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.

A junction-level skip index for NR5A1 distinguishes adrenocortical carcinoma from normal adrenal

Kenji Ohe

oekenji@gmail.com

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

Tomoko Tanaka

Fukuoka University

Toshihiko Yanase

Seiwakai Muta Hospital

Research Article

Keywords: NR5A1, SF-1, adrenocortical carcinoma, alternative splicing, exon skipping, splice junction, GTEx, TCGA, RNA-seq, steroidogenesis

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11681941/v1

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 Department of Bioregulatory Science of Life-Related Diseases of Fukuoka University received an unrestricted donation from MSD K.K.; this had no role in the design, execution, analysis or interpretation of the study. The authors declare no other competing interests.

A junction-level skip index for NR5A1 distinguishes adrenocortical carcinoma from normal adrenal

Kenji Ohe [1], Tomoko Tanaka [2], Toshihiko Yanase [3]

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

[2] Department of Regenerative Medicine and Transplantation, Faculty of Medicine, Fukuoka University, 7-45-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan

[3] Seiwa-kai, Muta Hospital, 3-9-1 Hoshikuma, Sawara-ku, Fukuoka 814-0163, Japan

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

One-sentence summary: in 270 normal adrenal samples (GTEx) and 73 adrenocortical carcinomas (TCGA-ACC), the proportion of NR5A1 transcripts joining an upstream exon directly to exon 7 rose from 0.38% to 0.90% (background-normalised $p = 1.1 \times 10^{-6}$), was not explained by sequencing depth, genome-wide splicing background or tumour purity, and correlated inversely with steroidogenic gene expression within tumours (rho = −0.33, p = 0.004); it did not correlate with Weiss grade.

Abstract

Steroidogenic factor 1 (SF-1, *NR5A1*) sets the transcriptional capacity of the adrenal cortex. Its 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 receptor. We asked how often this happens in human tissue, using public RNA-seq rather than new material.

We defined a skip index as the summed coverage of the exon 2→7, exon 3→7 and exon 4→7 junctions divided by the coverage of the exon 6→7 junction, computed per sample from Snaptron junction counts. Among *NR5A1*-expressing tissues in GTEx, the index was highest in ovary (median 4.27%, n = 180) and lowest in adrenal gland (0.38%, n = 270), which carried the largest denominator of any tissue examined. Most of the elevation in testis was accounted for by a genome-wide splicing background six times that of adrenal: normalised to that background, testis stood 1.3-fold above adrenal, against 7.0-fold for ovary.

In TCGA-ACC the index was 0.90% (n = 73), higher than normal adrenal ($p = 5.3 \times 10^{-5}$; background-normalised $p = 1.1 \times 10^{-6}$), and was not explained by read depth, splicing background or tumour purity. Within tumours it correlated inversely with steroidogenic gene expression (rho = −0.33, p = 0.004) but not with Weiss grade (rho = 0.14, p = 0.30).

Introduction

SF-1 is required for adrenal development and for expression of the rate-limiting steroidogenic enzymes, and its dosage drives adrenocortical proliferation: increasing SF-1

gene dosage is sufficient to trigger adrenocortical cell proliferation and tumour formation (Doghman et al., 2007). In adrenocortical carcinoma (ACC) SF-1 is a lineage marker of high diagnostic value, detected in essentially all adrenocortical tumours, and the strength of its expression is prognostic independently of stage (Sbiera et al., 2010). Amplification and overexpression of the gene have been reported chiefly in childhood adrenocortical tumours (Figueiredo et al., 2005; Pianovski et al., 2006).

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. Two such products have been recovered and sequenced from adrenocortical carcinoma cells treated with the CK2/CLK inhibitor CX-4945: exon 2→7, which is out of frame and terminates after 35 residues, and exon 3→7, which is in frame and deletes most of the ligand-binding domain. Neither encodes a full-length receptor. A third junction, exon 4→7, which is also out of frame, is reported in the transcriptome analysis of the same cells (Tanaka et al., 2019; manuscripts in preparation).

Whether this splicing occurs in human tissue outside that pharmacological setting, and whether it differs between normal adrenal and carcinoma, has not been examined. Public RNA-seq compilations index exon-exon junctions at per-sample resolution across tens of thousands of libraries, which makes the question answerable without new material. We asked three things: how the junctions are distributed across *NR5A1*-expressing tissues, whether they differ between normal adrenal and ACC, and whether their abundance relates to the steroidogenic output of the tumour.

Five junctions join an upstream exon to exon 7, and all abolish the full-length receptor

NR5A1 lies on the minus strand of chromosome 9 and its canonical transcript (ENST00000373588) has seven exons. Querying every junction overlapping the locus in the GTEx compilation returned 805 junctions, of which six share the exon 7 acceptor: the constitutive exon 6→7 junction and five junctions from exons 1 to 5 (Fig. 1). All six carry CT–AC dinucleotides, the minus-strand form of the canonical GT–AG motif, so all are spliceosomal.

Junction	Exons removed	Coding nucleotides removed	Reading frame	GENCODE annotated	GTEx samples	GTEx reads
exon 1→7	2–6	1,153	frameshift	no	30	41
exon 2→7	3–6	1,036	frameshift	no	223	352
exon 3→7	4–6	894	in frame	no	192	285
exon 4→7	5–6	268	frameshift	yes	892	2,083

Junction	Exons removed	Coding nucleotides removed	Reading frame	GENCODE annotated	GTEx samples	GTEx reads
exon 5→7	6	148	frameshift	no	187	252
exon 6→7	—	—	constitutive	yes	2,633	245,638

None of these junctions encodes a full-length receptor: four shift the reading frame and the fifth, exon 3→7, is in frame but deletes the Ftz-F1 box, the hinge and the N-terminal two thirds of the ligand-binding domain.

Annotation status does not track that consequence. The exon 4→7 junction is annotated — it appears in one of the nineteen *NR5A1* transcript models, ENST00001112754 — yet it removes 268 coding nucleotides and shifts the frame. Annotation records that a junction has been observed and assigned to a transcript model, which is a bibliographic fact rather than a functional one. We therefore defined the index by the functional consequence instead, as the junctions that join directly to exon 7 and remove at least two internal exons: exon 2→7, exon 3→7 and exon 4→7, divided by exon 6→7. The exon 5→7 junction removes a single exon and falls outside a definition built on consecutive skipping; exon 1→7 satisfies the definition but contributes 41 reads across 30 samples in all of GTEx, and including it changes no result reported here. Results for the narrower definition exon 2→7 + exon 3→7 are given as a sensitivity analysis throughout.

The index is lowest in adrenal, which carries the largest denominator

The index is a ratio and is undefined where *NR5A1* is not transcribed, so we restricted the analysis to tissues that express it and required at least 20 reads on the exon 6→7 junction. The threshold behaves as a specificity filter: of samples with at least one read on the exon 6→7 junction, 128 blood and 96 skin samples passed at a threshold of 1 but none and one respectively at 20, and brain fell from 347 to 7, consistent with *NR5A1* expression in the hypothalamic ventromedial nucleus being diluted in bulk tissue. Brain was excluded on that basis.

Tissue	n	Median index (%)	Mean (%)	Fraction > 0	Median denominator (reads)
Ovary	180	4.27	5.31	0.97	81
Testis	380	2.86	3.90	0.70	41
Spleen	255	0.40	0.49	0.65	313
Adrenal gland	270	0.38	0.63	0.72	347
Pituitary	224	0.00	1.32	0.46	52

Adrenal carried the largest denominator of any tissue examined, four times that of ovary and eight times that of testis, and the lowest index of the two gonadal tissues (Fig. 2A, B). Low detection in adrenal is therefore not a consequence of low *NR5A1* expression.

Detection depends on depth, so tissues were compared in a logistic model with log(denominator) as a covariate (coefficient 0.94, $p = 1.6 \times 10^{-11}$). Relative to adrenal, ovary was higher (coefficient 3.85, $p = 3.6 \times 10^{-15}$) and testis was higher (1.86, $p = 4.5 \times 10^{-8}$); pituitary (0.57, p = 0.07) and spleen (−0.23, p = 0.24) did not differ. Under the narrower definition the ovary and testis coefficients were 3.08 and 1.71, so the ordering does not depend on which junctions enter the numerator. Restricting to samples with denominators between 40 and 160, which removes the inflation that a small denominator produces in any ratio, left the same ordering (ovary 0.69 detected, testis 0.34, adrenal 0.19).

Genome-wide splicing background accounts for most of the testis signal

Testis is permissive for transcription and splicing across the genome, so an elevated index there need not be specific to *NR5A1*. We measured each sample's background as the fraction of junction reads falling on unannotated junctions across fifteen broadly expressed control genes, a quantity that has nothing to do with the *NR5A1* locus.

Tissue	Background (%)	*NR5A1* index (%)	Index / background (median per sample)
Ovary	0.081	4.27	49.2
Adrenal gland	0.051	0.38	7.0
Spleen	0.061	0.40	5.6
Testis	0.324	2.86	9.2
Pituitary	0.076	0.00	0.0

The testis background was six times that of adrenal (Fig. 2C). Expressed relative to each sample's own background, testis stood at 9.2 against 7.0 for adrenal (Mann-Whitney P = 0.006), whereas ovary stood at 49.2 ($P = 6.6 \times 10^{-55}$). Background therefore reduces the testis elevation from 7.5-fold to 1.3-fold but does not remove it, and we do not interpret the residual difference as specific to *NR5A1*. Ovary is not explained this way.

Adding log(background) as a covariate to the tissue model did not reduce the testis coefficient (1.86 to 1.87) and the covariate itself was not significant (p = 0.99). This is expected rather than contradictory: background is close to a tissue-level constant, so a model containing both cannot separate them, and the covariate carries only within-tissue variation. The ratio to background above is the informative comparison.

The index is higher in adrenocortical carcinoma than in normal adrenal

The same six junctions were quantified in TCGA-ACC. As in GTEx, exon 4→7 was the most abundant (51 samples, 268 reads), followed by exon 3→7 (26, 77) and exon 2→7 (26, 69), with exon 1→7 negligible (2, 3).

Aggregated per patient and applying the same denominator threshold, the index was 0.90% in 73 carcinomas against 0.38% in 270 normal adrenal samples (Mann-Whitney $p = 5.3 \times 10^{-5}$). Normalised to each sample's own splicing background the separation widened (7.0 against 20.8, $p = 1.1 \times 10^{-6}$). The two groups differ in magnitude rather than in detectability (Fig. 3): 0.72 of normal and 0.77 of tumour samples had a detectable junction, and group membership was not significant in a logistic model of detection ($p = 0.15$).

The comparison held across both numerator definitions and all three denominator thresholds tested.

Definition	Threshold	Normal median (%)	ACC median (%)	p	Background-normalised p
exon 2→7 + 3→7	10	0.00	0.00	6.0×10^{-5}	1.6×10^{-5}
exon 2→7 + 3→7	20	0.00	0.00	3.1×10^{-5}	7.6×10^{-6}
exon 2→7 + 3→7	50	0.00	0.00	1.4×10^{-5}	3.2×10^{-6}
+ exon 4→7	10	0.38	0.86	1.8×10^{-4}	5.1×10^{-6}
+ exon 4→7	20	0.38	0.90	5.3×10^{-5}	1.1×10^{-6}
+ exon 4→7	50	0.38	0.90	4.9×10^{-5}	9.5×10^{-7}

Three alternative explanations were examined and none accounts for the difference.

Splicing background. The median background was 0.051% in normal adrenal and 0.049% in ACC. The two differ ($p = 0.015$), but in the direction opposite to the one that would explain the result: background is lower in the carcinomas, which carry the higher skip index. A general increase in aberrant splicing, or a systematic difference in how junctions are called between the two cohorts, would have to act in the same direction as the finding to account for it, and it does not. Dividing each sample's index by its own background accordingly widens the separation rather than narrowing it.

Read depth. The index is a ratio of two junctions in the same gene, the same library and the same coverage profile, so depth is expected to act on numerator and denominator alike. Within GTEx adrenal the index fell slightly with denominator depth ($\rho = -0.13$, $P = 0.03$), and median denominators were comparable between the groups (347 and 284), so depth cannot produce the difference.

Tumour purity. ABSOLUTE purity was available for 71 of 73 tumours, with a median of 0.87. The index did not correlate with purity (rho = 0.08, p = 0.49), and the regression slope was not different from zero (p = 0.57). Because normal adrenal carries a *lower* index than tumour, any admixture of normal tissue moves a tumour sample towards the normal value; contamination is therefore conservative with respect to the comparison made here.

Post-mortem versus surgical tissue. Normal adrenal in GTEx is obtained post mortem and tumour tissue in TCGA at surgery, so agonal stress differs between the groups. Within GTEx adrenal the index rose with the expression of nine immediate-early genes (*FOS*, *FOSB*, *JUN*, *JUNB*, *EGR1*, *NR4A1*, *NR4A2*, *NR4A3* and *DUSP1*; composite $\rho = 0.33$, $P = 2.4 \times 10^{-8}$, adjusted for denominator depth, donor age and sex), and donor age and sex were themselves unrelated to the index ($\rho = -0.02$ for age and -0.03 for male sex). The same association was absent within TCGA-ACC (composite $\rho = 0.05$, $P = 0.68$, 73 patients), with a confidence interval that excludes the GTEx estimate. Sample quality pointed the same way. Within GTEx adrenal, adjusting for denominator depth, the index rose with ischaemic time ($\rho = +0.31$, $P = 2.2 \times 10^{-7}$) and fell with RNA integrity number ($\rho = -0.26$, $P = 1.3 \times 10^{-5}$) and with the exonic mapping rate ($\rho = -0.25$, $P = 3.4 \times 10^{-5}$); none of the three was related to the index in ovary, pituitary or testis (all $P \geq 0.15$). These measures were not available to us for TCGA-ACC, and whether they overlap with the immediate-early gene association was not tested. Agonal stress and post-mortem degradation therefore act to raise the index in the normal group, and cannot account for the higher index in carcinoma.

The index does not track Weiss grade

Within ACC the index showed no relation to histological grade. Weiss score was available for 55 patients. Spearman correlation with the index as a continuous variable was 0.14 (p = 0.30). Dichotomised at 5.5, the medians were 0.65% (n = 26) and 1.06% (n = 29), Mann-Whitney p = 0.25. In a logistic model of detection the Weiss coefficient was 0.20 (p = 0.22), and adding log(background) left it at 0.18 (p = 0.25).

Under the narrower definition the result was the same in direction and significance (rho = 0.15, p = 0.27; Mann-Whitney p = 0.10).

This null result is not attributable to zero inflation. That was the limitation of the narrower definition, under which both grade groups had a median of zero and 44% of patients had any detectable junction. Including exon 4→7 raised detection to 77% and gave a non-zero median in both groups, which is the condition under which a graded association would be detectable; the correlation did not emerge. With 26 and 29 patients per group the comparison can exclude a difference of roughly 0.8 standard deviations but not a smaller one.

Within tumours, a higher index accompanies lower steroidogenic gene expression

If these junctions remove functional SF-1, tumours carrying more of them should express less of what SF-1 drives. This comparison is made within TCGA-ACC alone, so it is not affected by differences between cohorts.

A steroidogenic score was formed from the standardised log expression of *STAR*, *CYP11A1*, *CYP17A1*, *CYP21A2* and *HSD3B2*, each normalised to ten control genes in the same sample. The index correlated inversely with that score (rho = −0.33, p = 0.004, n = 73).

Gene	rho	p
CYP11A1	−0.357	0.002
CYP21A2	−0.308	0.008
CYP17A1	−0.284	0.015
MC2R	−0.126	0.29
HSD3B2	−0.097	0.41
STAR	0.003	0.98
NR5A1	−0.058	0.62

Three of six steroidogenic genes reached nominal significance (Fig. 4); of these, only *CYP11A1* survives Bonferroni correction for the seven tests shown. The index did not correlate with *NR5A1* expression itself (rho = −0.06), so the association is not a consequence of how much *NR5A1* the tumour transcribes.

The relation survived adjustment. Regressing the steroidogenic score on the index gave a coefficient of −0.093 (p = 0.016); adding *NR5A1* expression, stage and purity individually left it between −0.073 and −0.099 (p = 0.013 to 0.022), and the fully adjusted model gave −0.079 (p = 0.013, n = 69). Dedifferentiation of high-grade tumours is the obvious confounder and does not account for it, consistent with the absence of any relation between the index and Weiss grade.

One post hoc observation deserves recording without being built upon. Resolved by junction, the association follows the reading frame rather than abundance. The two frameshifting junctions correlated with the steroidogenic score (exon 2→7, rho = −0.26, p = 0.027; exon 4→7, rho = −0.31, p = 0.008) while the in-frame exon 3→7 did not (rho = −0.06, p = 0.59), although exon 2→7 and exon 3→7 were detected in exactly the same fraction of patients (0.356), so the difference is not one of statistical power. The in-frame product retains both zinc modules of the DNA-binding domain and the AF-2 motif. This pattern was found after the main analysis and requires independent confirmation; the index used throughout this work was not redefined around it.

Survival: expression is stage-independent, the index is not

These analyses are exploratory. They were run after the comparisons above, on 73 patients with 24 deaths and 34 recurrence events, and the p values are not corrected for the number of tests performed in this study.

NR5A1 expression, normalised to control genes in the same sample, was associated with outcome in every model fitted. Adjusted for stage, the hazard ratio per unit of natural-log expression was 3.17 for death (95% CI 1.50–6.71, P = 0.0026; 71 patients, 23 events) and 3.38 for recurrence (1.93–5.90, P = 1.9×10^{-5}; 67 patients, 33 events). Stage itself carried hazard ratios of 2.7 and 2.4 in these models. These values sit within the range reported for

SF-1 protein and mRNA by Sbiera et al. (2010), whose stage-adjusted hazard ratios were 2.46 for death by immunohistochemistry and 4.69 in an mRNA cohort. The present estimate is an independent cohort measured by a different quantity.

The index behaved differently. It was not associated with overall survival after adjustment for stage (HR 1.08, p = 0.49), and for recurrence the association fell from p = 0.095 unadjusted to p = 0.22 with stage. Expressed as a depth-normalised absolute amount rather than a ratio, it was associated with recurrence alone (HR 1.15, p = 0.014) but not after stage adjustment (p = 0.060). Dichotomised at the median it was not associated with either endpoint (p = 0.42 and p = 0.36). An association appeared only in models that also contained *NR5A1* expression, which is a sign of instability rather than of an independent effect. We therefore record no independent prognostic association for the index.

Refitting the same models with the index on a log scale and with tumour purity added as a covariate gave the same answer. For recurrence the hazard ratio per ten-fold index was 1.56 unadjusted (95% CI 0.95–2.58, P = 0.081; 68 patients, 34 events) and 1.45 adjusted for stage and purity (0.82–2.55, P = 0.20; 65 patients, 32 events); for overall survival it was 1.34 (0.76–2.37, P = 0.32; 73 patients, 24 events) and 1.12 (0.59–2.11, P = 0.74; 69 patients, 22 events). Replacing the index by its rank, which removes any dependence on the shape of its distribution, left both endpoints non-significant (recurrence P = 0.15 unadjusted and P = 0.32 adjusted; overall survival P = 0.37 and P = 0.91). Patients lacking stage or purity were excluded from the adjusted models.

The index does not track splicing-factor levels, global skipping or tumour phenotype

We asked what differs between ACC tumours to produce the observed range of the index. Within TCGA-ACC (73 patients) we correlated the index with depth-normalised expression of 30 genes: seven SR proteins, nine kinases (*CLK1–CLK4*, *SRPK1*, *SRPK2* and three CK2 subunits), six serine/threonine phosphatase subunits, two hnRNPs and six U2 snRNP components. No gene survived Benjamini–Hochberg correction (smallest q = 0.23) (Fig. 5B). The largest correlations were *CSNK2A1* (ρ = +0.29, P = 0.013), *U2AF1* (ρ = −0.27, P = 0.021) and *PPP2CB* (ρ = +0.27, P = 0.023). Mean ρ was positive in all five functional groups (+0.02 to +0.11), so the groups did not separate in opposing directions.

Phosphorylation state is not visible in RNA-seq, so we also used autoregulatory splicing as a proxy for splicing-factor activity. For each of 14 SR protein, CLK and SRPK genes we grouped junctions that share a splice site, took the most abundant junction of each group as the reference, and measured the read fraction carried by the remaining junctions; in SR protein genes this event is dominated by the NMD-targeted poison exons. Thirteen genes had enough reads, and none survived correction (smallest q = 0.13) (Fig. 5A). *SRSF5* (ρ = +0.30, P = 0.010) and *CLK2* (ρ = +0.30, P = 0.022) were the largest, and both kept a positive coefficient after adjustment for stage, tumour purity and *NR5A1* expression (P = 0.0011 and P = 0.026).

The index is also not a reading of global splicing. Across 48 broadly expressed genes we built 474 groups of junctions sharing a splice site and defined a global exon-skipping index

as the read fraction on the non-shortest junction of each group. This index did not rise with the *NR5A1* skip index: the correlation was negative and did not survive adjustment ($\rho = -0.23$, $P = 0.048$; adjusted $P = 0.16$) (Fig. 5C).

Splice-site strength nonetheless governed skipping in these same data. We scored each bypassed splice site with position weight matrices built from the constitutive junctions of the same 48 genes (220 5′ and 220 3′ sites), assigning intron orientation from the GT–AG motif (1321 of 1337 junctions assignable). Skipping fell as site strength rose, at both ends: 5′ sites $\rho = -0.35$ ($P = 2.2 \times 10^{-7}$, 207 groups), 3′ sites $\rho = -0.22$ ($P = 7.7 \times 10^{-4}$, 222 groups), combined $\rho = -0.28$ ($P = 2.4 \times 10^{-9}$, 429 groups) (Fig. 5D). The median skipping fraction was 77.8% for the weakest tertile of sites against 29.7% for the strongest. Stratifying by site strength did not, however, reveal a relation to the *NR5A1* index: all three tertiles correlated negatively with it (weak $\rho = -0.27$, middle -0.13, strong -0.36). What is established here is the general rule, in tumour tissue and with a matrix derived from the same data; where the skipped exons of NR5A1 themselves fall within such a distribution is scored against a reference set of transcripts in a companion manuscript (Tanaka et al., in preparation) and is not assessed here.

A scan of the 12 numeric clinical and genomic variables recorded for acc_tcga found no correlate of the index either (smallest $q = 0.38$): fraction of genome altered $\rho = -0.10$ ($P = 0.42$), mutation count $\rho = -0.11$ ($P = 0.34$), Weiss score $\rho = +0.14$ ($P = 0.30$, 55 patients) and the pathologist's mitotic count per 50 high-power fields $\rho = -0.0005$ ($P = 0.997$, 50 patients). The index therefore varies along an axis that none of the measured molecular or histological features of these tumours describes.

Discussion

The quantity measured here is the proportion of *NR5A1* transcripts that cannot encode a full-length receptor. It is low in normal adrenal, which transcribes *NR5A1* more heavily than any other tissue examined, and roughly twice as high in adrenocortical carcinoma. Within carcinomas it moves with the steroidogenic programme and not with histological grade.

What the result is not. It is not a grade marker: the correlation with Weiss score was absent under both definitions and at every threshold, and this was not for want of detection once exon 4→7 was included. It is not a prognostic marker independent of stage. It is not a consequence of how much *NR5A1* a tumour transcribes, since the index and *NR5A1* expression are uncorrelated within both normal adrenal and tumour. It is not a reflection of general splicing disorder, since the control-gene background is slightly lower in the carcinomas, and it is not explained by tumour purity.

The relation to steroidogenic gene expression is the finding that connects this to function. Tumours with more of these junctions express less *CYP11A1*, *CYP21A2* and *CYP17A1*, after adjustment for stage, purity and *NR5A1* expression. The direction matches what has been observed pharmacologically: CX-4945 induces the same junctions in H295R cells and collapses steroidogenic output. Two of the six genes tested, *STAR* and *HSD3B2*, showed no

relation here although both fall under CX-4945 in cells, and that discrepancy is unexplained.

Our findings also bear on a longstanding disagreement about *NR5A1* levels in ACC. Gene-level expression could not be compared between GTEx and TCGA in this framework. Expressed relative to the summed coverage of ten housekeeping genes, the housekeeping genes themselves differed between 272 normal adrenal samples and 79 carcinomas by ratios (median ACC over median adrenal) of 0.38 to 1.71, with $P < 0.01$ for eight of the ten (Mann-Whitney; *PGK1* $P = 0.50$ and *HPRT1* $P = 0.18$ were the exceptions) (Supplementary File S8, Supplementary Table S8). Because these genes also form the denominator, the ratios describe how the composition of the housekeeping pool differs between the cohorts, and that difference spans more than four-fold. The *NR5A1* ratio of 0.73 ($P = 4.8 \times 10^{-4}$) lies inside that range. Most steroidogenic genes fell well below it (*HSD3B2* 0.02, *MC2R* 0.17, *CYP17A1* 0.24, *STAR* 0.26, *CYP21A2* 0.31), which shows that differences of that size are visible against this background, whereas *CYP11A1* (0.44) lay close to its lower end and the *NR5A1* difference within it. We therefore make no claim about *NR5A1* expression levels between normal adrenal and ACC. Within TCGA-ACC, where the comparison is internal, expression behaved as Sbiera et al. (2010) reported: present in all tumours, and prognostic by its level.

This distinction between what a within-sample ratio can support and what a cross-cohort absolute level cannot is, in our view, the methodological point of the study. The index is a ratio of two junctions in the same gene, library and coverage profile. Depth and library chemistry act on numerator and denominator together and largely cancel. Degradation does not cancel completely: in normal adrenal the index rose with ischaemic time and fell with RNA integrity, but that effect raises the normal value and so works against the finding. Absolute expression has no such protection, and the housekeeping comparison above shows how much it matters in practice.

Why adrenal should hold this splicing low, and what releases it in carcinoma, is not addressed by these data. The ovary stands out in the opposite direction and is not explained by splicing background, which suggests the junctions are under tissue-specific control rather than arising as noise.

The comparison made here is between normal adrenal cortex and carcinoma, which is not the comparison that arises in practice. The diagnostic difficulty in adrenal pathology lies between adenoma and carcinoma, and it is concentrated in the tumours scoring 2 or 3 on the Weiss system and in the oncocytic variants, for which separate criteria are applied. A tissue marker that separated the two would be read on the resected specimen rather than on a biopsy, since adrenal biopsy is recommended against in the diagnostic work-up of suspected adrenocortical carcinoma (Fassnacht et al., 2018); it would therefore inform classification and follow-up rather than the decision to operate. Whether the index separates adenoma from carcinoma is untested here, and we could not test it: adenoma transcriptomes with junction-level coverage are not publicly available. The adenoma series we could identify are microarray-based, or are enriched for circular RNA by ribonuclease R treatment, or are single-nucleus libraries biased towards transcript ends, and the one substantial RNA-seq series of adrenocortical adenomas, 52 adenomas and 7 carcinomas

sequenced as paired-end libraries at a depth that would support junction-level analysis, is deposited under controlled access (Di Dalmazi et al., 2020). What can be said from the present data is narrow but specific. The index is not a candidate for grading or for prognosis, since it tracks neither Weiss score nor mitotic count nor outcome, and it is not a surrogate for *NR5A1* expression, for splicing-factor abundance or for global splicing. If it has a diagnostic use it is the binary one, and establishing that requires adenoma material measured in the same way.

Limitations

The central comparison is between cohorts. TCGA-ACC contains no normal adrenal tissue and GTEx contains no carcinoma, so normal and tumour necessarily come from different studies, different tissue sources (post-mortem against surgical) and different library preparations. The control-gene background rate, which is the quantity a general technical difference would move, differed only slightly between them (0.051% against 0.049%) and in the direction opposite to the finding, and the index is a within-sample ratio; but no within-cohort replication of the normal-versus-tumour contrast is available here.

Gene-level expression could not be compared across cohorts, as set out in the Discussion. Statements about *NR5A1* abundance in this work are therefore confined to within-cohort comparisons.

Adrenocortical adenoma was not examined. TCGA-ACC contains no benign adrenocortical tumours, so whether the index separates adenoma from carcinoma, which is the clinically useful question, is untested.

The junctions are inferred from short-read alignments. The products of exon 2→7 and exon 3→7 have been confirmed by Sanger sequencing in a cell line, but the exon 4→7 and exon 1→7 junctions have not been confirmed in human tissue, and no protein corresponding to any of these transcripts has been demonstrated.

The observation that the association with steroidogenic expression follows the reading frame was made after the main analysis and is reported as such. It rests on three junction-level correlations in 73 patients.

Many statistical tests were performed. Only the normal-versus-tumour comparison ($p = 1.4 \times 10^{-6}$ to 1.9×10^{-4} across six specifications) is robust to any reasonable correction. The steroidogenic association ($p = 0.004$) and all survival analyses should be read as exploratory and in need of independent confirmation.

The mechanism behind the higher index in ACC is unidentified. Phosphorylation state, the layer at which SR protein activity is set, is not visible in RNA-seq, so the possibility that reduced SR protein phosphorylation underlies the shift was not tested here; only mRNA abundance and autoregulatory splicing, two proxies for it, were. The autoregulatory-splicing index of *SRSF4*, the gene implicated by that hypothesis, fell in the predicted direction but did not survive correction ($\rho = -0.24$, $P = 0.065$, $q = 0.18$). With 73 patients the search has limited reach: a Spearman correlation of about 0.32 or larger is detectable at

80% power, and smaller effects are not, so the absence of a surviving correlate does not exclude factors of modest effect.

Tests of whether the skipping restrains tumour growth were negative in direction as well as in significance. Were the skipping a feedback response to SF-1 excess, the index should rise with *NR5A1* expression, and it does not ($\rho = -0.08$). Were it curbing growth, the index should fall as proliferation rises; instead all six proliferation genes examined correlated positively with it, two of them robustly to the modelling choice after adjustment (*AURKA*, rank-based P = 0.0043; *BUB1*, P = 0.014) while *MKI67* and *CCNB1* were not and the pathologist's mitotic count was flat. We have no interpretation for that pattern. The absolute shift is in any case small: the median index is 0.38% in normal adrenal against 0.90% in ACC, so even complete loss of the skipped transcripts would alter productive *NR5A1* mRNA by about half a percentage point, which is unlikely to be consequential against the several-fold *NR5A1* overexpression reported in these tumours.

Methods

Data sources. Exon-exon junction counts were obtained from Snaptron, compilation `gtexv2` for GTEx and `tcgav2` for TCGA. Clinical variables for TCGA-ACC (Weiss score, stage, hormone excess, overall and disease-free survival) were taken from cBioPortal study `acc_tcga`. Tumour purity was taken from the TCGA PanCanAtlas ABSOLUTE master call set. Exon coordinates were taken from Ensembl for the canonical *NR5A1* transcript ENST00000373588. Some GTEx samples are listed in Snaptron under more than one rail identifier, with external identifiers that differ only in a trailing version number and identical junction counts; one identifier was kept for each sample, which removed 68 identifiers in all and 16 in the five tissues analysed here. No duplicate was found among the TCGA-ACC samples.

Junction definitions. Every junction overlapping chr9:124,479,231–124,509,456 (GRCh38) was retrieved and annotated by matching its intron boundaries, within ± 2 nucleotides, to those predicted from the exon model. *NR5A1* is on the minus strand, so the junction joining exon *n* to exon 7 spans from the genomic end of exon 7 plus one to the genomic start of exon *n* minus one. The six junctions sharing the exon 7 acceptor are listed in the Results.

Skip index. For each sample, index = (exon 2→7 + exon 3→7 + exon 4→7 coverage) / (exon 6→7 coverage). Samples with fewer than 20 reads on the denominator were excluded; thresholds of 1, 5, 10, 20, 50 and 100 were examined and are reported where relevant. For TCGA, counts were summed across runs belonging to the same patient before the ratio was taken. The narrower definition (exon 2→7 + exon 3→7) is reported as a sensitivity analysis.

Splicing background. For each sample, background = unannotated junction reads / total junction reads, summed over fifteen broadly expressed control genes (*ACTB*, *GAPDH*, *PGK1*, *SDHA*, *YWHAZ*, *PPIA*, *RPL13A*, *TBP*, *HPRT1*, *B2M*, *CLTC*, *VCL*, *MYH9*, *FLNA*, *EEF2*), requiring at least 200 total reads. Background-normalised index = index / background.

Gene-level expression. Gene-level coverage was retrieved from the Snaptron `genes` endpoint and divided, per sample, by the summed coverage of ten of the control genes

above (*ACTB*, *GAPDH*, *PGK1*, *SDHA*, *YWHAZ*, *PPIA*, *RPL13A*, *TBP*, *HPRT1*, *B2M*); samples with a sum below 1000 were excluded. This quantity is used for within-cohort comparisons, and across cohorts only to show, in the Discussion, that it cannot support them. In survival models it entered as its natural logarithm.

Statistics. Group comparisons used the Mann-Whitney U test and the Kruskal-Wallis test; associations between continuous variables used Spearman correlation. Detection (index > 0) was modelled by logistic regression with log(denominator) and, where stated, log(background) as covariates. Continuous outcomes were modelled by ordinary least squares. Survival was modelled by Cox proportional hazards; the proportional hazards assumption was examined by comparing hazard ratios in the first and second halves of follow-up. Analyses were performed in Python with scipy, statsmodels and pandas. No correction for multiple comparisons was applied; the number of tests performed is stated in Limitations.

Trans-acting factors, global skipping and splice-site strength. Expression of 30 splicing-regulatory genes was taken from the Snaptron `genes` endpoint and normalised as above. Autoregulatory splicing was measured for 14 SR protein, CLK and SRPK genes: annotated junctions of each gene were grouped by shared start or shared end coordinate, the junction carrying the most reads in each group was taken as the reference, and the index was the summed reads on the remaining junctions divided by the group total, requiring at least 20 reads per group and 100 reads per patient. For the global exon-skipping index the same grouping was applied to 48 broadly expressed genes with the shortest junction of each group as the reference, requiring at least 500 reads per patient. Splice-site sequences were taken from the hg38 assembly through the UCSC `getData/sequence` API, and intron orientation was assigned from the terminal dinucleotides (GT–AG for the plus orientation, CT–AC for the minus). Position weight matrices were built from the 9-nt 5′ sites (positions −3 to +6) and 23-nt 3′ sites (−20 to +3) of constitutive junctions carrying at least 200 reads, with Laplace smoothing and a log-odds score against a 27/23/23/27% ACGT background; no external splice-site model was used. Each group's bypassed site was the non-shared end of its reference junction. Clinical and genomic variables came from the cBioPortal `clinical-data` endpoint for acc_tcga, and the 12 variables with numeric values in at least 30 patients were tested. Associations with the index were adjusted for stage, ABSOLUTE purity and *NR5A1* expression by ordinary least squares, and robustness to the modelling choice was checked by repeating each fit with log-transformed and with rank-transformed index values and after removing the three highest index values. Immediate-early gene expression was summarised as the mean z-score of log expression of FOS, FOSB, JUN, JUNB, EGR1, NR4A1, NR4A2, NR4A3 and DUSP1, normalised as above, and its association with the index was assessed by rank-based partial correlation adjusting for denominator depth and, in GTEx, for donor age and sex. GTEx sample-quality variables (RNA integrity number, SMRIN; ischaemic time, SMTSISCH; exonic mapping rate, SMEXNCRT) were taken from the Snaptron sample metadata and related to the index by the same rank-based partial correlation, adjusting for denominator depth, with P from the t-test of the rank regression coefficient (n − 3 degrees of freedom); permutation tests (10,000 permutations) gave $P = 1 \times 10^{-4}$, the smallest value attainable, for all three variables in adrenal.

Code. Data retrieval, the per-sample table and Figures 1–4 are produced by a single script that queries the public endpoints directly (Supplementary File S1), and every statistic computed from the per-sample GTEx and TCGA-ACC values by a second that reads only the supplementary tables (Supplementary File S7), with the analyses of trans-acting factors, global skipping and splice-site strength in a further script (Supplementary File S3) and the sample-quality table in another (Supplementary File S6). Figures were drawn with matplotlib.

Figures

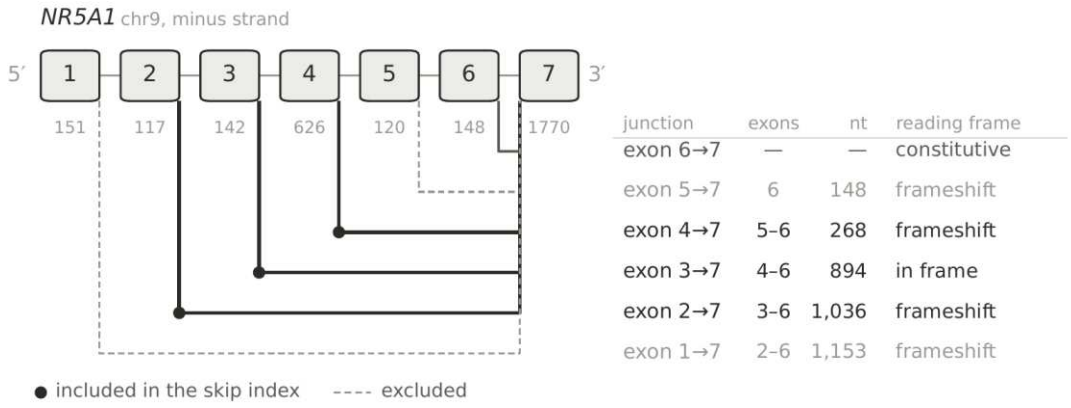


Figure 1. Junctions joining an upstream exon directly to exon 7 at the *NR5A1* locus. Exons of the canonical transcript (ENST00000373588) are drawn 5′ to 3′ as evenly spaced boxes, not to genomic scale; the number below each box is its length in nucleotides. *NR5A1* lies on the minus strand of chromosome 9, so the diagram is reversed relative to genomic coordinates. Lines below the gene mark the six junctions that share the exon 7 acceptor, listed at right with the exons each removes and the consequence for the reading frame of the NM_004959.5 coding sequence. Solid lines with a filled circle are the three junctions that enter the skip index; dashed lines are the junctions excluded from it, namely exon 1→7, which satisfies the definition but contributes 43 reads across 32 GTEx samples, and exon 5→7, which removes a single exon.

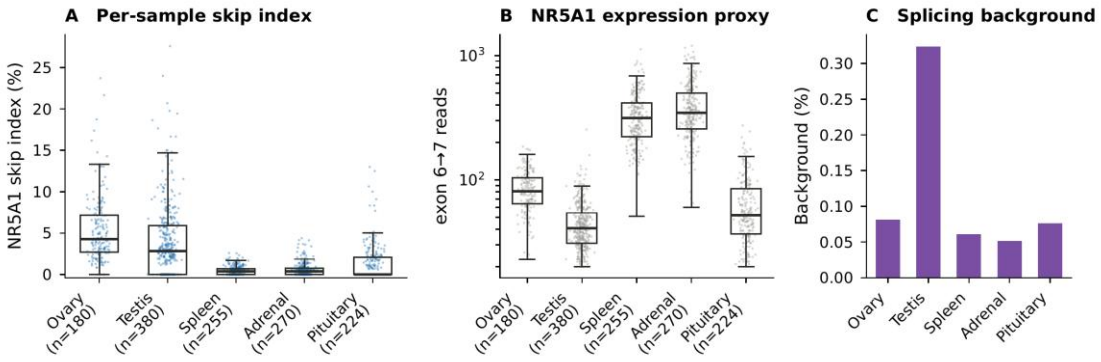


Figure 2. The skip index across *NR5A1*-expressing tissues in GTEx. (A) Per-sample skip index, defined as the summed coverage of the exon 2→7, exon 3→7 and exon 4→7 junctions divided by the coverage of the exon 6→7 junction. Samples with fewer than 20 reads on the denominator were excluded. Boxes show the median and interquartile range, whiskers extend to 1.5 × IQR, and each point is one sample. (B) The denominator itself, a proxy for *NR5A1* expression, on a logarithmic scale: adrenal carries the largest denominator and the lowest index. (C) Median splicing background per tissue, measured as the fraction of junction reads falling on unannotated junctions across fifteen control genes elsewhere in the genome.

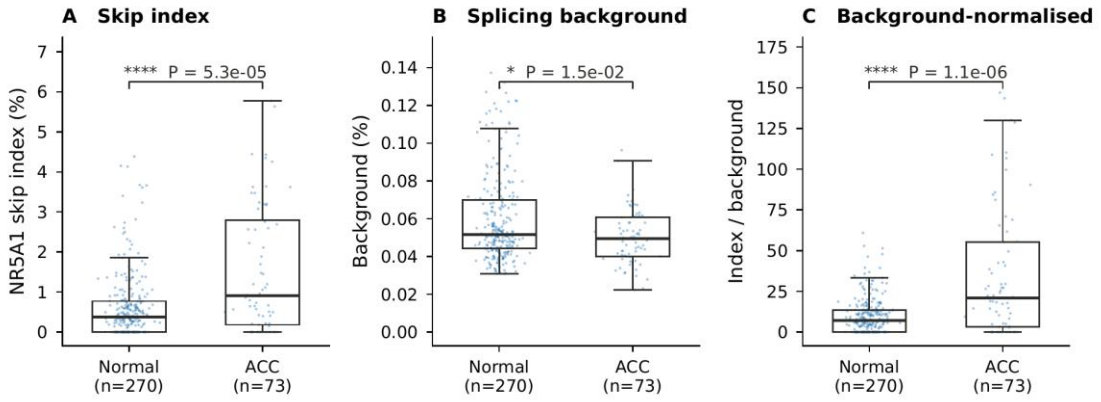


Figure 3. The skip index in normal adrenal and adrenocortical carcinoma. Normal adrenal samples are from GTEx, carcinomas from TCGA-ACC, both quantified with the same junction definitions and the same denominator threshold. (A) Skip index. (B) Splicing background, which is slightly lower in the carcinomas and therefore cannot account for their higher index in (A). (C) Skip index divided by each sample's own background. P values are from two-sided Mann-Whitney U tests.

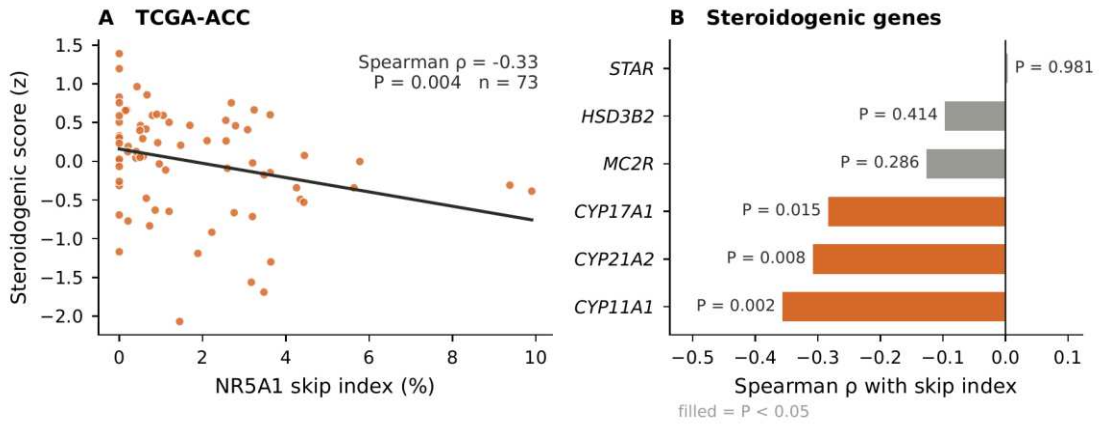


Figure 4. Within adrenocortical carcinoma, a higher skip index accompanies lower steroidogenic gene expression. (A) Skip index against a steroidogenic score formed from the standardised log expression of *STAR*, *CYP11A1*, *CYP17A1*, *CYP21A2* and *HSD3B2*, each normalised to ten control genes in the same sample. The line is an ordinary least-squares fit; the statistic given is Spearman's rho. (B) Spearman correlation between the skip index and each steroidogenic gene individually. Filled bars are nominally significant at $P < 0.05$; of these, only *CYP11A1* survives Bonferroni correction for the tests shown.

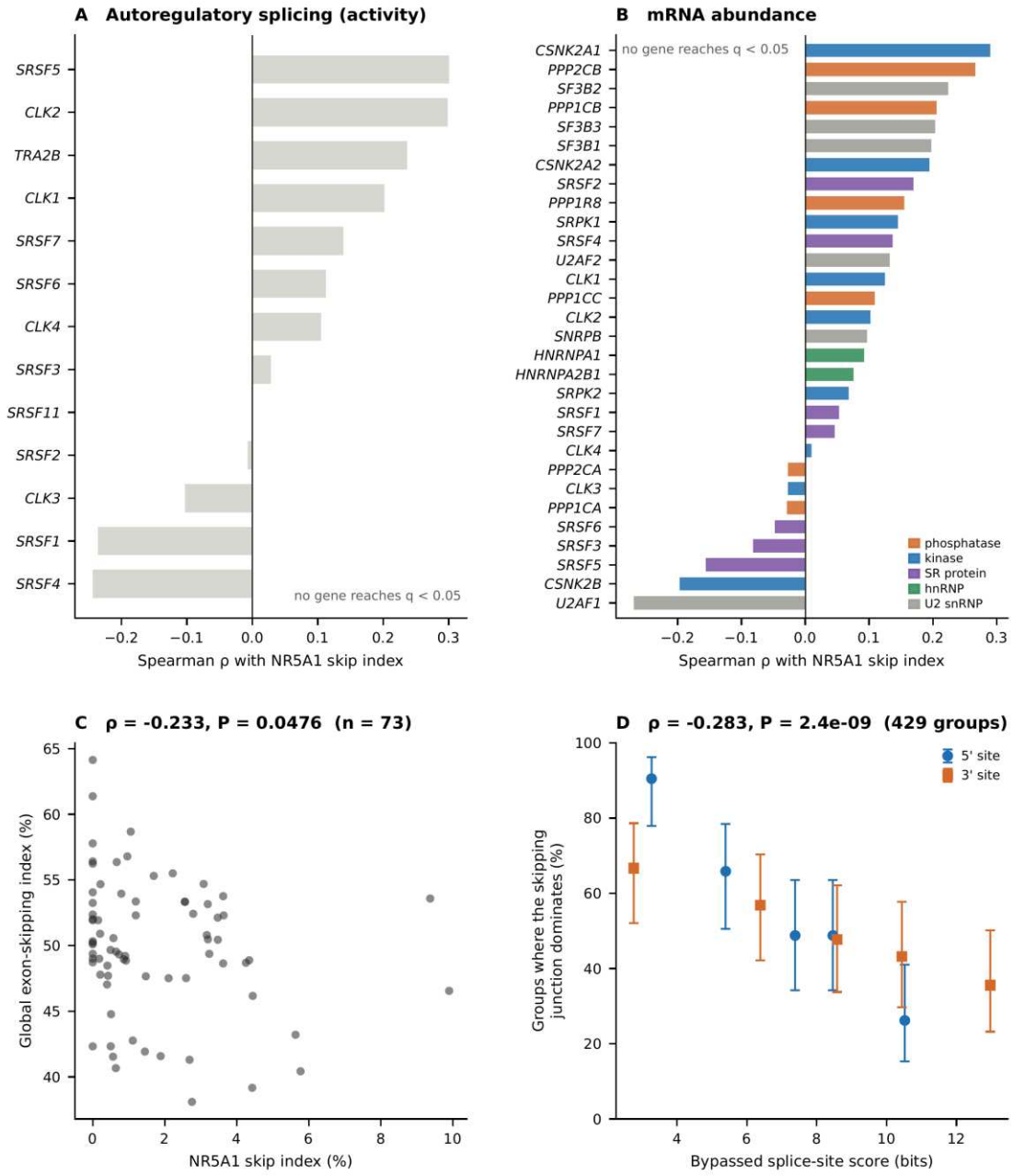


Figure 5. The skip index is not explained by splicing-factor levels or by global splicing, while splice-site strength governs skipping. (A) Spearman correlation between the *NR5A1* skip index and the autoregulatory-splicing index of 13 SR protein, CLK and SRPK genes in TCGA-ACC (73 patients). Bars are drawn solid where q < 0.05 after Benjamini–Hochberg correction; none reach it. (B) The same correlation for depth-normalised mRNA abundance of 30 splicing-regulatory genes. (C) The *NR5A1* skip index against the global exon-skipping index computed over 474 junction groups in 48 broadly expressed genes, one point per patient. The correlation printed on the panel is unadjusted; it does not survive adjustment for denominator depth, tumour purity, stage and NR5A1 expression

(adjusted P = 0.16). (D) Skipping fraction of each of 429 junction groups against the position weight matrix score of its bypassed splice site, in bits.

Data availability, contributions and declarations

Data availability. All data analysed are public. GTEx and TCGA junction counts are served by Snaptron (https://snaptron.cs.jhu.edu). Clinical annotation for TCGA-ACC is available through cBioPortal (https://www.cbioportal.org, study `acc_tcga`). Tumour purity estimates are from the TCGA PanCanAtlas ABSOLUTE call set at the GDC Data Portal. Values generated in this study are provided as Supplementary Tables S1–S8: per-sample junction and background counts for every sample with at least one exon 6→7 read (S1), correlations with splicing-factor mRNA abundance (S2) and with autoregulatory splicing (S3), the 429 junction groups with their splice-site scores (S4), the per-patient table underlying Figure 5 (S5), the immediate-early gene scores used in the tissue-handling comparison (S6), the sample-quality variables with the index for GTEx adrenal, ovary, pituitary and testis (S7), and gene-level coverage of the housekeeping and steroidogenic genes in GTEx adrenal and TCGA-ACC (S8).

Code availability. Supplementary File S1 is a self-contained Python script that queries only the public endpoints listed above, removes duplicated sample entries, and writes Supplementary Table S1 and Figures 1–4. Supplementary File S7 computes from Supplementary Tables S1, S6 and S7 every statistic in the Results that rests on per-sample GTEx and TCGA-ACC values, and needs no network access. Supplementary File S2 draws Figure 1 and needs no network access. Supplementary File S3 is the self-contained script for the analyses shown in Figure 5 and writes Supplementary Tables S2–S5; Supplementary File S5 draws Figure 5 from those tables without network access. Supplementary File S4 fits every survival model reported here from Supplementary Tables S1 and S5 together with clinical data retrieved from cBioPortal. Supplementary File S6 is a self-contained script for the sample-quality analysis that writes Supplementary Table S7, and Supplementary File S8 performs the gene-level comparison between cohorts and writes Supplementary Table S8. None requires credentials.

Author contributions. K.O. conceived the study, performed the analysis and wrote the manuscript. T.T. generated the CX-4945 experimental data in H295R cells on which the junction definitions used here are based, and reviewed and edited the manuscript. T.Y. contributed clinical and endocrinological expertise to the interpretation of the findings, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding. Grants-in-Aid for Scientific Research from MEXT/JSPS (18K09212 to K.O.; 16K09815 to T.T.); 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 Department of Bioregulatory Science of Life-Related Diseases of Fukuoka University received an unrestricted donation from MSD K.K.; this had no role in

the design, execution, analysis or interpretation of the study. The authors declare no other 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. 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.
2. Sbiera S, Schmull S, Assie G, et al. High diagnostic and prognostic value of steroidogenic factor-1 expression in adrenal tumors. *J Clin Endocrinol Metab*. 2010;95(10):E161–E171.
3. Figueiredo BC, Cavalli LR, Pianovski MA, et al. Amplification of the steroidogenic factor 1 gene in childhood adrenocortical tumors. *J Clin Endocrinol Metab*. 2005;90(2):615–619.
4. Pianovski MA, Cavalli LR, Figueiredo BC, et al. SF-1 overexpression in childhood adrenocortical tumours. *Eur J Cancer*. 2006;42(8):1040–1043.
5. Wilks C, Gaddipati P, Nellore A, Langmead B. Snaptron: querying splicing patterns across tens of thousands of RNA-seq samples. *Bioinformatics*. 2018;34(1):114–116.
6. 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.
7. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318–1330.
8. Zheng S, Cherniack AD, Dewal N, et al. Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell*. 2016;29(5):723–736.
9. Cerami E, Gao J, Dogrusoz U, et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–404.
10. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012;30(5):413–421.
11. 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.
12. Ohe K, Tanaka T, Horita Y, Harada Y, Yamasaki T, Abe I, Tanabe M, Nomiyama T, Kobayashi K, Enjoji M, Yanase T. Circular IRE-type RNAs of the NR5A1 gene are formed in adrenocortical cells. *Biochem Biophys Res Commun*. 2019;512(1):1–6.
13. Tanaka T, Ohe K, Nomiyama T, Yanase T. OR02-2 CX-4945 as a potential drug for adrenocortical carcinoma that induces multiple exon-skipping and circular RNA of NR5A1. *J Endocr Soc*. 2019;3(Supplement_1):OR02-2.
14. Di Dalmazi G, Altieri B, Scholz C, Sbiera S, Luconi M, Waldman J, Kastelan D, Ceccato F, Chiodini I, Arnaldi G, Riester A, Osswald A, Beuschlein F, Sauer S, Fassnacht M, Appenzeller S, Ronchi CL. RNA sequencing and somatic mutation status of adrenocortical tumors: novel pathogenetic insights. *J Clin Endocrinol Metab*. 2020;105(12):e4459–e4473.

15. Fassnacht M, Dekkers OM, Else T, Baudin E, Berruti A, de Krijger RR, Haak HR, Mihai R, Assie G, Terzolo M. European Society of Endocrinology Clinical Practice Guidelines on the management of adrenocortical carcinoma in adults, in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*. 2018;179(4):G1–G46.

References were verified against the primary sources on 31 October 2026.

Supplementary Files

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

- OheNR5A1skipindexv3supplementary.zip