Supplementary File S1

IGF2BP Family Expression and Survival in Lung Adenocarcinoma and Lung Squamous Cell Carcinoma

Supplementary tables contain full OS models (S1–S2), correlations (S3–S4), all 300 Hallmark results (S5a–S5d), continuous-expression models (S6), proportional-hazards diagnostics (S7), cohort characteristics and endpoint accounting (S8a–S8e), exploratory spline analyses (S9a–S9b), and external cohort results (S10a–S10b, S11a–S11b, S12–S14). Figures S1–S4 show OS and PFI curves; Figure S5 shows spline curves. Numerical results are unchanged.

Table S1. Full overall survival estimates in LUAD

Variable	Univ. N/ deaths	Univariate HR (95% CI)	p	Adjusted HR (95% CI)	p
Age >65 vs ≤65 years	492/179	1.194 (0.889–1.604)	0.238	1.277 (0.937–1.738)	0.121
Female vs male	492/179	0.897 (0.669–1.203)	0.468	0.877 (0.643–1.195)	0.406
Ever vs never smoker	478/169	0.932 (0.613–1.416)	0.740	0.930 (0.601–1.438)	0.744
Stage III–IV vs I–II	484/177	2.496 (1.825–3.413)	<0.001	2.286 (1.652–3.165)	<0.001
IGF2BP1 high vs low	492/179	1.365 (1.015–1.834)	0.039	1.166 (0.839–1.619)	0.361
IGF2BP2 high vs low	492/179	1.363 (1.015–1.829)	0.039	1.134 (0.824–1.560)	0.441
IGF2BP3 high vs low	492/179	1.595 (1.184–2.149)	0.002	1.486 (1.056–2.092)	0.023

Adjusted N=471; deaths=168. All seven variables enter one model simultaneously, using Efron ties and complete cases. No EGFR mutation variable is included. Clinical p values are descriptive; multiplicity-adjusted gene q values are in main Table 1. Univariate sample sizes vary by variable.

Table S2. Full overall survival estimates in LUSC

Variable	Univ. N/ deaths	Univariate HR (95% CI)	p	Adjusted HR (95% CI)	p
Age >65 vs ≤65 years	489/212	1.267 (0.951–1.688)	0.106	1.332 (0.988–1.796)	0.060
Female vs male	489/212	0.831 (0.603–1.145)	0.257	0.834 (0.603–1.152)	0.271
Ever vs never smoker	477/209	0.597 (0.264–1.352)	0.216	0.599 (0.259–1.383)	0.230
Stage III–IV vs I–II	485/210	1.527 (1.109–2.105)	0.010	1.677 (1.207–2.329)	0.002
IGF2BP1 high vs low	488/211	0.738 (0.561–0.970)	0.030	0.769 (0.578–1.023)	0.072
IGF2BP2 high vs low	488/212	0.868 (0.662–1.138)	0.304	0.942 (0.700–1.268)	0.695
IGF2BP3 high vs low	488/212	0.751 (0.572–0.986)	0.039	0.871 (0.647–1.172)	0.362

Adjusted N=470; deaths=206. All seven variables enter one model simultaneously, using Efron ties and complete cases. No EGFR mutation variable is included. Clinical p values are descriptive; multiplicity-adjusted gene q values are in main Table 1. Univariate sample sizes vary by variable.

Table S3. All unique Pearson correlations in LUAD

Normalized RSEM values without further transformation; pairwise complete observations. Each pair is shown once. BH q values adjust all 36 unique off-diagonal pairs within this histology, including age from the official TCGA-CDR.

Variable A	Variable B	N	Pearson r	p	BH q
IGF2BP1	IGF2BP2	492	0.353	<0.001	<0.001
IGF2BP1	IGF2BP3	492	0.551	<0.001	<0.001
IGF2BP1	EGFR	492	0.149	<0.001	0.003
IGF2BP1	KRAS	492	0.166	<0.001	0.001
IGF2BP1	TP53	492	-0.061	0.180	0.324
IGF2BP1	ALK	492	-0.045	0.322	0.464
IGF2BP1	BRAF	492	0.057	0.207	0.343
IGF2BP1	Age	492	-0.043	0.341	0.472
IGF2BP2	IGF2BP3	492	0.405	<0.001	<0.001
IGF2BP2	EGFR	492	0.223	<0.001	<0.001
IGF2BP2	KRAS	492	0.189	<0.001	<0.001
IGF2BP2	TP53	492	-0.076	0.091	0.206
IGF2BP2	ALK	492	0.005	0.909	0.965
IGF2BP2	BRAF	492	0.152	<0.001	0.003
IGF2BP2	Age	492	0.005	0.913	0.965
IGF2BP3	EGFR	492	0.345	<0.001	<0.001
IGF2BP3	KRAS	492	0.123	0.006	0.018
IGF2BP3	TP53	492	-0.162	<0.001	0.001
IGF2BP3	ALK	492	-0.034	0.453	0.583
IGF2BP3	BRAF	492	0.158	<0.001	0.002
IGF2BP3	Age	492	-0.117	0.009	0.024
EGFR	KRAS	492	0.011	0.810	0.941
EGFR	TP53	492	-0.123	0.007	0.018
EGFR	ALK	492	-0.024	0.595	0.729
EGFR	BRAF	492	0.061	0.177	0.324
EGFR	Age	492	-0.002	0.965	0.965
KRAS	TP53	492	-0.057	0.209	0.343
KRAS	ALK	492	-0.004	0.934	0.965
KRAS	BRAF	492	0.080	0.078	0.187
KRAS	Age	492	-0.023	0.607	0.729
TP53	ALK	492	0.067	0.139	0.278
TP53	BRAF	492	-0.003	0.940	0.965
TP53	Age	492	0.073	0.107	0.227
ALK	BRAF	492	-0.040	0.372	0.496
ALK	Age	492	-0.048	0.291	0.436
BRAF	Age	492	-0.053	0.242	0.379

Table S4. All unique Pearson correlations in LUSC

Normalized RSEM values without further transformation; pairwise complete observations. Each pair is shown once. BH q values adjust all 36 unique off-diagonal pairs within this histology, including age from the official TCGA-CDR.

Variable A	Variable B	N	Pearson r	p	BH q
IGF2BP1	IGF2BP2	487	0.270	<0.001	<0.001
IGF2BP1	IGF2BP3	487	0.141	0.002	0.007
IGF2BP1	TP53	487	0.104	0.021	0.059
IGF2BP1	CDKN2A	487	-0.068	0.136	0.326
IGF2BP1	SOX2	487	0.116	0.011	0.035
IGF2BP1	PIK3CA	487	0.105	0.020	0.059
IGF2BP1	NOTCH1	487	0.043	0.345	0.565
IGF2BP1	Age	488	-0.065	0.153	0.345
IGF2BP2	IGF2BP3	487	0.339	<0.001	<0.001
IGF2BP2	TP53	487	-0.027	0.549	0.766
IGF2BP2	CDKN2A	487	-0.013	0.774	0.883
IGF2BP2	SOX2	487	0.479	<0.001	<0.001
IGF2BP2	PIK3CA	487	0.409	<0.001	<0.001
IGF2BP2	NOTCH1	487	0.058	0.204	0.386
IGF2BP2	Age	488	0.009	0.850	0.900
IGF2BP3	TP53	487	-0.011	0.800	0.883
IGF2BP3	CDKN2A	487	-0.028	0.540	0.766
IGF2BP3	SOX2	487	0.343	<0.001	<0.001
IGF2BP3	PIK3CA	487	0.274	<0.001	<0.001
IGF2BP3	NOTCH1	487	0.026	0.573	0.766
IGF2BP3	Age	488	-0.062	0.171	0.362
TP53	CDKN2A	487	-0.011	0.810	0.883
TP53	SOX2	487	0.058	0.200	0.386
TP53	PIK3CA	487	0.050	0.273	0.492
TP53	NOTCH1	487	0.196	<0.001	<0.001
TP53	Age	488	-0.044	0.335	0.565
CDKN2A	SOX2	487	-0.004	0.929	0.933
CDKN2A	PIK3CA	487	-0.014	0.761	0.883
CDKN2A	NOTCH1	487	-0.074	0.104	0.268
CDKN2A	Age	488	-0.012	0.798	0.883
SOX2	PIK3CA	487	0.528	<0.001	<0.001
SOX2	NOTCH1	487	0.004	0.933	0.933
SOX2	Age	488	-0.116	0.011	0.035
PIK3CA	NOTCH1	487	-0.032	0.482	0.754
PIK3CA	Age	488	-0.025	0.574	0.766
NOTCH1	Age	488	-0.013	0.768	0.883

Table S5a. Complete Hallmark results in LUAD part 1

Each cell reports NES followed by the within-comparison FDR q value. Positive NES: high expression; negative NES: low expression. No significance-based selection. Hallmark 2026.1.Hs; 1,000 phenotype permutations; seed 42.

Hallmark gene set	IGF2BP1 NES and q	IGF2BP2 NES and q	IGF2BP3 NES and q
ADIPOGENESIS	-1.125; q=1.000	-1.965; q=0.024	-1.599; q=0.531
ALLOGRAFT REJECTION	-0.723; q=1.000	0.815; q=0.746	0.518; q=0.960
ANDROGEN RESPONSE	-0.778; q=1.000	1.155; q=0.581	-0.821; q=0.919
ANGIOGENESIS	1.006; q=0.609	1.167; q=0.593	0.744; q=0.985
APICAL JUNCTION	0.828; q=0.819	1.249; q=0.599	0.812; q=0.964
APICAL SURFACE	0.808; q=0.782	1.014; q=0.641	0.640; q=0.951
APOPTOSIS	-0.787; q=1.000	1.179; q=0.609	0.786; q=0.958
BILE ACID METABOLISM	-0.908; q=1.000	-1.482; q=0.317	-1.238; q=1.000
CHOLESTEROL HOMEOSTASIS	1.373; q=0.310	1.147; q=0.562	1.240; q=0.534
COAGULATION	0.896; q=0.747	-0.807; q=0.797	-1.130; q=1.000
COMPLEMENT	-0.633; q=1.000	0.926; q=0.652	-0.524; q=1.000
DNA REPAIR	1.427; q=0.268	1.189; q=0.630	1.690; q=0.082
E2F TARGETS	1.659; q=0.116	1.683; q=0.227	1.894; q=0.032
EPITHELIAL MESENCHYMAL TRANSITION	1.156; q=0.476	1.438; q=0.607	1.127; q=0.648
ESTROGEN RESPONSE EARLY	-0.953; q=1.000	-1.130; q=0.541	-1.155; q=1.000
ESTROGEN RESPONSE LATE	1.014; q=0.628	0.997; q=0.642	-0.940; q=0.956
FATTY ACID METABOLISM	-1.340; q=1.000	-2.065; q=0.020	-1.607; q=1.000
G2M CHECKPOINT	1.807; q=0.066	1.823; q=0.165	2.012; q=0.015
GLYCOLYSIS	1.534; q=0.189	1.736; q=0.216	1.854; q=0.037
HEDGEHOG SIGNALING	-0.735; q=1.000	0.896; q=0.674	-0.828; q=0.966
HEME METABOLISM	-1.273; q=1.000	-1.421; q=0.282	-1.192; q=1.000
HYPOXIA	1.336; q=0.310	1.341; q=0.611	1.220; q=0.528
IL2 STAT5 SIGNALING	-0.771; q=1.000	0.786; q=0.763	-0.849; q=0.991
IL6 JAK STAT3 SIGNALING	-0.746; q=1.000	0.949; q=0.641	0.542; q=0.984
INFLAMMATORY RESPONSE	-0.745; q=1.000	0.959; q=0.651	-0.605; q=1.000

Table S5b. Complete Hallmark results in LUAD part 2

Each cell reports NES followed by the within-comparison FDR q value. Positive NES: high expression; negative NES: low expression. No significance-based selection. Hallmark 2026.1.Hs; 1,000 phenotype permutations; seed 42.

Hallmark gene set	IGF2BP1 NES and q	IGF2BP2 NES and q	IGF2BP3 NES and q
INTERFERON ALPHA RESPONSE	-0.422; q=0.986	1.262; q=0.622	0.819; q=1.000
INTERFERON GAMMA RESPONSE	-0.610; q=0.989	1.057; q=0.657	0.738; q=0.952
KRAS SIGNALING DN	1.221; q=0.411	0.827; q=0.753	0.883; q=0.980
KRAS SIGNALING UP	-0.645; q=1.000	1.047; q=0.642	0.644; q=0.986
MITOTIC SPINDLE	1.906; q=0.051	1.999; q=0.075	2.174; q=0.009
mTORC1 SIGNALING	1.857; q=0.058	1.593; q=0.325	2.100; q=0.012
MYC TARGETS V1	1.549; q=0.195	1.287; q=0.620	1.678; q=0.080
MYC TARGETS V2	1.676; q=0.119	1.393; q=0.559	1.715; q=0.094
MYOGENESIS	0.811; q=0.811	-1.010; q=0.559	-0.981; q=1.000
NOTCH SIGNALING	-1.177; q=1.000	0.784; q=0.742	-0.915; q=0.931
OXIDATIVE PHOSPHORYLATION	0.688; q=0.866	-1.480; q=0.256	-0.512; q=0.959
P53 PATHWAY	-1.303; q=1.000	-1.014; q=0.604	-1.242; q=1.000
PANCREAS BETA CELLS	1.097; q=0.533	-0.662; q=0.919	1.032; q=0.729
PEROXISOME	0.750; q=0.826	-1.566; q=0.272	-0.943; q=1.000
PI3K AKT mTOR SIGNALING	1.480; q=0.224	1.135; q=0.551	1.440; q=0.259
PROTEIN SECRETION	-0.809; q=1.000	-0.489; q=0.968	0.820; q=1.000
REACTIVE OXYGEN SPECIES PATHWAY	0.512; q=0.959	-1.138; q=0.595	0.701; q=0.960
SPERMATOGENESIS	1.685; q=0.134	1.407; q=0.603	1.713; q=0.081
TGF BETA SIGNALING	-0.580; q=0.965	0.973; q=0.655	-0.765; q=0.905
TNFA SIGNALING VIA NFKB	-0.915; q=1.000	1.287; q=0.681	-0.566; q=1.000
UNFOLDED PROTEIN RESPONSE	1.944; q=0.077	1.019; q=0.661	1.639; q=0.093
UV RESPONSE DN	-0.759; q=1.000	1.204; q=0.644	-0.819; q=0.867
UV RESPONSE UP	1.350; q=0.313	-1.054; q=0.598	1.112; q=0.630
WNT BETA CATENIN SIGNALING	1.239; q=0.410	0.761; q=0.751	-1.052; q=0.978
XENOBIOTIC METABOLISM	-0.666; q=1.000	-1.304; q=0.393	-1.055; q=1.000

LUAD: 492 participants (246 high/246 low), 20,191 ranked genes. All 50 sets passed matched size limits of 15–500. Nominal permutation p values, matched-set sizes, leading-edge genes, ranked lists, and full-precision outputs are included in Supplementary Data S1.

Table S5c. Complete Hallmark results in LUSC part 1

Each cell reports NES followed by the within-comparison FDR q value. Positive NES: high expression; negative NES: low expression. No significance-based selection. Hallmark 2026.1.Hs; 1,000 phenotype permutations; seed 42.

Hallmark gene set	IGF2BP1 NES and q	IGF2BP2 NES and q	IGF2BP3 NES and q
ADIPOGENESIS	-1.665; q=0.456	-1.703; q=0.370	-1.666; q=0.547
ALLOGRAFT REJECTION	-0.578; q=0.929	-1.710; q=0.710	-1.402; q=0.261
ANDROGEN RESPONSE	-1.522; q=0.700	-1.458; q=0.322	-1.547; q=0.264
ANGIOGENESIS	-0.757; q=0.904	-1.289; q=0.335	-1.562; q=0.387
APICAL JUNCTION	1.004; q=0.600	0.777; q=0.920	-1.260; q=0.382
APICAL SURFACE	0.852; q=0.794	-0.908; q=0.601	-1.052; q=0.519
APOPTOSIS	-1.083; q=0.733	-1.342; q=0.325	-1.420; q=0.301
BILE ACID METABOLISM	0.698; q=0.867	-1.443; q=0.315	-1.086; q=0.516
CHOLESTEROL HOMEOSTASIS	-0.732; q=0.871	1.214; q=0.667	0.863; q=0.761
COAGULATION	-1.305; q=0.677	-1.422; q=0.317	-1.615; q=0.483
COMPLEMENT	-1.151; q=0.645	-1.678; q=0.209	-1.577; q=0.447
DNA REPAIR	1.245; q=0.402	0.599; q=0.914	1.341; q=0.422
E2F TARGETS	1.586; q=0.317	1.428; q=0.515	2.029; q=0.047
EPITHELIAL MESENCHYMAL TRANSITION	0.520; q=0.959	-0.911; q=0.618	-1.285; q=0.369
ESTROGEN RESPONSE EARLY	-1.436; q=0.629	-1.137; q=0.433	-1.444; q=0.320
ESTROGEN RESPONSE LATE	-1.165; q=0.669	-1.195; q=0.417	-1.226; q=0.372
FATTY ACID METABOLISM	-1.480; q=0.645	-1.258; q=0.355	-0.971; q=0.590
G2M CHECKPOINT	1.642; q=0.339	1.711; q=0.444	2.018; q=0.026
GLYCOLYSIS	-0.853; q=0.856	1.024; q=0.817	0.777; q=0.726
HEDGEHOG SIGNALING	1.386; q=0.283	1.335; q=0.513	0.978; q=0.694
HEME METABOLISM	-0.927; q=0.804	-1.308; q=0.350	-1.243; q=0.366
HYPOXIA	-0.878; q=0.853	0.791; q=1.000	-1.086; q=0.495
IL2 STAT5 SIGNALING	-0.697; q=0.881	-1.511; q=0.284	-1.158; q=0.438
IL6 JAK STAT3 SIGNALING	-0.776; q=0.912	-1.665; q=0.150	-1.522; q=0.268
INFLAMMATORY RESPONSE	-0.984; q=0.830	-1.523; q=0.305	-1.500; q=0.268

Table S5d. Complete Hallmark results in LUSC part 2

Each cell reports NES followed by the within-comparison FDR q value. Positive NES: high expression; negative NES: low expression. No significance-based selection. Hallmark 2026.1.Hs; 1,000 phenotype permutations; seed 42.

Hallmark gene set	IGF2BP1 NES and q	IGF2BP2 NES and q	IGF2BP3 NES and q
INTERFERON ALPHA RESPONSE	-0.678; q=0.872	-1.668; q=0.177	-1.429; q=0.313
INTERFERON GAMMA RESPONSE	-0.938; q=0.827	-1.701; q=0.250	-1.414; q=0.267
KRAS SIGNALING DN	1.471; q=0.293	1.170; q=0.672	0.988; q=0.745
KRAS SIGNALING UP	-0.945; q=0.858	-1.391; q=0.310	-1.550; q=0.298
MITOTIC SPINDLE	1.471; q=0.251	1.362; q=0.545	1.178; q=0.558
mTORC1 SIGNALING	0.846; q=0.756	0.945; q=0.892	1.664; q=0.099
MYC TARGETS V1	1.080; q=0.568	0.866; q=0.971	1.714; q=0.092
MYC TARGETS V2	1.458; q=0.233	1.588; q=0.471	1.796; q=0.074
MYOGENESIS	0.811; q=0.764	-0.796; q=0.724	-1.416; q=0.284
NOTCH SIGNALING	1.045; q=0.580	0.695; q=0.980	-0.880; q=0.671
OXIDATIVE PHOSPHORYLATION	-1.198; q=0.783	-1.043; q=0.510	0.810; q=0.733
P53 PATHWAY	-1.354; q=0.640	1.098; q=0.741	-0.981; q=0.597
PANCREAS BETA CELLS	1.488; q=0.325	0.682; q=0.938	0.901; q=0.763
PEROXISOME	-1.072; q=0.708	-1.307; q=0.331	-0.853; q=0.690
PI3K AKT mTOR SIGNALING	-0.746; q=0.886	-1.180; q=0.416	0.845; q=0.734
PROTEIN SECRETION	-1.381; q=0.669	-0.970; q=0.585	-0.534; q=0.951
REACTIVE OXYGEN SPECIES PATHWAY	-1.829; q=0.315	-1.144; q=0.444	-0.781; q=0.742
SPERMATOGENESIS	1.720; q=0.401	0.779; q=0.981	1.160; q=0.522
TGF BETA SIGNALING	-1.285; q=0.647	-1.060; q=0.511	-1.933; q=0.163
TNFA SIGNALING VIA NFKB	-1.180; q=0.693	-1.409; q=0.310	-1.556; q=0.337
UNFOLDED PROTEIN RESPONSE	1.222; q=0.399	0.678; q=0.889	1.252; q=0.499
UV RESPONSE DN	-0.837; q=0.847	-0.680; q=0.838	-1.257; q=0.367
UV RESPONSE UP	1.264; q=0.415	-0.954; q=0.584	-0.900; q=0.665
WNT BETA CATENIN SIGNALING	1.513; q=0.362	1.463; q=0.585	-0.801; q=0.739
XENOBIOTIC METABOLISM	-1.191; q=0.729	-1.382; q=0.298	-1.287; q=0.388

LUSC: 488 participants per gene (244 high/244 low), 20,242 ranked genes. All 50 sets passed matched size limits of 15–500. Nominal permutation p values, matched-set sizes, leading-edge genes, ranked lists, and full-precision outputs are included in Supplementary Data S1.

Table S6. Continuous expression sensitivity analysis

The three genes entered each model simultaneously as standardized $\log_2$(RSEM+1), with age >65, female sex, ever smoking, and stage III–IV. Standardization used all nonmissing expression values within histology. HRs represent a one-standard-deviation increase; model participants match the primary complete-case models.

Cohort	Gene	N	Deaths	HR per SD (95% CI)	p	BH q
LUAD	IGF2BP1	471	168	1.279 (1.080–1.516)	0.004	0.027
LUAD	IGF2BP2	471	168	1.009 (0.846–1.203)	0.919	0.919
LUAD	IGF2BP3	471	168	1.106 (0.912–1.340)	0.306	0.527
LUSC	IGF2BP1	470	206	0.931 (0.802–1.081)	0.349	0.527
LUSC	IGF2BP2	470	206	1.049 (0.896–1.228)	0.553	0.664
LUSC	IGF2BP3	470	206	0.931 (0.802–1.082)	0.352	0.527

BH q values adjust the six continuous-expression gene tests as a family separate from the six primary median-split tests. These secondary results do not validate an alternative marker or establish a nonlinear threshold.

Table S7. Proportional hazards diagnostics

Entries are p values from scaled Schoenfeld-residual tests using rank and Kaplan–Meier time transformations, applied to the same fitted lifelines models. Each covariate was tested separately; no global-test result is claimed. N/deaths: LUAD 471/168; LUSC 470/206.

Cohort	Variable	Primary rank	Primary KM	Continuous rank	Continuous KM
LUAD	IGF2BP1	0.738	0.984	0.496	0.690
LUAD	IGF2BP2	0.132	0.179	0.268	0.306
LUAD	IGF2BP3	0.899	0.914	0.419	0.436
LUAD	Age >65 vs ≤65 years	0.614	0.338	0.671	0.390
LUAD	Female vs male	0.146	0.213	0.186	0.243
LUAD	Ever vs never smoker	0.874	0.850	0.982	0.744
LUAD	Stage III–IV vs I–II	0.282	0.287	0.224	0.221
LUSC	IGF2BP1	0.974	0.683	0.867	0.570
LUSC	IGF2BP2	0.497	0.626	0.394	0.497
LUSC	IGF2BP3	0.735	0.366	0.302	0.192
LUSC	Age >65 vs ≤65 years	0.425	0.535	0.406	0.514
LUSC	Female vs male	0.768	0.645	0.797	0.697
LUSC	Ever vs never smoker	0.811	0.927	0.791	0.908
LUSC	Stage III–IV vs I–II	0.323	0.334	0.292	0.309

None of these individual tests detected a departure at $p<0.05$. Nonsignificant diagnostics do not prove proportional hazards.

Table S8a. Cohort accounting

Cohort accounting	LUAD	LUSC
Source rows	494	489
Unique participant barcodes	493	489
Rows excluded with all three IGF2BPs missing	2	0
Expression cohort participants	492	489
All three IGF2BPs available	492	486
Primary complete-case model participants	471	470
Deaths in primary model	168	206
Participants excluded from primary model	21	19

LUAD: 494 source rows represent 493 participants; one expression-empty duplicate row and one other expression-empty participant were excluded. Primary-model exclusions: LUAD 13 smoking-only, 7 stage-only, and 1 with both missing; LUSC 12 smoking, 4 stage, and 3 missing one gene. Source clinical values were recovered by exact barcode linkage, not imputed.

Table S8b. Descriptive characteristics

Characteristic	LUAD cohort	LUAD model	LUSC cohort	LUSC model
N	492	471	489	470
Age mean (SD), years	65.3 (10.0)	65.2 (9.9)	67.2 (8.6)	67.4 (8.4)
Age >65 years	255 (51.8%)	241 (51.2%)	300 (61.3%)	292 (62.1%)
Female	267 (54.3%)	258 (54.8%)	127 (26.0%)	124 (26.4%)
Ever smoker	408 (82.9%)	401 (85.1%)	459 (93.9%)	452 (96.2%)
Never smoker	70 (14.2%)	70 (14.9%)	18 (3.7%)	18 (3.8%)
Smoking unknown	14 (2.8%)	0 (0.0%)	12 (2.5%)	0 (0.0%)
Stage III–IV	104 (21.1%)	100 (21.2%)	90 (18.4%)	85 (18.1%)
Stage I–II	380 (77.2%)	371 (78.8%)	395 (80.8%)	385 (81.9%)
Stage unknown	8 (1.6%)	0 (0.0%)	4 (0.8%)	0 (0.0%)

Percentages use the full column denominator, including missing values. Cohort denotes all participants with at least one IGF2BP measurement; model denotes complete cases for the seven-covariate OS model.

Table S8c. Expression thresholds

Cohort	Gene	N	Median RSEM	Log_2 mean	Log_2 SD
LUAD	IGF2BP1	492	5.100	3.6386	3.3128
LUAD	IGF2BP2	492	349.005	8.2803	1.6464
LUAD	IGF2BP3	492	213.245	7.4435	2.4265
LUSC	IGF2BP1	488	48.060	5.6976	3.2871
LUSC	IGF2BP2	488	1638.070	10.2405	1.7884
LUSC	IGF2BP3	488	946.065	9.2013	1.9797

High: >median; low: ≤median. No ties occurred. High/low counts were 246/246 in LUAD and 244/244 in LUSC for every gene. Log_2(RSEM+1) means and sample SDs specify continuous-model scaling.

Table S8d. Endpoint comparisons

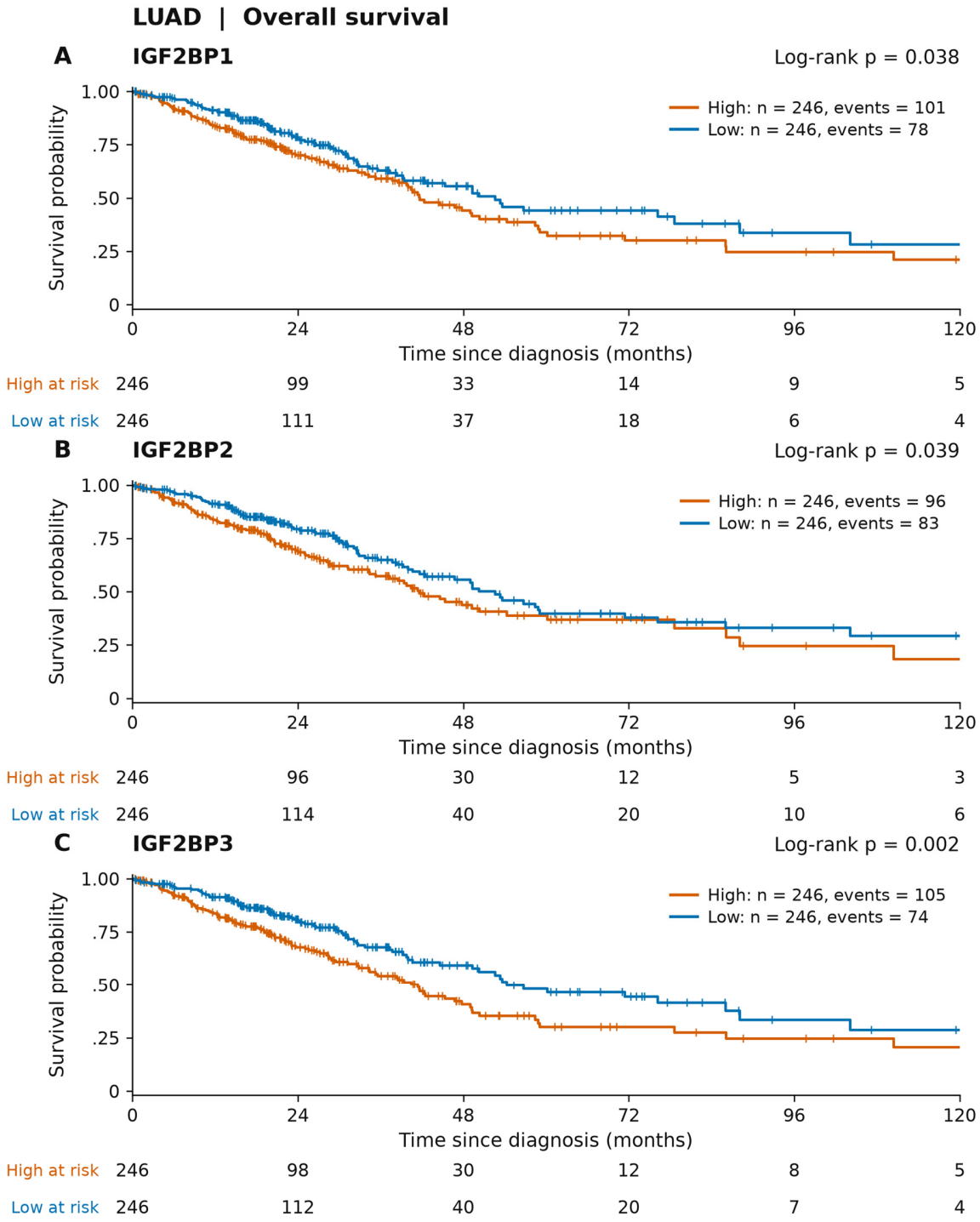
Cohort	End-point	Gene	N	Events	High/low events	Log-rank p	BH q
LUAD	OS	IGF2BP1	492	179	101/78	0.038	0.046
LUAD	OS	IGF2BP2	492	179	96/83	0.039	0.046
LUAD	OS	IGF2BP3	492	179	105/74	0.002	0.012
LUAD	PFI	IGF2BP1	492	200	100/100	0.508	0.610
LUAD	PFI	IGF2BP2	492	200	103/97	0.087	0.236
LUAD	PFI	IGF2BP3	492	200	109/91	0.022	0.133
LUSC	OS	IGF2BP1	488	211	97/114	0.029	0.046
LUSC	OS	IGF2BP2	488	212	102/110	0.304	0.304
LUSC	OS	IGF2BP3	488	212	98/114	0.039	0.046
LUSC	PFI	IGF2BP1	488	144	69/75	0.295	0.442
LUSC	PFI	IGF2BP2	488	143	74/69	0.845	0.845
LUSC	PFI	IGF2BP3	488	144	66/78	0.118	0.236

BH correction uses six comparisons per endpoint across both histologies. LUSC gene-specific samples differ by one missing expression value, so event totals can differ despite identical group sizes. Exact risk counts and estimates are in Supplementary Data S1 and Figures S1–S4.

Table S8e. Reverse Kaplan–Meier follow-up estimates

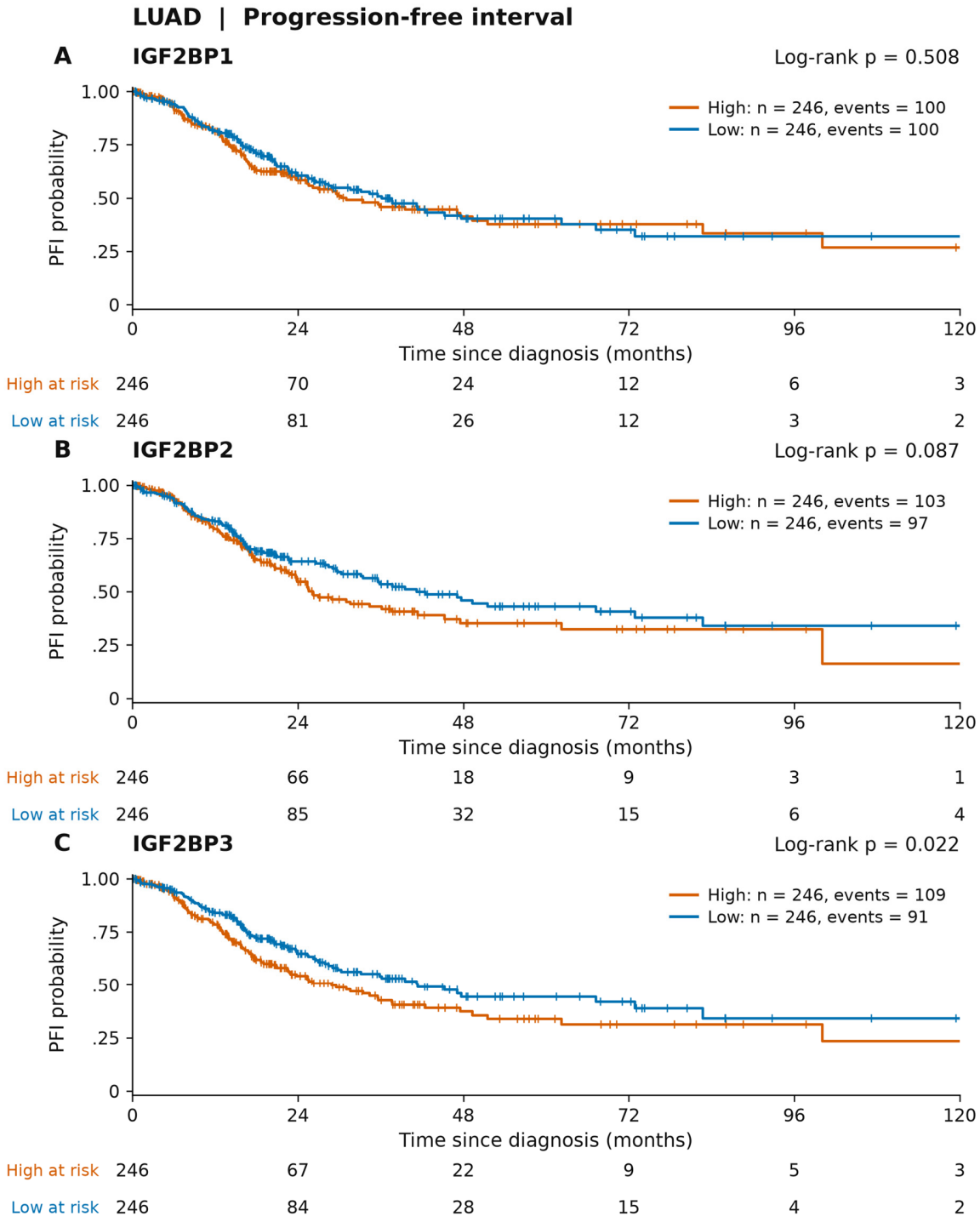
Cohort	OS cohort N	Deaths	Reverse KM follow-up days (95% CI)
LUAD	492	179	863 (791–944)
LUSC	489	212	1092 (960–1259)

Figure S1 Overall survival in LUAD



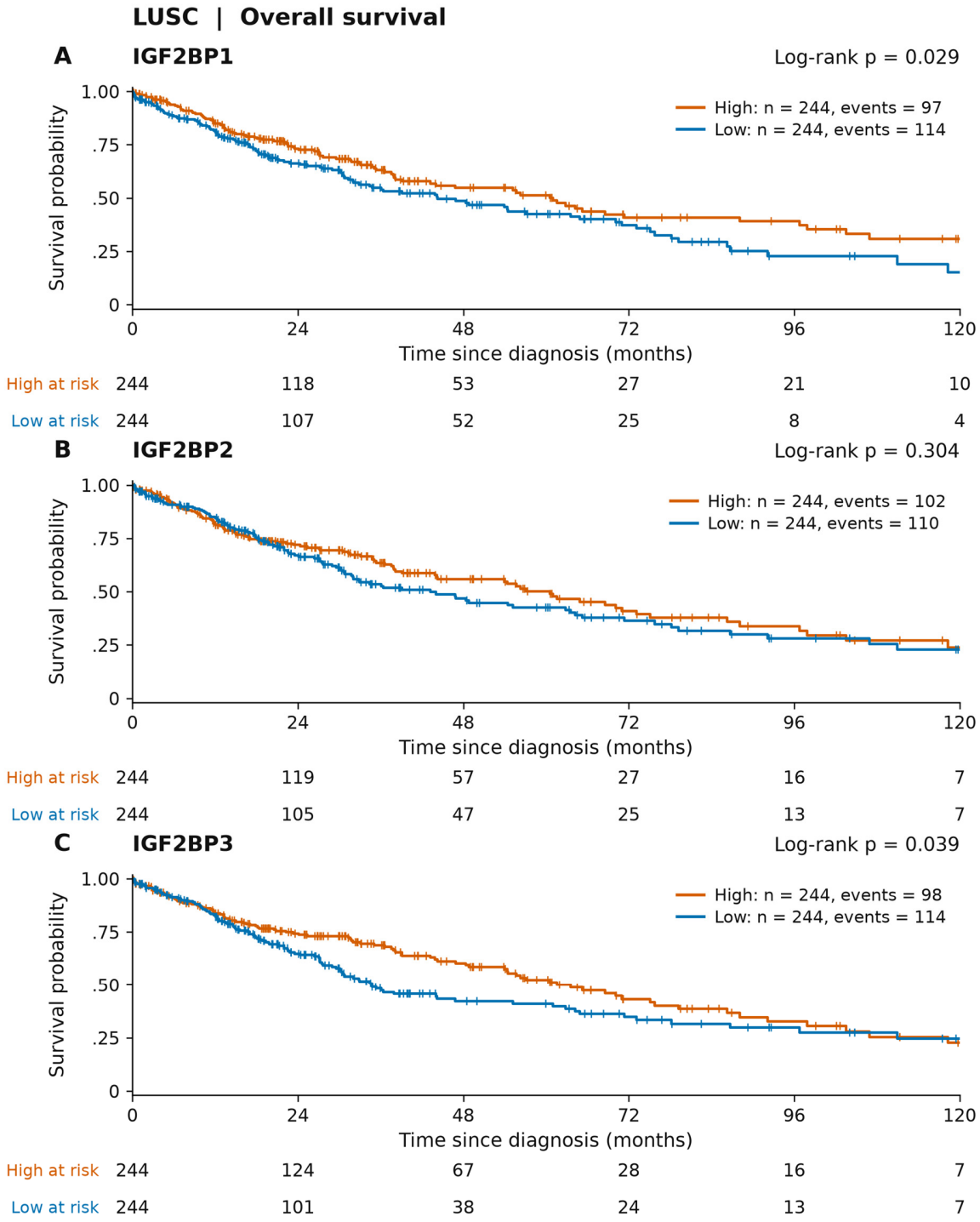
A–C: IGF2BP1, IGF2BP2, IGF2BP3. Orange: high; blue: low. Ticks indicate censoring. Group N/events and numbers at risk are shown. Display ends at 120 months; nominal log-rank tests use all follow-up. Endpoint-family BH q values are in Tables S8a–S8e. Endpoints come from TCGA-CDR.

Figure S2 Progression free interval in LUAD



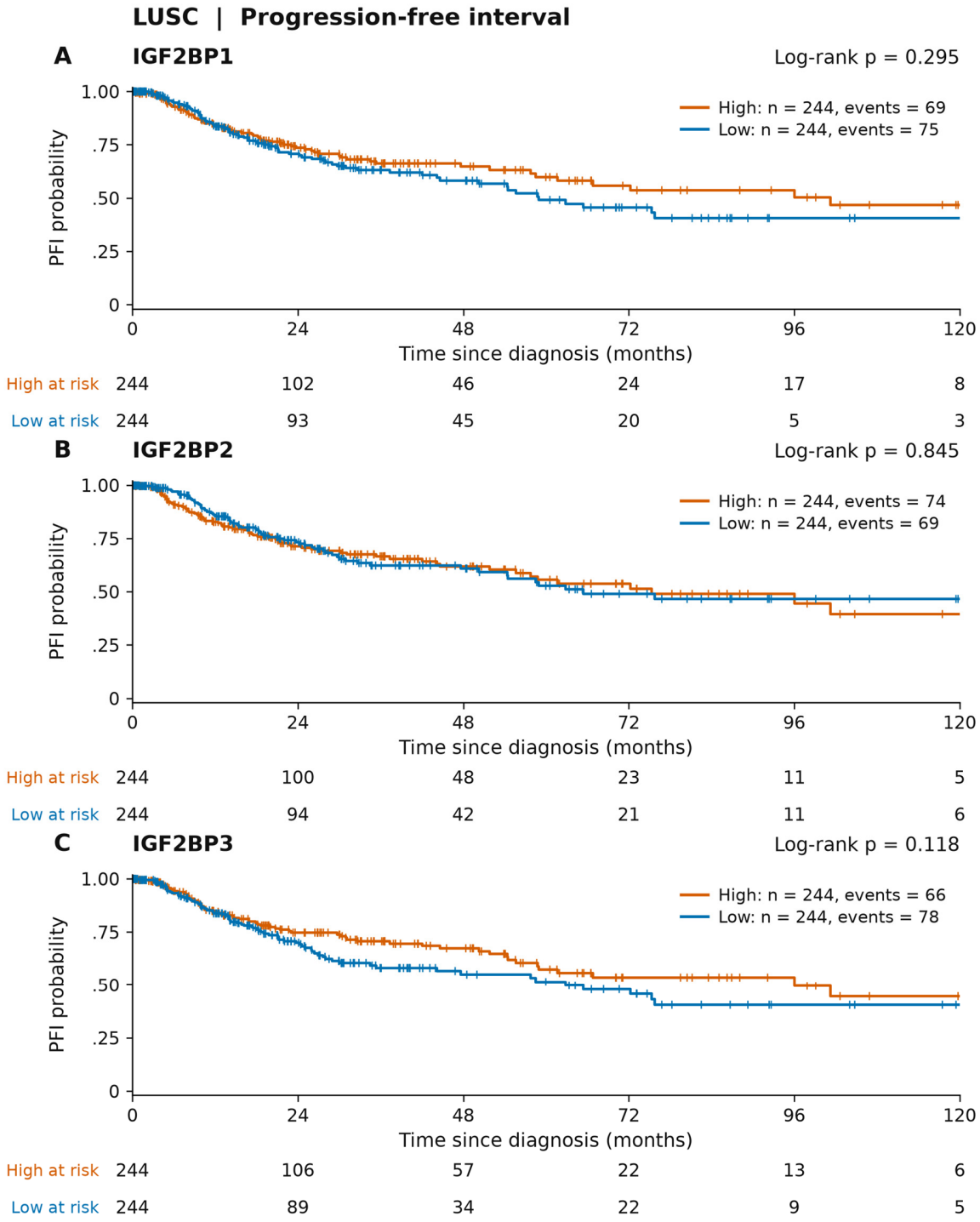
A–C: IGF2BP1, IGF2BP2, IGF2BP3. Orange: high; blue: low. Ticks indicate censoring. Group N/events and numbers at risk are shown. Display ends at 120 months; nominal log-rank tests use all follow-up. Endpoint-family BH q values are in Tables S8a–S8e. Endpoints come from TCGA-CDR.

Figure S3 Overall survival in LUSC



A–C: IGF2BP1, IGF2BP2, IGF2BP3. Orange: high; blue: low. Ticks indicate censoring. Group N/events and numbers at risk are shown. Display ends at 120 months; nominal log-rank tests use all follow-up. Endpoint-family BH q values are in Tables S8a–S8e. Endpoints come from TCGA-CDR.

Figure S4 Progression free interval in LUSC



A–C: IGF2BP1, IGF2BP2, IGF2BP3. Orange: high; blue: low. Ticks indicate censoring. Group N/events and numbers at risk are shown. Display ends at 120 months; nominal log-rank tests use all follow-up. Endpoint-family BH q values are in Tables S8a–S8e. Endpoints come from TCGA-CDR.

Table S9a. Post-review restricted cubic spline diagnostics

Each target gene was fitted with a three-knot restricted cubic spline in a separate model. The other two genes retained continuous linear terms; age >65, female sex, ever smoking, and stage III–IV were unchanged. Complete cases match the original models. Nonlinearity compares spline versus linear (1 df); overall association compares the spline versus omission of the target gene (2 df). BH adjustments use six tests separately for each family. These are post-review exploratory analyses and do not replace the primary analyses.

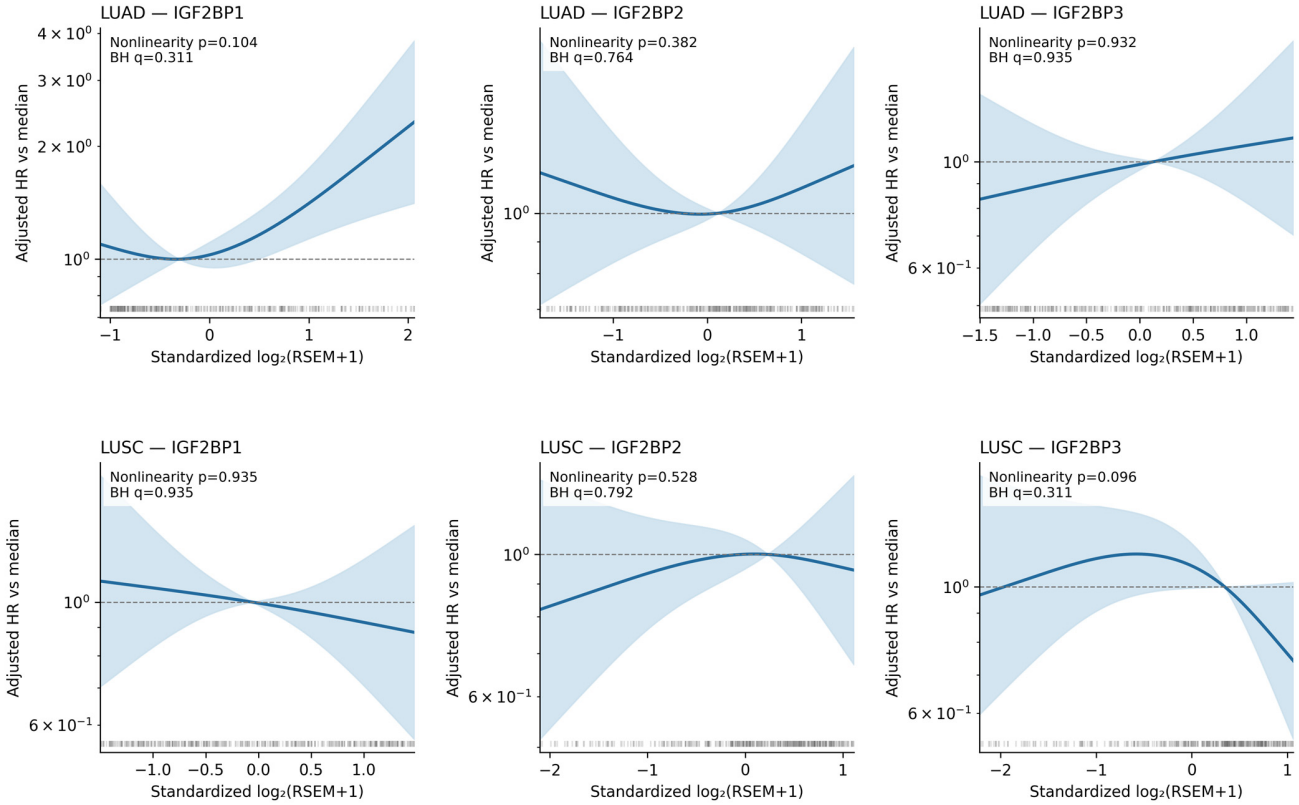
Cohort	Gene	N	Deaths	Nonlinear p	BH q	Overall p	BH q
LUAD	IGF2BP1	471	168	0.104	0.311	0.005	0.029
LUAD	IGF2BP2	471	168	0.382	0.764	0.679	0.686
LUAD	IGF2BP3	471	168	0.932	0.935	0.587	0.686
LUSC	IGF2BP1	470	206	0.935	0.935	0.643	0.686
LUSC	IGF2BP2	470	206	0.528	0.792	0.686	0.686
LUSC	IGF2BP3	470	206	0.096	0.311	0.163	0.490

Table S9b. Spline knot locations

Cohort	Gene	10th percentile z	50th percentile z	90th percentile z
LUAD	IGF2BP1	-1.098349	-0.303771	1.572740
LUAD	IGF2BP2	-1.399243	0.114965	1.190897
LUAD	IGF2BP3	-1.352373	0.129545	1.305018
LUSC	IGF2BP1	-1.320456	-0.056566	1.323539
LUSC	IGF2BP2	-1.490604	0.230009	0.959225
LUSC	IGF2BP3	-1.654975	0.345492	0.950932

z denotes standardized $\log_2(\text{RSEM}+1)$, using the original histology-specific scaling constants. Quantiles were computed in the same complete-case cohort without outcome-driven selection. For knots $a<b<c$, the additional nonlinear basis is $[(z-a)_+^3 - (c-a)/(c-b)(z-b)_+^3 + (b-a)/(c-b)(z-c)_+^3]/(c-a)^2$. The model includes z and this basis; $u_+=\max(u,0)$. All full-precision tests, coefficients, covariance-derived curves, and original scaling constants are supplied in Supplementary Data S1.

Figure S5 Post-review expression–risk shape analysis



Three-knot restricted cubic spline curves for each IGF2BP gene in LUAD (top) and LUSC (bottom). HRs are relative to the complete-case median of the target gene, conditional on the other two genes and four clinical covariates. Shading denotes pointwise 95% CIs; rugs show the observed expression distribution. Curves are displayed between the 5th and 95th percentiles. Nonlinearity p values compare spline and linear models with one additional degree of freedom; q values adjust six nonlinearity tests. No nonlinearity test met nominal $p<0.05$ or BH $q<0.05$. Displayed shapes do not identify a validated threshold, and nonsignificance does not prove linearity.

Table S10a. External cohort accounting

Post-review external association assessment. Source totals include other histologies and normal tissues; they are not analyzed tumor counts. Clinical core adjustment: age >65, female sex, all three gene z scores. Complete-case OS requires positive follow-up.

Cohort	Histology	Eligible	Core N / deaths	Expanded N / deaths
GSE50081	LUAD	127	127 / 51	115 / 46
GSE50081	LUSC	42	42 / 16	34 / 12 *
GSE30219	LUAD	85	83 / 43	83 / 43
GSE30219	LUSC	61	61 / 43	61 / 43
GSE31210	LUAD	204	204 / 30	204 / 30

GSE50081: 181 source samples; 127 pure LUAD and 42 strictly labeled LUSC retained; 12 other/ambiguous histologies excluded, including one "squamous cell carcinoma X2". GSE30219: 307 source samples (293 tumors/14 normal); ADC 85 and SQC 61 retained, other tumors excluded; 2 LUAD follow-up times of zero excluded from Cox models. GSE31210: 246 source samples; 20 normal excluded, 226 LUAD; 22 source-flagged prognosis exclusions leave 204. Expanded GSE50081 samples lose 12 LUAD and 8 LUSC for missing smoking. * LUSC expanded fit failed and its estimates are unavailable.

Table S10b. External preprocessing and clinical covariates

Cohort	Input scale and processing	Expanded clinical covariates
GSE50081	RMA log2; no further log	Stage II vs I; ever smoking
GSE30219	RMA log2; no further log	T2–4 vs T1; N1–3 vs N0
GSE31210	MAS5; log2(intensity+1)	Stage II vs I; ever smoking

GPL570 uniquely annotated probes were median-aggregated on log2 scale: IGF2BP1 223689_at, 227377_at, 241574_s_at; IGF2BP2 218847_at, 223963_s_at; IGF2BP3 1569033_at, 203819_s_at, 203820_s_at, 216493_s_at. Full eight-gene mapping and within-eligible-stratum means/sample SDs are in probe_mapping.csv and standardization.csv. No TCGA scaling or coefficient transport was performed. Source survival units were not pooled.

Table S11a. External continuous-expression Cox models core adjustment

HRs per one eligible-stratum SD of processed log2 expression. All three IGF2BPs enter simultaneously. Core: age >65 and female sex; expanded: additionally available stage/smoking or T/N variables as specified in Tables S10a–S10b. Nominal 95% CIs. Each model family has 15 intended gene tests; failed expanded tests enter the correction denominator at p=1 but have no reported estimate/p/q. Efron ties. All null results are retained.

Cohort / histology	Gene	N / deaths	HR (95% CI)	p	BH q
GSE50081 LUAD	IGF2BP1	127/51	1.633 (1.200–2.221)	0.0018	0.0134
GSE50081 LUAD	IGF2BP2	127/51	1.315 (0.974–1.775)	0.0736	0.2208
GSE50081 LUAD	IGF2BP3	127/51	0.878 (0.634–1.215)	0.4323	0.6364
GSE50081 LUSC	IGF2BP1	42/16	1.189 (0.746–1.893)	0.4667	0.6364
GSE50081 LUSC	IGF2BP2	42/16	1.393 (0.583–3.328)	0.4559	0.6364
GSE50081 LUSC	IGF2BP3	42/16	0.949 (0.554–1.626)	0.8490	0.9096
GSE30219 LUAD	IGF2BP1	83/43	0.954 (0.719–1.265)	0.7412	0.9096
GSE30219 LUAD	IGF2BP2	83/43	2.160 (1.556–2.996)	4.07e-06	6.10e-05
GSE30219 LUAD	IGF2BP3	83/43	1.184 (0.847–1.655)	0.3236	0.6067
GSE30219 LUSC	IGF2BP1	61/43	1.197 (0.885–1.619)	0.2429	0.5205
GSE30219 LUSC	IGF2BP2	61/43	0.689 (0.476–0.998)	0.0486	0.1824
GSE30219 LUSC	IGF2BP3	61/43	0.987 (0.680–1.432)	0.9432	0.9432
GSE31210 LUAD	IGF2BP1	204/30	1.210 (0.897–1.632)	0.2119	0.5205
GSE31210 LUAD	IGF2BP2	204/30	0.962 (0.651–1.422)	0.8470	0.9096
GSE31210 LUAD	IGF2BP3	204/30	1.599 (1.090–2.345)	0.0163	0.0815

Table S11b. External continuous-expression Cox models expanded adjustment

HRs per one eligible-stratum SD of processed log2 expression. All three IGF2BPs enter simultaneously. Core: age >65 and female sex; expanded: additionally available stage/smoking or T/N variables as specified in Tables S10a–S10b. Nominal 95% CIs. Each model family has 15 intended gene tests; failed expanded tests enter the correction denominator at p=1 but have no reported estimate/p/q. Efron ties. All null results are retained.

Cohort / histology	Gene	N / deaths	HR (95% CI)	p	BH q
GSE50081 LUAD	IGF2BP1	115/46	1.705 (1.204–2.412)	0.0026	0.0196
GSE50081 LUAD	IGF2BP2	115/46	1.273 (0.908–1.785)	0.1615	0.4039
GSE50081 LUAD	IGF2BP3	115/46	0.791 (0.556–1.125)	0.1922	0.4118
GSE50081 LUSC	IGF2BP1	34/12	NA	NA	NA
GSE50081 LUSC	IGF2BP2	34/12	NA	NA	NA
GSE50081 LUSC	IGF2BP3	34/12	NA	NA	NA
GSE30219 LUAD	IGF2BP1	83/43	0.965 (0.730–1.275)	0.8000	1.0000
GSE30219 LUAD	IGF2BP2	83/43	2.134 (1.546–2.947)	4.14e-06	6.21e-05
GSE30219 LUAD	IGF2BP3	83/43	1.141 (0.800–1.628)	0.4660	0.7767
GSE30219 LUSC	IGF2BP1	61/43	0.937 (0.659–1.333)	0.7186	1.0000
GSE30219 LUSC	IGF2BP2	61/43	0.732 (0.511–1.048)	0.0885	0.2658
GSE30219 LUSC	IGF2BP3	61/43	1.028 (0.698–1.512)	0.8905	1.0000
GSE31210 LUAD	IGF2BP1	204/30	1.329 (0.965–1.829)	0.0815	0.2658
GSE31210 LUAD	IGF2BP2	204/30	0.858 (0.572–1.287)	0.4593	0.7767
GSE31210 LUAD	IGF2BP3	204/30	1.407 (0.950–2.083)	0.0886	0.2658

GSE50081 LUSC: nonconvergence with sparse data and smoking separation warnings; no penalized replacement was introduced. The core result remains available. PH diagnostic departures and small event counts limit interpretation; significance in one cohort and not another is not a formal interaction test.

Table S12. Focused external LUSC coexpression

Both Pearson on processed log2 expression and Spearman rank correlation are shown. BH q uses all 36 unique pairs in the external nine-variable panel separately for each cohort, histology and method. The panel is IGF2BP1/2/3, EGFR, KRAS, TP53, SOX2, PIK3CA and age. All 360 coefficients across all five strata and both methods are supplied in external_correlations.csv; this table displays all four focused pairs, regardless of significance.

Cohort / N	Method	Pair	r / rho	p	BH q
GSE50081 / 42	Pearson	IGF2BP2 / SOX2	0.620	1.19e-05	2.14e-04
GSE50081 / 42	Pearson	IGF2BP2 / PIK3CA	0.271	0.0821	0.3400
GSE50081 / 42	Pearson	IGF2BP3 / SOX2	0.649	3.38e-06	1.22e-04
GSE50081 / 42	Pearson	IGF2BP3 / PIK3CA	0.077	0.6275	0.7530
GSE50081 / 42	Spearman	IGF2BP2 / SOX2	0.600	2.65e-05	4.77e-04
GSE50081 / 42	Spearman	IGF2BP2 / PIK3CA	0.006	0.9712	0.9712
GSE50081 / 42	Spearman	IGF2BP3 / SOX2	0.700	2.52e-07	9.08e-06
GSE50081 / 42	Spearman	IGF2BP3 / PIK3CA	-0.024	0.8798	0.9712
GSE30219 / 61	Pearson	IGF2BP2 / SOX2	0.553	3.86e-06	6.96e-05
GSE30219 / 61	Pearson	IGF2BP2 / PIK3CA	0.582	8.79e-07	3.17e-05
GSE30219 / 61	Pearson	IGF2BP3 / SOX2	0.475	1.10e-04	0.0013
GSE30219 / 61	Pearson	IGF2BP3 / PIK3CA	0.273	0.0335	0.1205
GSE30219 / 61	Spearman	IGF2BP2 / SOX2	0.675	2.35e-09	8.46e-08
GSE30219 / 61	Spearman	IGF2BP2 / PIK3CA	0.631	4.96e-08	8.92e-07
GSE30219 / 61	Spearman	IGF2BP3 / SOX2	0.449	2.81e-04	0.0020
GSE30219 / 61	Spearman	IGF2BP3 / PIK3CA	0.317	0.0127	0.0574

Table S13. Public HPA lung cancer immunohistochemistry

HPA version 26 XML, retrieved 25 November 2026. Antibody-specific aggregate staining counts for “Lung cancer”; histologies are mixed. “Detected” combines high, medium and low staining; it is a descriptive definition, not a diagnostic cutoff. Patients may overlap across antibodies. Do not pool rows or interpret their totals as unique patients.

Gene	Antibody	N	High	Medium	Low	Not detected	Detected
IGF2BP1	HPA021367	12	0	1	0	11	1/12
IGF2BP1	HPA062273	12	0	2	0	10	2/12
IGF2BP2	CAB017126	9	3	1	2	3	6/9
IGF2BP2	HPA035145	11	0	10	1	0	11/11
IGF2BP3	HPA076951	12	5	2	3	2	10/12

Sources: IGF2BP1 ENSG00000159217; IGF2BP2 ENSG00000073792; IGF2BP3 ENSG00000136231, https://www.proteinatlas.org/. Complete source XML records and SHA-256 hashes are archived. Antibody-dependent differences, limited sample size, mixed histology, and absence of matched OS/transcript data prevent prognostic, quantitative transcript–protein or mechanistic validation. HPA survival analyses were not used. No new IHC experiment was performed.

Table S14. External proportional hazards diagnostic flags

Scaled Schoenfeld-residual tests on all estimable models, with rank and Kaplan–Meier (KM) time transforms. This table shows every covariate/transform with nominal $p<0.05$; all diagnostic p values, including nonsignificant values, are in external_PH_diagnostics.csv. These multiple descriptive tests were not BH-adjusted, do not establish a mechanism, and caution against interpreting HRs as proven constant effects. No global PH test is claimed.

Cohort / histology	Model	Transform	Variable	p
GSE30219 LUAD	core	rank	IGF2BP1_z	0.0270
GSE30219 LUAD	core	km	IGF2BP1_z	0.0104
GSE30219 LUAD	expanded	rank	IGF2BP1_z	0.0328
GSE30219 LUAD	expanded	km	IGF2BP1_z	0.0143
GSE30219 LUAD	expanded	km	IGF2BP3_z	0.0326
GSE30219 LUSC	expanded	rank	IGF2BP1_z	0.0446
GSE31210 LUAD	core	rank	IGF2BP2_z	0.0054
GSE31210 LUAD	core	rank	female	0.0300
GSE31210 LUAD	core	km	IGF2BP2_z	0.0067
GSE31210 LUAD	core	km	female	0.0181
GSE31210 LUAD	expanded	rank	IGF2BP2_z	0.0012
GSE31210 LUAD	expanded	rank	stage_II	0.0044
GSE31210 LUAD	expanded	km	IGF2BP2_z	0.0025
GSE31210 LUAD	expanded	km	stage_II	0.0210
GSE50081 LUAD	expanded	rank	IGF2BP3_z	0.0469
GSE50081 LUAD	expanded	rank	ever_smoker	0.0154
GSE50081 LUAD	expanded	km	ever_smoker	0.0384

The GSE50081 LUSC expanded model was unestimable and has no PH diagnostic. Small event counts (as few as 16 in an estimable core model) and model complexity limit power and precision. Independent Efron score/information and BH calculations verified the nine estimable external Cox fits; the verification report is included in Supplementary Data S1.