and set their intranuclear distribution [7,8] and for SR proteins whose level changes splice-site choice [9]. Correlations are rank partial correlations adjusted for log junction depth, RNA integrity number, ischaemic time and exonic mapping rate; P comes from the t test of the rank-regression coefficient, and q is the Benjamini-Hochberg value [10] within tissue over the twenty quantities tested.

Gene	Adrenal (n = 270)	Ovary (n = 180)	Pituitary (n = 224)	Testis (n = 376)	Combined P
CLK2	−0.283	−0.233	−0.219	−0.196	1.1×10^{-13}
CLK4	−0.194	−0.264	−0.275	−0.189	3.7×10^{-13}
SRSF6	−0.206	−0.339	−0.324	−0.212	1.9×10^{-17}
CSNK2A1	−0.082	−0.068	−0.225	−0.165	—

All twelve values for the three factors reached q < 0.05 within their tissue, with P from 2.7×10^{-6} to 0.0019; a 10,000-fold permutation gave the same P. The combined column is a signed Stouffer combination over the four tissues, which are separate sets of donors. These are the values under the primary normalisation; how far they depend on it is the subject of the next paragraphs.

How much depends on the normaliser. Abundance above is coverage divided by the summed coverage of ten housekeeping genes. That denominator is a choice, and in two of the four tissues it is itself related to the index. Adjusted for depth and the three quality variables, the housekeeping total rose with the skip index in ovary (r = +0.288, $P = 1.1 \times 10^{-4}$) and in testis (+0.249, $P = 1.1 \times 10^{-6}$), and was flat in adrenal gland (+0.059, P = 0.33) and pituitary (+0.051, P = 0.45). A denominator that tracks the index pushes every ratio built on it in the opposite direction, whether or not the numerator carries anything.

We therefore repeated the twelve tests under four further denominators — the 48 broadly expressed genes of the global-index panel, total reads for the sample, mapped reads, and the exon 6−7 junction itself — and under models that do and do not carry the denominator as a covariate. Total and mapped reads involve no choice of genes.

In ovary every denominator was related to the index: the 48-gene total at +0.320 ($P = 1.5 \times 10^{-5}$), total reads at +0.296 ($P = 6.5 \times 10^{-5}$) and mapped reads at +0.297 ($P = 6.1 \times 10^{-5}$). It is therefore not a property of the ten-gene panel. In ovary, with the exon 6−7 count already in the model, a larger library still carries a higher skip index: the rare skipped junctions are detected disproportionately in deep libraries, and that happens in ovary and not in adrenal gland or pituitary. In testis the gene-based denominators were related to the index (+0.249 and +0.164) while library size was not (+0.049, P = 0.35).

The direction held throughout. Between ten and twelve of the twelve cells were negative under every scheme, and the only cell that turned positive was CLK2 in testis, at +0.02 to +0.07 and never significant. What changed was which cells reached significance.

- In adrenal gland and pituitary no denominator was related to the index, and all three factors held under every denominator and model: adrenal CLK2 −0.25 to −0.33, CLK4 −0.19 to −0.25, SRSF6 −0.21 to −0.27; pituitary CLK2 −0.15 to −0.22, CLK4 −0.22 to −0.27, SRSF6 −0.27 to −0.32, all P ≤ 0.02.
- In ovary, where the denominator has to be carried in the model, SRSF6 held (−0.27 to −0.31, $P \le 3 \times 10^{-4}$) and so did CLK4 (−0.20 to −0.25, P ≤ 0.008), while CLK2 did not (−0.09 to −0.16, P 0.03 to 0.2).

- In testis, with library size as the denominator, none of the three reached significance: CLK2 +0.017 (P = 0.7), CLK4 −0.074 (P = 0.2), SRSF6 −0.041 (P = 0.4). The testis values in the table above rest on a gene-based denominator that is itself related to the index.

We therefore state the result for adrenal gland, pituitary and ovary, and not for testis.

The association did not depend on the share of NR5A1-expressing cells. Adding NR5A1 abundance as a covariate left all twelve correlations negative with P < 0.012, and in adrenal strengthened them (CLK2 −0.283 to −0.340).

Restricting to samples with RNA integrity of at least 7 divided the tissues the same way. In adrenal gland and pituitary the correlations held or grew (adrenal CLK2 −0.362, $P = 3 \times 10^{-7}$; pituitary SRSF6 −0.414, $P = 6 \times 10^{-5}$). In ovary only SRSF6 survived (−0.266, P = 0.005). In testis all three fell to near zero (−0.03 to −0.12, all P ≥ 0.24) in the 107 samples that remained, a fall larger than the loss of power alone would give. Two independent checks — restricting on RNA integrity, and changing the denominator — remove the testis association and leave the other three tissues standing.

CSNK2A1, the catalytic subunit of CK2 and the other target of CX-4945, was not associated in adrenal or ovary. On these data the CLK arm of that pharmacology, not the CK2 arm, is the one that tracks NR5A1 skipping in normal tissue.

The within-tissue direction is opposite to the between-tissue one. Ovary carries 1.19 times the CLK2 of adrenal, 1.88 times the CLK4 and 1.35 times the SRSF6, and also the higher skip index. Factor abundance explains variation inside a tissue; it does not explain the level a tissue sits at.

The association is specific to NR5A1

The three factors do not track genome-wide skipping in a consistent direction, and their correlation with the NR5A1 index survives adjustment for it.

If less CLK2, CLK4 or SRSF6 simply made every weak exon easier to skip, the global skipping index should follow the same direction in every tissue. It did not. The correlation with the global index ran from −0.27 to +0.65 across the twelve cells and changed sign between tissues for all three factors: CLK4 reached +0.646 in testis ($P = 4.1 \times 10^{-45}$) and −0.134 in pituitary, CLK2 −0.273 in ovary and +0.202 in testis, SRSF6 +0.202 in adrenal gland and −0.196 in pituitary.

Gene	Tissue	With the global index	With the NR5A1 index	NR5A1 index, given the global index
CLK2	Adrenal	+0.030 (0.63)	−0.292 (1.4×10^{-6})	−0.298 (8.4×10^{-7})
CLK2	Ovary	−0.273 (2.6×10^{-4})	−0.061 (0.42)	−0.069 (0.37)
CLK2	Pituitary	+0.072 (0.29)	−0.214 (0.0014)	−0.241 (3.2×10^{-4})
CLK2	Testis	+0.202 (8.8×10^{-5})	−0.087 (0.096)	−0.074 (0.16)

Gene	Tissue	With the global index	With the NR5A1 index	NR5A1 index, given the global index
CLK4	Adrenal	+0.115 (0.061)	−0.186 (0.0024)	−0.203 (9.2×10^{-4})
CLK4	Ovary	−0.012 (0.87)	−0.136 (0.073)	−0.136 (0.073)
CLK4	Pituitary	−0.134 (0.048)	−0.273 (4.2×10^{-5})	−0.250 (2.0×10^{-4})
CLK4	Testis	+0.646 (4.1×10^{-45})	−0.097 (0.062)	−0.066 (0.20)
SRSF6	Adrenal	+0.202 (9.4×10^{-4})	−0.208 (6.7×10^{-4})	−0.239 (8.9×10^{-5})
SRSF6	Ovary	+0.084 (0.27)	−0.213 (0.0047)	−0.212 (0.0050)
SRSF6	Pituitary	−0.196 (0.0036)	−0.329 (6.5×10^{-7})	−0.294 (1.0×10^{-5})
SRSF6	Testis	+0.007 (0.90)	−0.112 (0.031)	−0.112 (0.031)

Rank partial correlation (P), adjusted for log junction depth, the log normaliser and the three quality variables; the last column adds the global skipping index and the reads behind it.

Adding the global index left all twelve correlations negative and eight of the twelve at $P < 0.05$. The six adrenal and pituitary cells all kept $P \le 9.2 \times 10^{-4}$; in ovary SRSF6 held (−0.212, $P = 0.0050$) while CLK4 ($P = 0.073$) and CLK2 ($P = 0.37$) did not; in testis only SRSF6 reached significance (−0.112, $P = 0.031$). The association with NR5A1 is therefore not a by-product of a shift in genome-wide exon definition, and the tissues in which it stands are the same ones that survive a change of normaliser.

One qualification limits how far this can be pushed. The 48 genes behind the global index are highly expressed housekeeping genes whose splice sites are mostly strong, whereas the NR5A1 exons bypassed here sit at weak sites. The comparison separates NR5A1 from average exons; it does not separate NR5A1 from weak exons in general.

What does not account for the ovary

The ovary carries the highest index of the four tissues and the explanation for it remains open; this section records what was tested and ruled out.

Within the 180 ovary samples the index was unrelated to sample quality: RNA integrity, ischaemic time and exonic mapping rate all gave $P \ge 0.15$, in contrast to adrenal, where all three were associated. It was also unrelated to NR5A1 abundance itself (−0.043, $P = 0.57$) and to the steroidogenic genes STAR, CYP11A1, CYP17A1 and HSD3B2 (all $q > 0.2$ after adjustment for depth and quality).

NR0B1 (DAX-1), which tracks the index in adrenal, did not do so in ovary (−0.045, $P = 0.56$). NR2F2 (COUP-TFII), the repressor that marks the thecal and interstitial compartment rather than the granulosa compartment, was 7.4-fold more abundant in ovary than in adrenal ($P = 5.7$

$\times 10^{-72}$) but was not associated with the index within ovary once depth and quality were taken into account (−0.124, P = 0.10; −0.063, P = 0.40 with NR5A1 abundance added). The repressor landscape of the ovary differs from that of the adrenal, but neither repressor accounts for the index within it.

Public RNA-seq of purified granulosa-lineage cells does not settle whether those cells carry the ovarian level. Four SRA studies were available, pooled per biological sample and compared with GTEx ovary samples matched on denominator range.

Study	Material	n	Skip index	Depth-matched GTEx ovary	P
SRP048674	purified granulosa cells, 12 donors	12	2.10%	3.94%	0.0035
SRP021912	preovulatory cumulus and mural granulosa	6	4.26%	3.64%	0.80
SRP029207	cumulus cells	5	7.50%	4.32%	0.32
SRP178507	granulosa cells, polycystic ovary syndrome	2	3.48%	4.03%	—

Only the largest study fell below whole ovary; the other three did not differ from it or ran above. The studies used different granulosa subtypes and different library preparations, and three have fewer than seven samples, so they do not support a conclusion either way.

Donor age, menopausal status and follicular activity. None of the three is related to the index. Across the 180 ovary samples the index did not change with donor age: Spearman rho +0.045 unadjusted (P = 0.55), and +0.014 after adjustment for denominator depth, RNA integrity, ischaemic time and exonic mapping rate (P = 0.86, 95% CI −0.13 to +0.16). Grouping donors as premenopausal (under 50, n = 72), transitional (in their fifties, n = 58) and postmenopausal (60 and over, n = 50), a quasi-binomial model of skipped reads out of the denominator, with log denominator as a covariate, gave an odds ratio of 1.08 for the oldest band against the youngest (0.84–1.38, P = 0.55).

The GTEx pathology notes record follicles, follicular cysts, corpora lutea, oocytes, corpora albicantia and atrophy. Classifying each ovary from its note, 31 of 180 carried a record of follicular activity, 47 of corpora albicantia and 80 of atrophy. The index did not differ by any of the three: odds ratio 0.85 for follicular activity (0.65–1.12, P = 0.26), 1.08 for corpora albicantia (P = 0.51) and 0.95 for atrophy (P = 0.62).

The informative comparison is against adrenal. Ovaries with no recorded follicular activity still carried far more skipping than female adrenal samples: odds ratio 4.60 (2.89–7.34, $P = 1.4 \times 10^{-10}$, 149 ovaries against 105 adrenals). Ovaries with recorded follicular activity gave 6.12 (2.87–13.06, $P = 2.9 \times 10^{-6}$), a confidence interval that overlaps the first. Whatever raises the index in ovary is present in ovaries in which the pathologist found no follicle, no corpus luteum and no oocyte.

One check limits how far the gene-expression side of this can be taken. The pathology notes themselves behave as expected with age: the share of ovaries classified as showing follicular activity falls from 55% in the twenties to 43% in the thirties, 28% in the forties, 3% in the fifties and none above seventy. The follicle marker genes do not. AMH, the standard marker of follicular reserve [11], was unrelated to donor age (rho = −0.02, P = 0.76), as were INHA, FOXL2, HSD17B1 and CYP19A1, and none of them separated the ovaries the pathologist called active from the rest (all P > 0.15). Gene-level follicular content is therefore not measurable in these samples: the RNA aliquot and the examined section are different pieces of the same ovary, and a few follicles contribute little RNA against the stroma. The conclusion above rests on the pathology notes, not on the marker genes.

Two transcriptional repressors track the index in the adrenal gland

In adrenal, and in no other tissue, both NR0B1 and NR2F2 fell as the index rose.

NR0B1 (DAX-1) represses SF-1-mediated transactivation [12] and is expressed with SF-1 across steroidogenic tissue. NR2F2 (COUP-TFII) is a second repressor of the same promoters [13,14]. Adjusted for depth, RNA integrity, ischaemic time and exonic mapping rate, NR0B1 gave −0.177 and NR2F2 −0.204 in adrenal, both $q < 0.05$ within tissue. Neither reached significance in ovary, and the direction was inconsistent in pituitary and testis.

The direction is the one that makes a repressor and the skipping alternative ways of limiting SF-1 output: samples with less repressor carried more skipping. These are correlations in bulk tissue and do not establish that either repressor acts on splicing.

Two cautions apply to the repressor result. In pituitary and testis many of the twenty quantities tested moved negatively together, which points to a common factor such as the share of NR5A1-expressing cells rather than to the individual genes; adrenal and ovary did not show that pattern. And NR0B1 is not independent of the zona reticularis markers described next: on its own it gave −0.177 (P = 0.0038) and the reticularis score −0.186 (P = 0.0023), but with each in the other's model they fall to −0.106 (P = 0.084) and −0.121 (P = 0.049). These are two views of one axis, not two findings.

Adrenal zonation. Scoring the three cortical zones from within-tissue standardised log expression — CYP11B2, DACH1, VSNL1 and NR4A2 for the zona glomerulosa, CYP11B1, CYP17A1 and FDX1 for the fasciculata, CYB5A and SULT2A1 for the reticularis [15] — and adjusting for denominator depth, RNA integrity, ischaemic time, exonic mapping rate, donor age and sex, the index fell as the zona reticularis score rose: r = −0.219 ($P = 3.3 \times 10^{-4}$), with SULT2A1 at −0.210 ($P = 6.1 \times 10^{-4}$) and CYB5A at −0.203 ($P = 9.3 \times 10^{-4}$). CYP17A1 gave −0.144 (P = 0.019). The zona glomerulosa score did not reach significance (+0.086, P = 0.17), nor did CYP11B2 (+0.104, P = 0.092) or CYP11B1 (+0.072, P = 0.24), and DACH1 and VSNL1 were flat.

An earlier pass, adjusted for depth alone, had the index rising with the glomerulosa score (+0.223, P = 2.3×10^{-4}). That association was carried by NR4A2 and CYP11B2, both of which track the state of the donor at death rather than the size of a zone. NR4A2 is an immediate-early gene, a class that post-mortem interval is known to alter in GTEx tissue [16], and across the 270 adrenal samples it rose with the Hardy death classification (rho = +0.573, P = 5.6×10^{-25}), with ischaemic time (+0.502) and with the autolysis score (+0.446), and fell with RNA integrity (−0.479). With death classification and autolysis score added to the model the glomerulosa score falls to +0.074 (P = 0.23), CYP11B2 to +0.091 (P = 0.14) and NR4A2 to +0.122 (P = 0.048). We therefore report the reticularis result and not the glomerulosa one.

NR5A1 abundance itself behaves the same way. Against the index it gave +0.183 with depth alone (P = 0.0026) and +0.027 once the quality variables were included (P = 0.66). An analysis of this locus that does not carry the GTEx handling variables will find associations that are not there.

The adrenal result also has to be read against the sample-handling finding of the companion analysis: in adrenal alone, the index rose with ischaemic time and fell with RNA integrity. The repressor correlations above survive adjustment for those variables, but adrenal is the tissue in which handling matters most.

No evidence that the skipping fine-tunes steroidogenesis through SF-1 dosage

The amount of skipping present in normal tissue is too small to move SF-1 dosage appreciably, and the associations that are present do not behave as a dosage effect would.

The index s is skipping reads over exon 6−7 reads, so the full-length share of NR5A1 transcripts is 1/(1 + s). A rise of one percentage point in s therefore lowers functional SF-1 by at most about 1%, and if a steroidogenic gene followed SF-1 dosage alone it could fall by at most about 1% per point. We fitted, within each tissue, log(target) on the index with log NR5A1 abundance, log denominator and the three quality variables, using HC3 standard errors, and read the slope as a percentage change per percentage point.

In ovary every slope lay within about one percentage point of zero and none reached significance; the confidence intervals exclude effects beyond about −3% per point for STAR, CYP11A1 and CYP21A2. HSD17B1 came closest to the dosage bound at −0.9% (P = 0.055) and went to zero when the three splicing factors were added.

In adrenal the slopes were large and mostly positive, the wrong sign for a dosage effect, and they collapsed when CLK2, CLK4 and SRSF6 entered the model.

Target	Slope per point	95% CI	P	With CLK2, CLK4, SRSF6
CYP21A2	+12.9%	+3.1 to +23.7	0.0091	+5.7% (0.25)
MC2R	+23.8%	−2.0 to +56.5	0.074	+3.8% (0.76)
CYP19A1	+21.0%	−6.4 to +56.4	0.15	−3.4% (0.80)
STAR	+9.5%	−0.5 to +20.5	0.063	+1.4% (0.77)
HSD17B1	−6.1%	−11.2 to −0.6	0.030	−0.2% (0.94)

Target	Slope per point	95% CI	P	With CLK2, CLK4, SRSF6
CYP17A1	−5.2%	−12.7 to +2.8	0.20	−10.5% (0.010)

The parsimonious reading is that the index and these genes share an upstream state captured by the three factors, rather than that the index acts on them. CYP17A1 is the exception: it grew stronger on adjustment. That is one gene of ten with no correction for multiple testing and is reported as exploratory.

Testis was the one tissue where associations exceeded the dosage bound and survived adjustment: CYP17A1 −3.0% per point ($q = 0.002$), CYP19A1 −5.6% ($q < 0.001$) and MC2R −6.5% ($q = 0.002$), still −2.5% to −4.1% with the three factors in the model. A truncated protein acting against full-length SF-1 would fit that pattern, but so would variation in Leydig-cell content, and the testis correlations with the three factors themselves did not survive restriction to high-quality samples. We do not interpret it further.

In pituitary the steroidogenic targets are barely expressed and the canonical SF-1 targets of that tissue were not examined, so pituitary is not interpreted here.

An independent dataset reproduces the adrenal level and shows the junctions in H295R

Searching the whole public SRA for NR5A1 coverage returned 1,436 runs, among them 64 of the adrenocortical carcinoma line NCI-H295R deposited by an unrelated group, which carry the skipping junctions at four times the level of depth-matched adrenal tissue.

We queried the Snaptron srav3h compilation, which holds junction summaries for about 316,000 public human runs processed through the same pipeline as the GTEx and TCGA data used above, and kept every run with at least 20 reads on exon 6→7.

Adrenal tissue in that set gave a pooled index of 0.36% with 69% of runs carrying no skipping read, against 0.57% and 28% for GTEx adrenal. A second processing of different donors therefore returns the same low adrenal level, which is the clearest external check available on the measurement itself.

SRP124954, 64 runs of NCI-H295R with knockdown and scrambled controls deposited for a study of nicotinamide nucleotide transhydrogenase, gave a median index of 3.38% and a pooled index of 3.62%, with only 2 of 64 runs carrying no skipping read.

Source	n	Denominator	Median	Pooled	P vs H295R
NCI-H295R (SRP124954)	64	62–178	3.38%	3.62%	—
GTEx adrenal, same range	24	62–178	0.62%	0.90%	6.5×10^{-8}
TCGA adrenocortic	23	62–178	0.86%	2.06%	0.00016

Source	n	Denominator	Median	Pooled	P vs H295R
al carcinoma, same range					

The junctions that were first recovered from this cell line under CX-4945 treatment are thus present in it at baseline, in data generated by another laboratory for an unrelated purpose. The comparison is between a proliferating cell line and bulk tissue and cannot be read as a statement about carcinoma in a patient.

The search found no sex cord-stromal tumour. Runs whose metadata carried granulosa, Sertoli or Leydig wording were normal cells or spermatogenesis studies, and the published RNA-seq of adult granulosa cell tumours is either held under controlled access or was generated by 3′-end sequencing, which does not report junctions. Whether these junctions occur in sex cord-stromal tumours cannot be answered from public data.

Discussion

Two separate things set the NR5A1 skip index, and this work identifies one of them.

The level a tissue sits at is not explained here. Ovary carries eleven times the adrenal index while skipping exons genome-wide only 1.16 times as often, and it carries more CLK2, CLK4 and SRSF6 than adrenal, not less. Whatever fixes the tissue level runs opposite to the factor abundance that governs variation inside a tissue, and remains unidentified.

The variation within a tissue does track those three factors, in the direction the pharmacology predicts. CX-4945 inhibits both CK2 and the CLK family and induces these junctions in H295R cells [4,5]. In normal tissue the CLK arm is the one that shows: CLK4 is associated in adrenal gland, pituitary and ovary and CLK2 in adrenal gland and pituitary, SRSF6 is a CLK substrate [7] and the most robust of the three, and CSNK2A1 is not associated in adrenal or ovary. A reasonable reading is that the phosphorylation state of SR proteins, which RNA-seq cannot measure directly, sets how often these weak exons are bypassed, and that the abundance of the kinases and of SRSF6 is a partial proxy for it.

The effect looks specific rather than genome-wide. The three factors do not track global exon skipping in a consistent direction across tissues, and their association with NR5A1 survives adjustment for it. What this does not establish is specificity to NR5A1 as against weak splice sites in general, because the panel behind the global index is made of strong-site housekeeping genes. Scoring the bypassed sites and building the global index from weak sites only would settle it and has not been done.

On function, the arithmetic is the main constraint. At 0.4% in adrenal and 4% in ovary, this splicing cannot be a dosage lever: a one-point rise removes about 1% of full-length transcript. The associations observed in adrenal are larger than that, mostly positive, and disappear on adjusting for the three factors, which places the index and the steroidogenic genes downstream of a shared state rather than in series. In normal tissue the index reads more like an output of the splicing environment than an input to the steroidogenic programme. That is consistent with the pharmacology without contradicting it: what a drug does by driving skipping to a high level says nothing about whether the low level found in normal tissue does anything.

The adrenal is the tissue that holds this splicing lowest, carries the largest NR5A1 denominator, and is the one where two SF-1 repressors track the index. Why the adrenal should hold it low, and what releases it in carcinoma, is not answered by these data.

Limitations

Every result here is a correlation in bulk post-mortem tissue, and none of the associations establishes direction.

The samples are bulk tissue, so an association between a factor and the index may reflect the proportions of cell types rather than regulation inside any cell. Adding NR5A1 abundance as a covariate, which is the best available proxy for the share of NR5A1-expressing cells, did not remove the three factor associations, but it is an imperfect proxy.

mRNA abundance stands in for protein activity. The layer at which SR protein function is actually set, phosphorylation, is invisible to RNA-seq, so a factor with no abundance association may still matter. SR protein genes also autoregulate through unproductive splicing [17,18], so their mRNA level is itself a splicing outcome and not only a cause of one.

Abundance here is a ratio, and the choice of denominator matters. In ovary and testis every denominator available in these data — a ten-gene or a 48-gene housekeeping total, total reads, mapped reads — is itself related to the skip index, so in those two tissues the factor associations cannot be separated from library composition as cleanly as in adrenal gland and pituitary, where no denominator is. We give the range across five denominators rather than a single figure. A spike-in or an absolute quantification would settle what a ratio cannot.

The testis associations with CLK2, CLK4 and SRSF6 fell to near zero when samples were restricted to RNA integrity of at least 7, by more than the loss of power accounts for. The four-tissue consistency therefore rests mainly on adrenal, ovary and pituitary.

In pituitary and testis many of the twenty quantities tested moved in the same direction, which is the signature of a shared factor such as cell composition rather than of individual genes. Testis in particular carries an unusually complex transcriptome, much of it contributed by the germ-cell lineage [19]. Gene-level interpretation in those two tissues should be withheld.

Many tests were performed. Benjamini-Hochberg correction was applied within tissue across the twenty quantities of the splicing-factor panel and within tissue across the steroidogenic targets, but not across the whole manuscript. The CYP17A1 result in adrenal and the testis steroidogenic associations are reported as exploratory.

The comparison with public SRA runs is across studies with different library preparations and depths. Depth was matched where a comparison is made, but library chemistry was not, and the H295R comparison sets a proliferating adrenocortical carcinoma line [20,21] against bulk tissue.

Methods

Data sources. Exon-exon junction counts and gene-level coverage were obtained from Snaptron [22], which serves the recount3 reprocessing [23]: compilation gtexv2 for GTEx [24] and srav3h for the public SRA [25], both on GRCh38. Sample quality variables (SMRIN,

SMTSISCH, SMEXNCRT) are GTEx sample attributes. Exon coordinates are those of the canonical NR5A1 transcript ENST00000373588.

Skip index. Six junctions share the exon 7 acceptor at chr9:124,483,006. The skip index is the summed coverage of exon 2−7, exon 3−7 and exon 4−7 divided by the coverage of exon 6−7, computed per sample, with samples carrying fewer than 20 reads on the denominator excluded. Junctions were matched to the expected coordinates with a tolerance of 2 bp.

Duplicate removal. Some GTEx samples are listed in Snaptron under more than one rail identifier, with external identifiers differing only in a trailing version number and identical junction counts. One identifier was kept per sample. The analyses here use 270 adrenal, 180 ovary, 224 pituitary and 380 testis samples.

Global skipping index. Annotated junctions of 48 housekeeping genes were grouped by shared donor or shared acceptor. The reference is the junction with the most reads; the remaining junctions of the group skip at least one exon. Per sample the index is the summed non-reference reads divided by all reads, over groups carrying at least 20 reads, for samples with at least 500 such reads. A variant taking the shortest junction as reference gave medians of 37–56%, indicating that the shortest junction is often not the constitutive one, and is not used.

Gene-level abundance. Coverage from the Snaptron genes endpoint was divided, per sample, by the summed coverage of ACTB, GAPDH, PGK1, SDHA, YWHAZ, PPIA, RPL13A, TBP, HPRT1 and B2M; samples with a sum below 1000 were excluded.

Normalisation sensitivity. The twelve factor-by-tissue tests were repeated under five denominators: the ten housekeeping genes used throughout, the 48 broadly expressed genes of the global-index panel, total reads and mapped reads from the GTEx sample annotation (SMRDTTL and SMMPPD), and the exon 6−7 junction count. Each was run in three forms: the ratio of coverage to the denominator; the same ratio with the log denominator added as a covariate; and the raw log coverage with the log denominator as a covariate. Each denominator was also tested against the skip index itself, within tissue and adjusted for depth and the three quality variables, to identify denominators that cannot be trusted as such. Total and mapped reads correlated at +0.98 to +1.00 within every tissue and are not independent of each other.

Statistics. Associations within tissue are rank partial correlations: ranks of the index and of the variable are regressed on the ranks of the covariates, and the correlation of the residuals is reported. P comes from the t test of the rank-regression coefficient on n − 2 − k degrees of freedom, not from a normal approximation. Covariates are log denominator, RIN, ischaemic time and exonic mapping rate unless otherwise stated. Permutation P values use 10,000 permutations of the residualised variable. Combination across tissues is a signed Stouffer combination. Group comparisons use the two-sided Mann-Whitney U test. Benjamini-Hochberg q values are computed within tissue over the quantities of the relevant panel. Dosage models are ordinary least squares of log(target + half the smallest non-zero value) on the index with HC3 standard errors. Analyses were performed in Python with numpy, pandas, scipy and statsmodels.

Availability and declarations

Data availability. All data analysed are public. Junction and gene-level summaries are served by Snaptron (https://snaptron.cs.jhu.edu), compilations gtexv2 and srav3h. The SRA studies used are SRP124954, SRP048674, SRP021912, SRP029207 and SRP178507.

Code availability. Analysis scripts will accompany this preprint as supplementary files; each queries only the public endpoints above and needs no credentials.

Author contributions. K.O. is the sole author: he conceived the study, performed all analyses and wrote the manuscript.

Acknowledgements. The CX-4945 pharmacology that motivates the splicing-factor panel comes from work with Tomoko Tanaka and Toshihiko Yanase, cited here and reported separately.

Funding. Grant-in-Aid for Scientific Research from JSPS (18K09212); the Suisyo Kenkyu Project of Fukuoka University (167005 and 197010); and the Center for Clinical and Translational Research of Kyushu University Hospital (A147). The funders had no role in study design, analysis, interpretation or the decision to publish.

Competing interests. The author declares no competing interests.

Ethics. This study used only publicly available, de-identified data and involved no new human material; no ethical approval was required.

References

1. Luo X, Ikeda Y, Parker KL. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell*. 1994;77(4):481–490.
2. Doghman M, Karpova T, Rodrigues GA, et al. Increased steroidogenic factor-1 dosage triggers adrenocortical cell proliferation and cancer. *Mol Endocrinol*. 2007;21(12):2968–2987.
3. Ferraz-de-Souza B, Lin L, Achermann JC. Steroidogenic factor-1 (SF-1, NR5A1) and human disease. *Mol Cell Endocrinol*. 2011;336(1–2):198–205.
4. Ohe K, Tanaka T, Horita Y, et al. Circular IRE-type RNAs of the NR5A1 gene are formed in adrenocortical cells. *Biochem Biophys Res Commun*. 2019;512(1):1–6.
5. Siddiqui-Jain A, Drygin D, Streiner N, et al. CX-4945, an orally bioavailable selective inhibitor of protein kinase CK2, inhibits prosurvival and angiogenic signaling and exhibits antitumor efficacy. *Cancer Res*. 2010;70(24):10288–10298.
6. Ohe K, Tanaka T, Yanase T. A junction-level skip index for NR5A1 distinguishes adrenocortical carcinoma from normal adrenal. Preprint, 2026.
7. Colwill K, Pawson T, Andrews B, et al. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J*. 1996;15(2):265–275.
8. Gui JF, Lane WS, Fu XD. A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature*. 1994;369(6482):678–682.
9. Jensen MA, Wilkinson JE, Krainer AR. Splicing factor SRSF6 promotes hyperplasia of sensitized skin. *Nat Struct Mol Biol*. 2014;21(2):189–197.

10. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57(1):289–300.
11. Dewailly D, Andersen CY, Balen A, et al. The physiology and clinical utility of anti-Müllerian hormone in women. *Hum Reprod Update*. 2014;20(3):370–385.
12. Ito M, Yu R, Jameson JL. DAX-1 inhibits SF-1-mediated transactivation via a carboxy-terminal domain that is deleted in adrenal hypoplasia congenita. *Mol Cell Biol*. 1997;17(3):1476–1483.
13. Buholzer CF, Arrighi JF, Abraham S, et al. Chicken ovalbumin upstream promoter-transcription factor is a negative regulator of steroidogenesis in bovine adrenal glomerulosa cells. *Mol Endocrinol*. 2005;19(1):65–75.
14. van den Driesche S, Walker M, McKinnell C, et al. Proposed role for COUP-TFII in regulating fetal Leydig cell steroidogenesis, perturbation of which leads to masculinization disorders in rodents. *PLoS One*. 2012;7(5):e37064.
15. Nishimoto K, Nakagawa K, Li D, et al. Adrenocortical zonation in humans under normal and pathological conditions. *J Clin Endocrinol Metab*. 2010;95(5):2296–2305.
16. Ferreira PG, Muñoz-Aguirre M, Reverter F, et al. The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nat Commun*. 2018;9(1):490.
17. Lareau LF, Inada M, Green RE, et al. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*. 2007;446(7138):926–929.
18. Ni JZ, Grate L, Donohue JP, et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing. *Genes Dev*. 2007;21(6):708–718.
19. Soumillon M, Necsulea A, Weier M, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep*. 2013;3(6):2179–2190.
20. Gazdar AF, Oie HK, Shackleton CH, et al. Establishment and characterization of a human adrenocortical carcinoma cell line that expresses multiple pathways of steroid biosynthesis. *Cancer Res*. 1990;50(17):5488–5496.
21. Rainey WE, Bird IM, Mason JI. The NCI-H295 cell line: a pluripotent model for human adrenocortical studies. *Mol Cell Endocrinol*. 1994;100(1–2):45–50.
22. Wilks C, Gaddipati P, Nellore A, et al. Snaptron: querying splicing patterns across tens of thousands of RNA-seq samples. *Bioinformatics*. 2018;34(1):114–116.
23. Wilks C, Zheng SC, Chen FY, et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. *Genome Biol*. 2021;22(1):323.
24. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318–1330.
25. Katz K, Shutov O, Lapoint R, et al. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res*. 2022;50(D1):D387–D390.

References are numbered in order of first citation. Every record except number 6 was checked against its PubMed entry or DOI. Number 6 is the companion analysis, in submission at the time of writing; its DOI will be added once it is posted.

Figures

Four figures accompany this manuscript. Tissues carry the same colour throughout: adrenal gland blue, ovary green, pituitary orange, testis purple. The four hues are taken from the Okabe-Ito set and every adjacent pair clears a colour-vision-deficiency separation check, but tissue is also named on every axis, so no panel depends on colour alone.

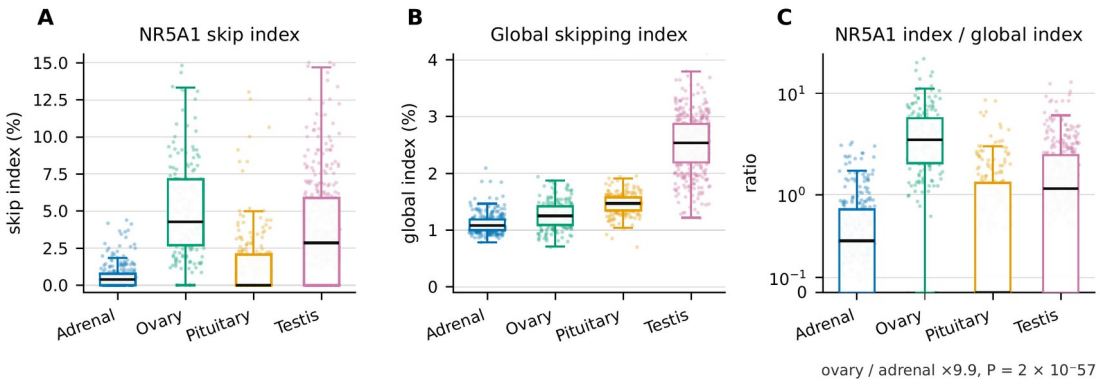


Figure 1. The tissue difference is specific to NR5A1. Each point is one GTEx sample; boxes show the median and interquartile range and whiskers extend to 1.5 × IQR, outliers not drawn. (A) NR5A1 skip index. (B) Global skipping index, the share of junction reads falling on non-reference junctions across 48 broadly expressed genes. (C) The per-sample ratio of the two, on a symmetric log scale. Ovary carries 11.3 times the NR5A1 index of adrenal gland but only 1.16 times the global index, and 9.9 times the ratio (Mann-Whitney P = 1.5×10^{-57}).

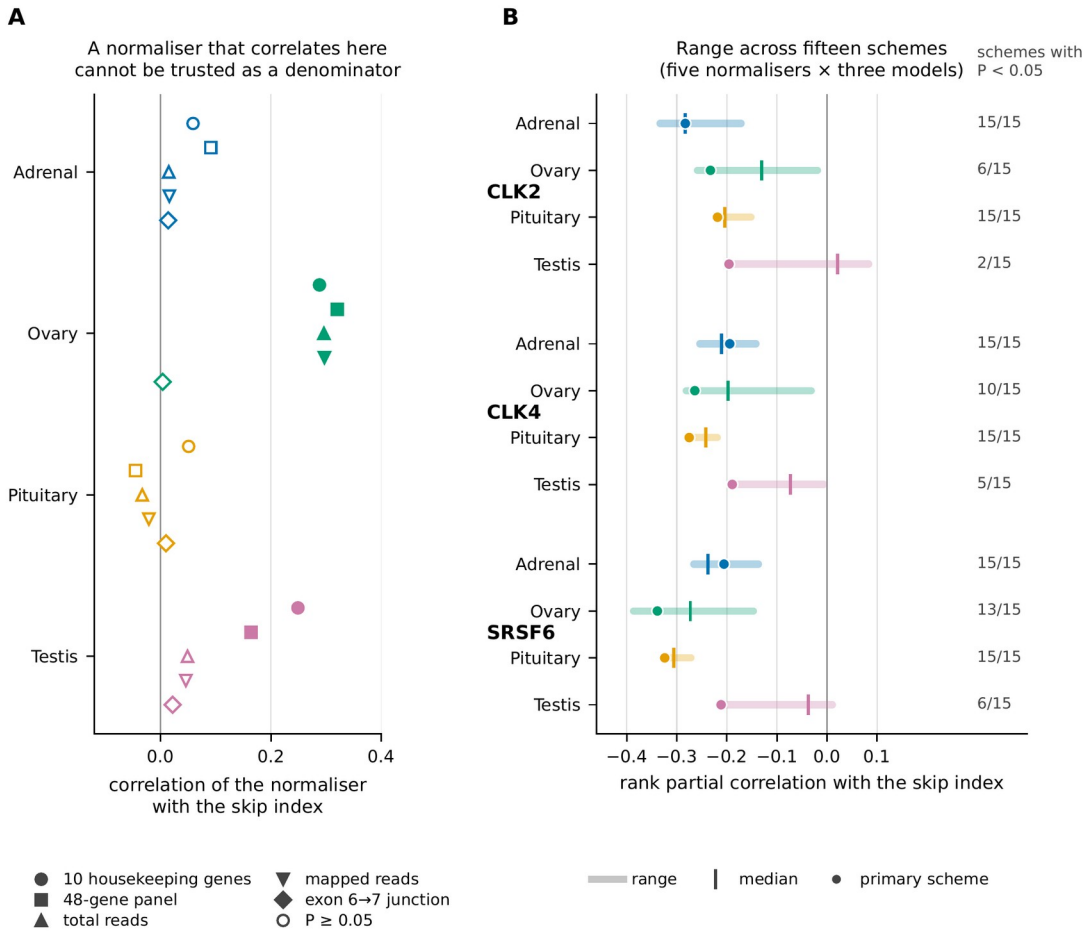


Figure 2. How much the factor associations depend on the normaliser. (A) Each candidate normaliser tested against the skip index itself, within tissue, adjusted for log junction depth and the three quality variables. Filled markers are P < 0.05. A normaliser that correlates here cannot serve as a denominator for an abundance measure, because the ratio will move opposite to the index whatever its numerator does. In ovary all four gene-based and library-size normalisers correlate; in testis the two gene-based ones do and library size does not; in adrenal gland and pituitary none does. (B) The twelve factor-by-tissue cells. The bar spans the range of the rank partial correlation across fifteen schemes (five normalisers × three models), the tick is the median and the filled circle the primary scheme reported in the table. The figure at right counts the schemes reaching P < 0.05. Adrenal gland and pituitary reach 15 of 15 for all three factors; ovary reaches 13 of 15 for SRSF6 and 10 of 15 for CLK4; testis reaches at most 6 of 15.

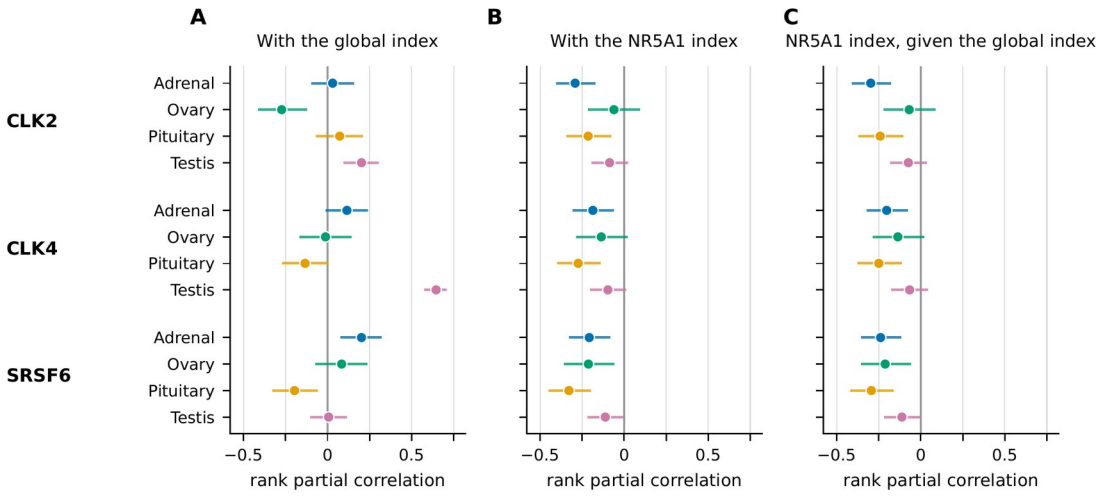


Figure 3. The association with NR5A1 is not a by-product of genome-wide skipping. Rank partial correlations with 95% confidence intervals, adjusted for log junction depth, the log normaliser and the three quality variables. (A) Each factor against the global skipping index. The sign changes between tissues for all three factors, so the factors do not move genome-wide skipping in a consistent direction. (B) Each factor against the NR5A1 skip index. (C) The same with the global skipping index added to the model. All twelve correlations remain negative and eight reach $P < 0.05$, the six adrenal and pituitary cells among them.

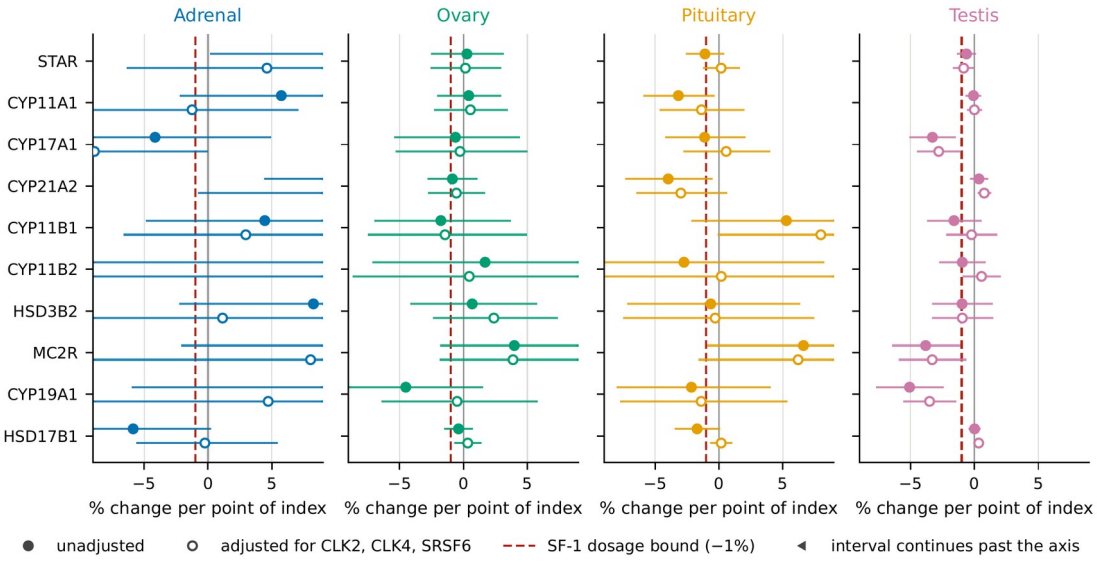


Figure 4. The skipping is too rare to set steroidogenic output through SF-1 dosage, except in testis. Percentage change in each steroidogenic transcript per percentage point of skip index, from ordinary least squares on log abundance with heteroskedasticity-consistent standard errors, with 95% confidence intervals. Filled circles are unadjusted, open circles adjusted for CLK2, CLK4 and SRSF6. The dashed line marks −1% per point, the most that loss of full-length SF-1 can account for when a skip index of s leaves a full-length share of 1/(1 + s). Triangles mark intervals that continue past the axis. Only in testis do estimates fall beyond the bound with intervals excluding zero, and only there do they survive adjustment: CYP17A1 −2.8%, MC2R −3.3% and CYP19A1 −3.5% per point.

Supplementary Files

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

- OheNR5A1fourtissuesupplementary.zip