

Comparative Evolutionary and Epigenetic Analysis of the Insulin-Like Growth Factor 2 P1 and P4 Promoters in Hepatocellular Carcinoma

Sailaubek Nurislam
nuriksailaubek5@gmail.com

ABSTRACT

Insulin-like growth factor 2 (IGF2) is transcribed from five promoters with distinct developmental and tissue-specific activities. P1 is the dominant promoter in healthy adult liver, whereas P4 is predominantly active during fetal development and is largely silent in adult liver. Because hepatocellular carcinoma (HCC) can reactivate fetal IGF2 transcriptional programs, promoter-specific regulation may be obscured by gene-level measurements. This study tested whether the adult-liver P1 and fetal P4 promoters represent distinct regulatory states by integrating comparative sequence analysis, CpG-island architecture, predicted transcription-factor binding, regulatory-locus annotation, and methylation-expression relationships in TCGA-LIHC. Overall sequence conservation was similar for P1 and P4, but their regulatory architecture and methylation-expression behavior differed. P4 retained CpG-island architecture across the examined primates and showed an inverse association between methylation and IGF2 expression, consistent with oncofetal reactivation. In contrast, P1 lacked a CpG island despite 68% GC content and showed a positive association between methylation and both total IGF2 expression and expression of the P1-specific first exon; this association persisted within HCC. These findings indicate that P1 and P4 exhibit promoter-specific epigenetic behavior in HCC that is not captured by a uniform promoter-methylation model. Independent validation and multivariable analysis are needed to determine the mechanisms underlying the unexpected P1 relationship.

Keywords: DNA methylation, hepatocellular carcinoma, IGF2, oncofetal reactivation, promoter evolution

I. INTRODUCTION

IGF2 encodes insulin-like growth factor II, a mitogenic growth factor closely related to IGF1. The human IGF2 locus spans approximately 30 kb on chromosome 11p15.5 and contains multiple non-coding leader exons upstream of the shared protein-coding exons. Five promoters—P1, P0, P2, P3, and P4—precede leader exons 1, 2, 4, 6, and 7, respectively, and generate transcripts with different 5' untranslated regions

[1]. Most promoters are active during fetal development, whereas P1 becomes the dominant promoter in the healthy adult liver.

IGF2 belongs to a small group of genes subject to genomic imprinting, in which expression depends on the parental origin of the allele. Under normal conditions, the paternal copy is active and the maternal copy is silenced [2]. This parent-of-origin control is maintained by DNA methylation. At the neighboring H19/IGF2 imprinting-control region, the unmethylated maternal allele binds CCCTC-binding factor (CTCF) and forms an insulator that restricts IGF2 promoters from accessing shared enhancers. Methylation of the paternal allele prevents this insulator configuration and permits paternal IGF2 expression. Alteration of this regulatory architecture can therefore change both allelic control and total IGF2 output.

Hepatocellular carcinoma (HCC) frequently shows an oncofetal shift in IGF2 transcription. P1 activity is reduced in a substantial proportion of tumors, while fetal promoters—especially P3, but also P0, P2, and P4—can be reactivated [3]. Fetal-promoter activation has also been observed in chronic hepatitis C and cirrhosis, indicating that the change can arise during chronic liver injury before overt malignancy [4]. Loss of parent-specific methylation at the IGF2 locus is common in HCC [5]. Targeted methylation of the P4 region reduced IGF2 expression in an experimental HCC model [6], although the TCGA-LIHC analyses in the present study are observational.

P1 and P4 may also differ in their regulatory inputs. CCAAT/enhancer-binding protein alpha (CEBPA) has been experimentally linked to liver-specific activation of P1 [7], and specificity protein 1 (SP1) is required for activity of the adult liver-specific P1 promoter [8]. CTCF regulates the wider H19/IGF2 locus through chromatin organization [9]. These three factors were selected for motif screening because they represent liver-specific promoter activation, binding to GC-rich promoter DNA, and higher-order locus regulation, respectively. Existing HCC literature has focused more heavily on P3 than on direct comparison of P1 and P4. Because P1 and P4 represent adult-liver and fetal transcriptional programs, respectively, we hypothesized that their divergent behavior in HCC would be accompanied by distinct regulatory architecture and methylation-expression relationships. We therefore integrated comparative sequence analysis, regulatory annotation, and TCGA-LIHC methylation-expression data to test whether P1 and P4 represent distinct regulatory states rather than interchangeable contributors to total IGF2 expression.

II. MATERIALS AND METHODS

2.1 Promoter Definition and Coordinate Harmonization

Promoter-specific transcription start sites (TSSs) were defined using National Center for Biotechnology Information (NCBI) Gene (Gene ID 3481), the P1-specific RefSeq transcript NM_001007139.6, the P4-specific transcript NM_001291861.3, RefSeqGene NG_008849.1, and the GRCh38 primary assembly.

Coordinates were converted to GRCh37/hg19 with the University of California, Santa Cruz (UCSC) LiftOver tool. The resulting hg19 TSS positions were chr11:2,170,796 for P1 and chr11:2,158,518 for P4. A symmetric 1,001-bp window centered on each verified TSS (± 500 bp) was used as the operational promoter region in downstream analyses. Promoter nomenclature follows [1].

2.2 Comparative Sequence and CpG-Island Analysis

Human P1 and P4 promoter-window sequences and orthologous or annotated comparator regions from chimpanzee (*Pan troglodytes*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), chicken (*Gallus gallus*), and zebrafish (*Danio rerio*) were retrieved from Ensembl. This panel spans a close primate, two commonly used rodent models, and two more distant vertebrate outgroups. Mouse and rat were both retained so that a shared rodent pattern could be distinguished from a species-specific result. P3-associated regions were included as fetal-promoter comparators, not as experimental controls, where they provided developmental-promoter context for P4. Pairwise alignments were performed in Jalview, and sequence identity was calculated for each human-promoter/comparator pair. Because the comparator regions are not all functionally equivalent promoters, identity percentages were treated as descriptive measures of sequence similarity rather than direct measures of conserved promoter function. Nucleotide BLAST (BLASTn) searches were also performed against *Pan troglodytes*, *Mus musculus*, and *Bos taurus* nucleotide databases using default parameters. Cattle was added as a non-rodent placental-mammal comparator to test whether strong matches extended beyond primates and rodents. Representative hits were evaluated by identity, E-value, query coverage, and genomic annotation.

CpG islands were identified with EMBOSS Cpgplot using default parameters. Island presence, length, and position relative to each analyzed sequence window were recorded. Cross-species comparison was used to assess whether promoter-associated CpG architecture was preserved, but the results were interpreted descriptively because island calls depend on sequence boundaries and threshold settings and do not by themselves establish conserved promoter function.

2.3 Predicted Binding Sites and Candidate Regulatory Loci

The analyzed sequences were scanned with the JASPAR motif database at an 80% relative profile-score threshold for predicted CEBPA, SP1, and CTCF binding sites. These motif matches were treated as predictions; no chromatin immunoprecipitation sequencing (ChIP-seq) or electrophoretic mobility shift assay (EMSA) was performed to validate protein-DNA binding in this study. Published experimental evidence was evaluated separately to distinguish direct promoter binding from higher-order chromatin regulation.

Three regulatory variants were annotated with RegulomeDB and the Encyclopedia of DNA Elements (ENCODE). RegulomeDB integrates functional evidence to score the regulatory potential of genetic

variants, whereas ENCODE provides experimentally mapped protein-binding and chromatin-state annotations. These variants were selected because they represent high-confidence regulatory loci within the broader IGF2 regulatory landscape and allow comparison of promoter-proximal P4-associated loci with a distal IGF2-locus comparator. rs10840452 and rs2000993 lie approximately 1.24 kb and 1.13 kb upstream of the P4 TSS, respectively; they are therefore described as P4-associated proximal regulatory loci, not as variants inside the ± 500 -bp P4 promoter window. rs3741208 lies approximately 22.3 kb upstream of the P1 TSS. Because it is far outside the ± 500 -bp P1 window, it was used only as a distal IGF2-locus comparator.

ENCODE peak counts represent the total number of overlapping ChIP-seq records for each variant; Figures 1–3 display only the top contributing transcription factors and tissue/cell-line categories and therefore do not sum to these totals.

2.4 TCGA-LIHC Methylation and Expression Analysis

Illumina HumanMethylation450 beta values for TCGA-LIHC were obtained through the University of California, Santa Cruz (UCSC) Xena browser. The methylation comparison included 377 primary tumors and 50 TCGA solid-tissue normal samples, referred to here as adjacent non-tumor liver samples. CpG probes overlapping the predefined P1 and P4 promoter windows were selected, and tumor–non-tumor differences were tested separately for 12 probe-level comparisons using the Mann–Whitney U test. P values from this family of 12 tests were adjusted with the Benjamini–Hochberg false discovery rate (FDR) procedure. Levene’s test was used to compare the dispersion of probe-level methylation beta values between HCC and adjacent non-tumor groups.

Gene-level IGF2 expression from RNA sequencing (RNA-seq) was obtained from the Xena IlluminaHiSeq dataset as $\log_2(\text{RSEM} + 1)$, where RSEM denotes RNA-Seq by Expectation-Maximization. Of the samples with methylation data, 371 HCC and 41 adjacent non-tumor samples had matched gene-level expression data. For promoter-level correlation analyses, mean promoter methylation was calculated for each sample by averaging beta values across all available CpG probes within the predefined P1 or P4 window. Spearman correlation was used to assess the association of mean P1 and mean P4 methylation with gene-level IGF2 expression in the pooled cohort and separately within HCC and adjacent non-tumor liver samples. The two primary pooled promoter-level correlations (P1 and P4) formed a separate hypothesis family and were evaluated against a Bonferroni-adjusted significance threshold of 0.025. Exon-level expression for the P1-specific first exon (chr11:2,170,356–2,170,833, hg19) was analyzed as an additional promoter-specific outcome; an equivalent P4-specific exon-level window was not available in the assembled dataset. Ninety-five percent confidence intervals were estimated using 5,000 bootstrap resamples.

2.5 Clinical Covariates and Statistical Scope

Within HCC samples, the pathologic stage was grouped as early (Stage I–II) or late (Stage III–IV). Stage associations were tested using the Mann–Whitney U test. Viral-hepatitis information was available only as an incomplete free-text serology field. Because the entries did not provide mutually exclusive confirmed infection categories, viral status was treated as a data-quality limitation and was not used for inferential analysis.

All analyses were performed in Python using pandas, scipy.stats, and statsmodels, with $p < 0.05$ as the nominal significance threshold unless a multiple-testing correction is specified. Benjamini–Hochberg FDR correction was applied to the 12 probe-level tumor–non-tumor methylation comparisons, whereas the two primary pooled promoter-level methylation–expression correlations used a Bonferroni-adjusted threshold of 0.025. The pathologic-stage analysis was exploratory, did not include a multivariable model, and its reported p value was not adjusted for multiple testing. Building a full model with stage, tumor purity, cell composition, and technical covariates was outside the scope of this study.

III. RESULTS

3.1 Comparative Sequence Analysis and CpG-Island Architecture

A CpG island was detected in human P4, chimpanzee P3 and P4, and chicken P1, whereas none was detected in human P1, the tested rodent regions, or zebrafish P1 (Table 1). The detected CpG islands ranged from 221 to 489 bp and were located at different positions relative to the analyzed promoter windows. Human P1 was notable because it lacked a CpG island despite 68% GC content. The pattern indicates that GC richness alone was insufficient to meet the CpG-island criteria and that predicted CpG-island architecture varied across the analyzed promoter-associated regions.

Table 1: CpG-island presence, length, and position in analyzed sequence windows. P3 regions are included as fetal-promoter comparators rather than experimental controls.

Species / region	Island	Length (bp)	Position
Human P1	No	–	–
Human P4	Yes	489	47–535
Chimpanzee P3	Yes	221	723–943
Chimpanzee P4 (island 1)	Yes	411	64–474
Chimpanzee P4 (island 2)	Yes	247	699–945
Chicken P1	Yes	223	602–824
Mouse P0	No	–	–

Species / region	Island	Length (bp)	Position
Mouse P3	No	–	–
Rat P1	No	–	–
Rat P3	No	–	–
Zebrafish P1	No	–	–

Mean pairwise identity was 48.8% for the human P1 window and 50.4% for the human P4 window (Table 2). Thus, the alignment did not indicate that either promoter was globally more conserved. Because the comparator regions are not all functionally equivalent promoters, these identity percentages were interpreted descriptively rather than as direct measures of conserved promoter function. Sequence conservation alone therefore did not distinguish P1 from P4. In contrast, their CpG-island architecture differed, identifying promoter structure rather than global sequence identity as a more informative axis of comparison.

Table 2: Pairwise sequence identity of human P1 and P4 windows to comparator sequences (Jalview).

Comparator sequence	P1 (%)	P4 (%)
Chimpanzee P4	53.25	51.70
Chimpanzee P3	46.11	53.99
Rat P1	45.32	43.18
Rat P3	48.91	48.96
Chicken P1	52.53	52.40
Mouse P0	53.99	48.23
Mouse P3	49.62	66.36
Zebrafish P1	40.43	38.38

3.2 Predicted Binding Sites and Regulatory Loci

Table 3: Predicted transcription-factor binding sites (JASPAR; 80% threshold).

Region	Factor	Score	Inside island
Human P1	CEBPA	0.837	No
Human P1	SP1	0.818	No
Human P4	CTCF	0.807	No
Human P4	SP1	0.966	No
Human P4	CEBPA	0.917	No
Mouse P3	SP1	1.000	No
Mouse P3	CEBPA	0.871	No
Chimpanzee P3	CEBPA	0.884	No
Chimpanzee P3	SP1	0.867	No

Predicted CEBPA and SP1 sites were identified in both human promoter windows, and a predicted CTCF sequence match was identified in P4 (Table 3). None of the listed sites fell inside the detected CpG-island boundaries. These motif matches are predictions and do not confirm protein binding. Published evidence supports direct promoter binding by CEBPA and SP1 at P1, whereas CTCF primarily regulates the wider H19/IGF2 locus through chromatin organization (Table 4).

The two P4-associated loci had similar ENCODE profiles and lie approximately 1.1-1.2 kb upstream of the P4 TSS, placing them in the same promoter-proximal regulatory neighborhood. By contrast, rs3741208 lies 22.3 kb upstream of P1 and was retained only as a distal IGF2-locus comparator. These annotations were therefore used to characterize regulatory context rather than to make a direct positional comparison between P1- and P4-proximal variants.

3.3 Regulatory Landscape of the Candidate Loci

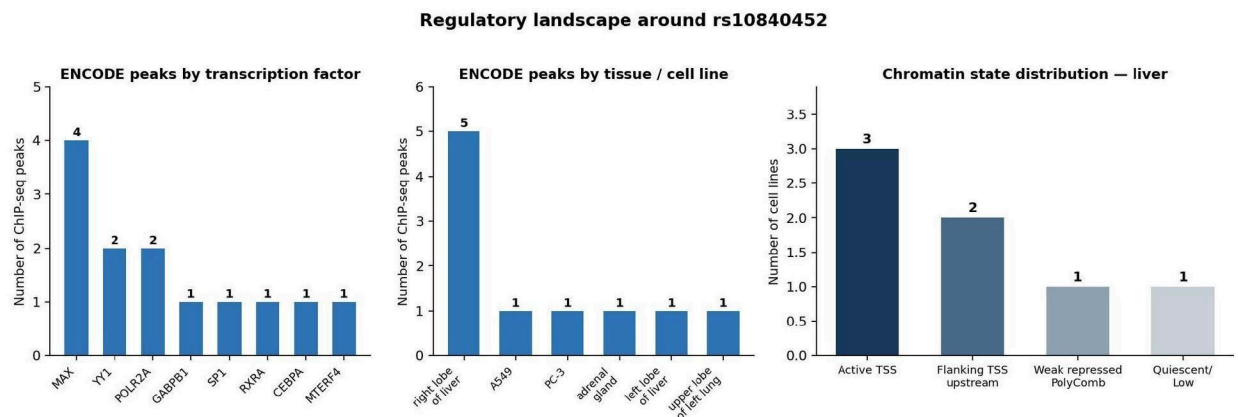


Figure 1: ENCODE transcription-factor, tissue/cell-line, and liver chromatin-state summaries around rs10840452. A 'ChIP-seq peak' indicates a site where a protein was experimentally detected bound to DNA in a given cell type; chromatin states (e.g., 'Active TSS', 'Flanking TSS upstream') are standardized labels describing the likely regulatory function of a genomic region based on combined epigenomic data.

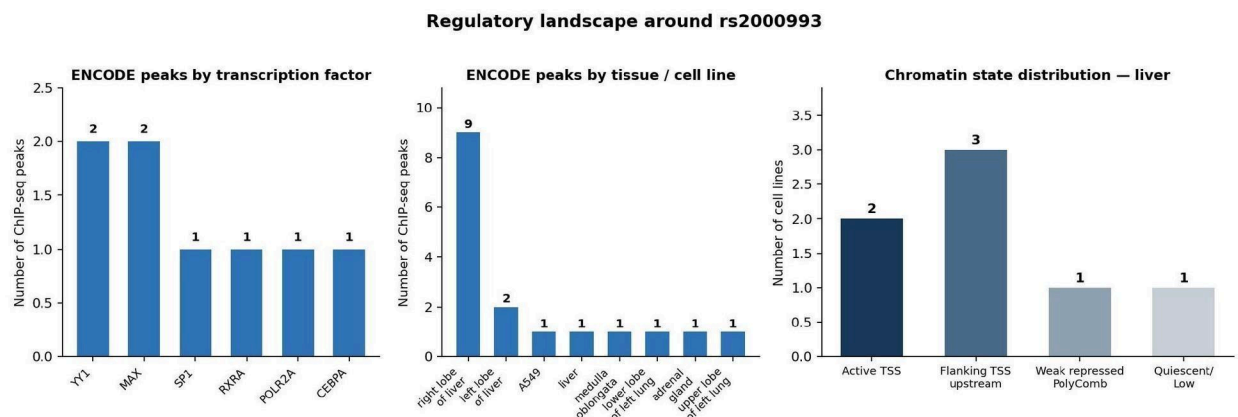


Figure 2: ENCODE transcription-factor, tissue/cell-line, and liver chromatin-state summaries around rs2000993.

rs10840452 and rs2000993 were enriched for signals involving active-transcription factors, including MAX, YY1, and POLR2A, and showed substantial representation in liver-related samples (Figures 1–2). Active or flanking TSS annotations were prominent in the liver chromatin summaries. These ENCODE annotations suggest that both variants reside within a transcriptionally active regulatory environment consistent with promoter-proximal activity. Their similar profiles are also consistent with the variants belonging to the same P4-associated regulatory neighborhood.

The rs3741208 profile differed from the two P4-associated loci: EZH2 and RAD21 were prominent, the detected tissue/cell-line signals were not liver-dominated, and Polycomb-associated states were represented (Figure 3). Because the locus is distal to P1, this contrast describes heterogeneity within the broader IGF2 region and should not be used to infer the chromatin state of the P1 promoter itself. Coordinates and RegulomeDB/ENCODE summary statistics for all three loci are listed in Table 5.

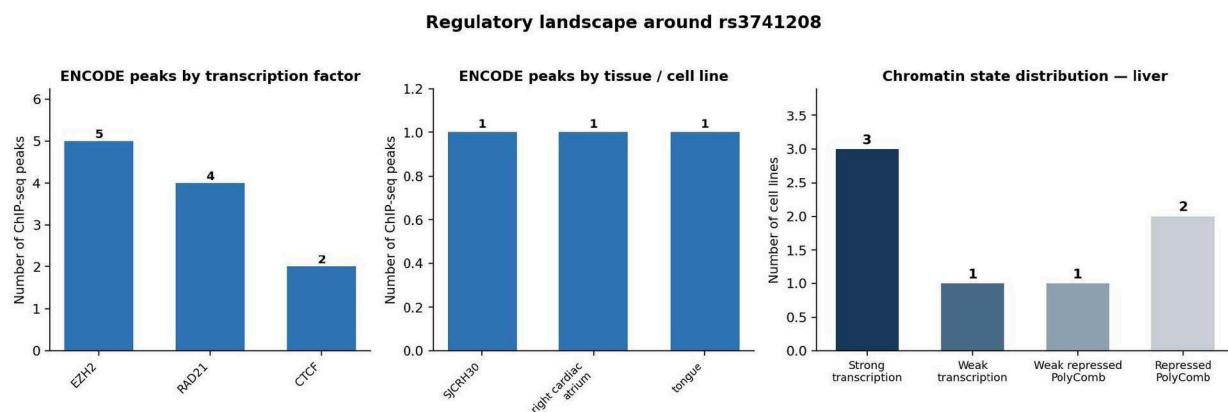


Figure 3: ENCODE summary around rs3741208. This locus is approximately 22.3 kb from P1 and is shown as a distal comparator, not a P1-proximal variant.

Table 4: Published evidence for selected regulatory factors at the IGF2 locus.

Factor	Direct binding?	Interpretation
CEBPA	Yes	Direct P1 binding and liver-specific activation [7].
SP1	Yes, at P1	A functional SP1-binding site is required for activity of the adult liver-specific P1 promoter; the P4 site identified here remains computationally predicted [8].
CTCF	Not established here	Binds the H19/IGF2 imprinting-control region and regulates the locus through chromatin organization [9].

Table 5: Candidate regulatory loci (hg19 coordinates).

Variant	Relationship to promoter	RegulomeDB rank	ENCODE peaks
rs10840452	P4-associated; ~1.24 kb upstream	1f	101
rs2000993	P4-associated; ~1.13 kb upstream	1f	115
rs3741208	Distal to P1; ~22.3 kb upstream	1f	274

Together, the sequence, CpG, motif, and regulatory-locus analyses showed that P1 and P4 cannot be distinguished by global sequence conservation alone, but they occupy different regulatory contexts. This motivated the central functional test: whether promoter methylation at P1 and P4 shows distinct relationships with IGF2 expression in HCC.

3.4 DNA Methylation, Expression, and Covariate Checks

IGF2 promoter-specific hypermethylated/hypomethylated CpGs: HCC vs. normal liver (TCGA-LIHC)

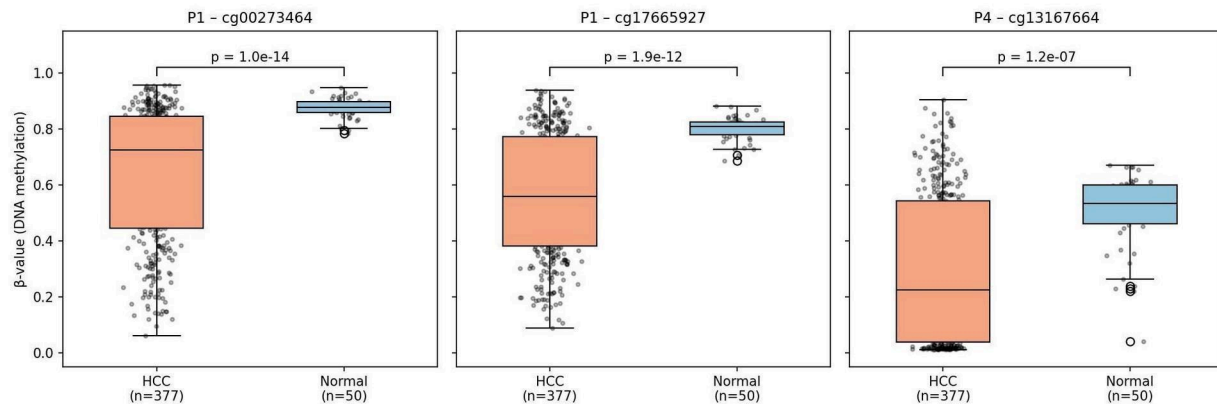


Figure 4: IGF2 promoter-specific differential methylation in HCC versus adjacent non-tumor liver (TCGA-LIHC). Methylation beta values at the three most differentially methylated promoter-window CpG probes are shown for HCC (n = 377) and adjacent non-tumor liver (n = 50). “Normal” in the plot denotes TCGA adjacent non-tumor liver samples.

The three most differentially methylated probes—P1 cg00273464, P1 cg17665927, and P4 cg13167664—were hypomethylated in HCC relative to adjacent non-tumor liver (Figure 4). HCC samples also showed substantially greater dispersion of probe-level methylation beta values than adjacent non-tumor samples (Levene’s test, $p < 1 \times 10^{-9}$).

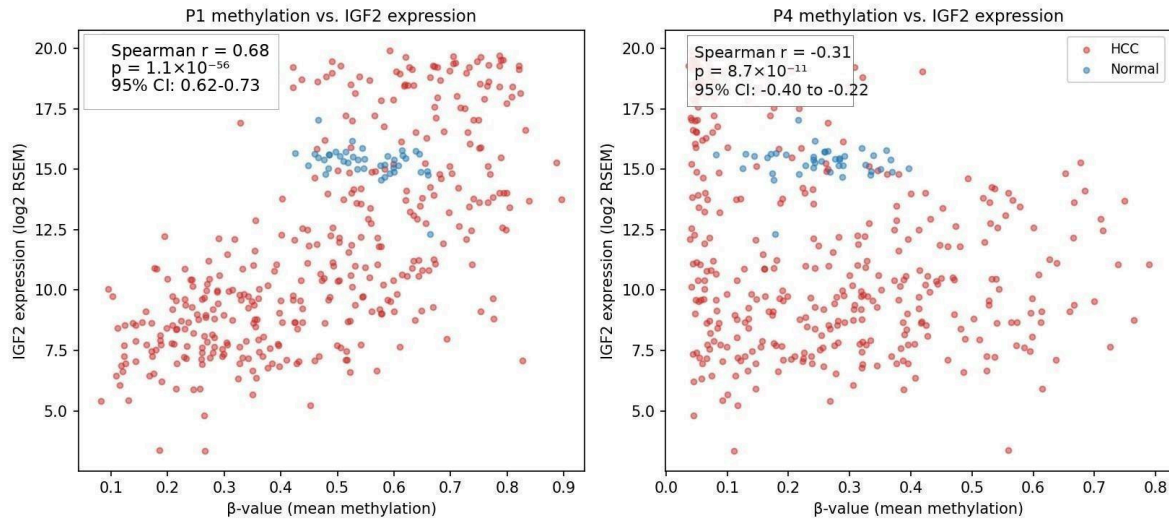


Figure 5: IGF2 gene-level expression versus mean P1 and P4 methylation in matched TCGA-LIHC samples. “Normal” denotes TCGA adjacent non-tumor liver samples. The legend in the right panel applies to both panels.

Across 412 samples with matched methylation and gene-level expression data, mean P4 promoter methylation was inversely associated with IGF2 expression (Spearman $\rho = -0.31$, $p = 8.7 \times 10^{-11}$, 95% CI -0.40 to -0.22). Mean P1 promoter methylation showed the opposite pattern, with a positive association with IGF2 expression ($\rho = 0.68$, $p = 1.1 \times 10^{-56}$, 95% CI 0.62 - 0.73) (Figure 5). These were the two primary promoter-level correlation tests and were evaluated against a Bonferroni-adjusted significance threshold of 0.025; both remained significant. The opposing directions identified P1 and P4 as distinct methylation-expression states rather than a single uniform promoter-response pattern.

Stratified analyses showed that both promoter-level associations persisted within HCC. Among 371 HCC samples, P1 methylation remained positively associated with IGF2 expression ($\rho = 0.691$, 95% CI 0.634 - 0.738 , $p = 5.71 \times 10^{-54}$), whereas P4 methylation remained inversely associated with expression ($\rho = -0.315$, 95% CI -0.410 to -0.212 , $p = 5.61 \times 10^{-10}$). In the smaller adjacent non-tumor subset ($n = 41$), the P1 association was negative ($\rho = -0.331$, 95% CI -0.602 to -0.009 , $p = 0.034$), whereas the P4 association was weak and not significant ($\rho = -0.126$, $p = 0.434$). Thus, the opposing HCC associations were not produced solely by pooling tumor and adjacent non-tumor samples. The adjacent non-tumor estimates remain uncertain because of the smaller sample size.

Positive associations were also observed between methylation and expression of the P1-specific first exon: $\rho = 0.34$ for cg00273464 ($p = 9.1 \times 10^{-13}$) and $\rho = 0.35$ for cg17665927 ($p = 1.6 \times 10^{-13}$) (Figure 6). This promoter-specific outcome supports the interpretation that the positive P1 association was not driven entirely by transcripts initiated from other IGF2 promoters. Within HCC, the associations remained significant but were smaller: $\rho = 0.211$ for cg00273464 ($p = 4.22 \times 10^{-5}$) and $\rho = 0.241$ for cg17665927 ($p = 2.62 \times 10^{-6}$). Neither association was significant in the adjacent non-tumor subset.

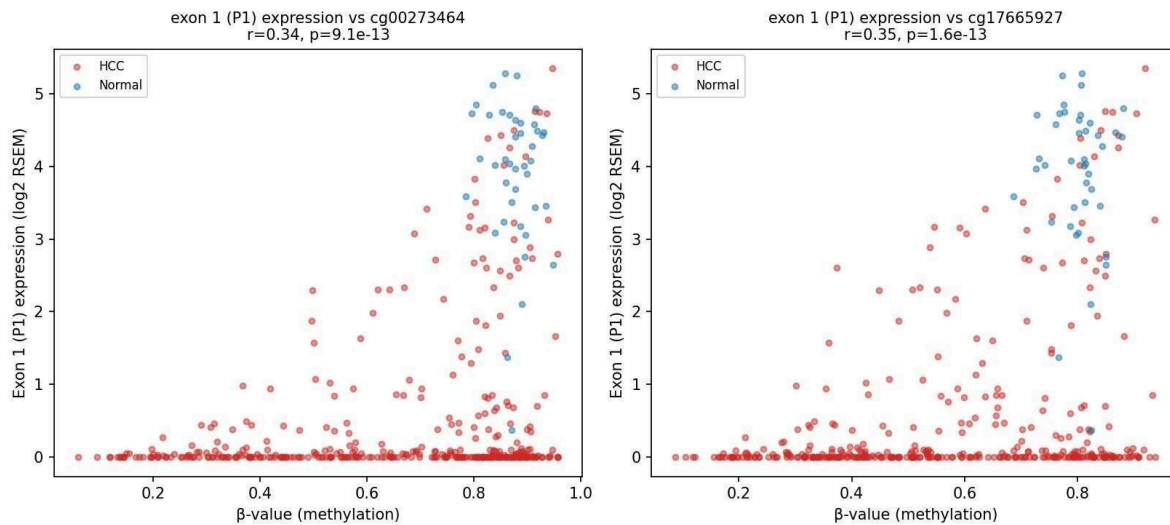


Figure 6: P1-specific exon 1 expression versus methylation at cg00273464 and cg17665927. “Normal” denotes TCGA adjacent non-tumor liver samples.

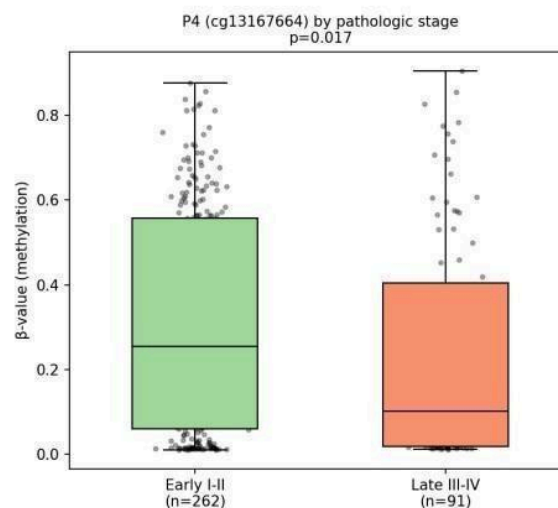


Figure 7: Exploratory HCC-only comparison of P4 cg13167664 methylation by pathologic stage.

P4 cg13167664 methylation was lower in late-stage than in early-stage HCC ($p = 0.017$) (Figure 7). This was an exploratory single-probe stage comparison; the p value is nominal and was not adjusted for multiple testing. The result suggests that pathologic stage may be associated with methylation at this probe, although it was not tested in a multivariable model.

Table 6: Representative BLASTn hits for the P1 and P4 windows.

Window	Species	Match	Identity; E-value
P4	<i>Bos taurus</i>	IGF2 coding sequence	77.48%; 3×10^{-93}
P4	<i>Mus musculus</i>	IGF2 exon 3	79.82%; 4×10^{-77}
P4	<i>Mus musculus</i>	IGF2 precursor	79.51%; 2×10^{-75}
P4	<i>Pan troglodytes</i>	Transcript X4	100.00%; 3×10^{-53}
P4	<i>Bos taurus</i>	Transcript 7	88.18%; 2×10^{-26}
P1	<i>Mus musculus</i>	Peg8/Igf2as mRNA	74.75%; 6×10^{-71}
P1	<i>Mus musculus</i>	Igf2os lncRNA	74.75%; 6×10^{-71}
P1	<i>Mus musculus</i>	DMR0	74.71%; 6×10^{-71}
P1	<i>Mus musculus</i>	IGF2AS lncRNA	74.75%; 6×10^{-71}
P1	<i>Mus musculus</i>	Pseudo-exon 1	85.41%; 6×10^{-46}

BLASTn produced strong P4-window matches in chimpanzee, mouse, and cattle, including an annotated mouse

IGF2 exon 3 region (Table 6). P1-window matches in mouse more often corresponded to antisense transcripts, the DMR0 differentially methylated region, or other sequences surrounding the imprinted locus than to exon 1 itself. The identified exon 1 match was annotated as a pseudo-exon and had lower query coverage than the surrounding matches. This pattern is consistent with greater lineage-specific divergence of the P1-associated exon sequence in rodents, while the wider imprinted locus remains conserved.

IV. DISCUSSION

The central finding of this study is that P1 and P4 exhibit opposing methylation-expression relationships in HCC despite similar overall sequence conservation. P4 followed the expected oncofetal pattern: methylation was reduced in tumors, and lower methylation was associated with higher IGF2 expression. This interpretation is consistent with experimental evidence that targeted methylation of the P4 region can suppress IGF2 expression [6]. The P4 result therefore connects its fetal regulatory identity with a canonical inverse methylation-expression relationship in HCC.

P1 behaved differently. Its promoter window lacked a CpG island despite high GC content, and higher methylation was positively associated with both total IGF2 expression and expression of the P1-specific

first exon. The association also persisted within HCC, making simple tumor-versus-non-tumor separation an insufficient explanation. This pattern does not fit the classical model in which promoter methylation reduces transcription. One possibility is that both P1 methylation and expression are influenced by another tumor feature, such as proliferation, tumor purity, or cell composition. Context-dependent relationships between IGF2 methylation and expression have been reported in breast cancer [10], but the mechanism in HCC remains unresolved. The combined results therefore support promoter-specific regulation: P4 is compatible with canonical oncofetal reactivation, whereas P1 reflects a distinct, context-dependent methylation-expression state.

The regulatory annotations provide supporting context rather than an independent central conclusion. The two P4-associated loci lie near the P4 promoter and showed active transcription-related ENCODE annotations, whereas rs3741208 lies approximately 22.3 kb upstream of P1 and describes the broader IGF2 regulatory landscape rather than the immediate P1 promoter. Viral infection and alcohol exposure can also influence methylation patterns [11, 12], but the available TCGA viral-serology field was insufficient for reliable stratified analysis. These factors should therefore be considered potential sources of biological heterogeneity rather than explanations established by the present data.

Several limitations constrain interpretation. The study uses a single cohort, and both RNA-seq and methylation were measured in bulk tissue, so tumor purity and cell composition may influence the observed associations. Although the P1 relationship persisted within HCC, its opposite direction in the smaller adjacent non-tumor subset suggests possible tissue-context dependence; only 41 adjacent non-tumor samples had matched methylation and expression data. P4-specific exon-level expression was unavailable. JASPAR motif matches and ENCODE annotations also do not demonstrate regulatory activity in TCGA tumors. Finally, the analysis did not include a multivariable model. Independent cohorts and models incorporating clinical, cellular-composition, and technical covariates are therefore needed to test the robustness and mechanism of the promoter-specific relationships reported here.

V. CONCLUSIONS

P1 and P4 exhibit distinct methylation-expression relationships in HCC despite similar overall sequence conservation. P4 shows the expected inverse association consistent with oncofetal reactivation, whereas P1 methylation is positively associated with both total IGF2 expression and P1-specific first-exon expression. These opposing patterns show that gene-level IGF2 measurements can obscure promoter-specific regulatory behavior and that a uniform promoter-methylation model is insufficient for this locus in HCC. Independent cohorts and multivariable analyses are needed to determine the mechanisms underlying the unexpected P1 relationship.

ACKNOWLEDGEMENTS

The author thanks Yash Verma, Gabriel Ortiz for guidance and feedback throughout the development of this research project.

REFERENCES

- [1] Rotwein, P. (2018). The complex genetics of human insulin-like growth factor 2 are not reflected in public databases. *Journal of Biological Chemistry*, 293(12), 4324–4333. <https://doi.org/10.1074/jbc.RA117.001573>
- [2] Chao, W., & D'Amore, P. A. (2008). IGF2: Epigenetic regulation and role in development and disease. *Cytokine & Growth Factor Reviews*, 19(2), 111–120. <https://doi.org/10.1016/j.cytogfr.2008.01.005>
- [3] Li, X., Nong, Z., Ekström, C., Larsson, E., Nordlinder, H., Hofmann, W. J., Trautwein, C., Odenthal, M., Dienes, H. P., Ekström, T. J., & Schirmacher, P. (1997). Disrupted IGF2 promoter control by silencing of promoter P1 in human hepatocellular carcinoma. *Cancer Research*, 57(10), 2048–2054.
- [4] Nardone, G., Romano, M., Calabrò, A., Pedone, P. V., de Sio, I., Persico, M., Budillon, G., Bruni, C. B., Riccio, A., & Zarrilli, R. (1996). Activation of fetal promoters of insulinlike growth factors II gene in hepatitis C virus-related chronic hepatitis, cirrhosis, and hepatocellular carcinoma. *Hepatology*, 23(6), 1304–1312. <https://doi.org/10.1002/hep.510230602>
- [5] Poirier, K., Chalas, C., Tissier, F., Couvert, P., Mallet, V., Carrié, A., Marchio, A., Sarli, D., Gicquel, C., Chaussade, S., Beljord, C., Chelly, J., Kerjean, A., & Terris, B. (2003). Loss of parental-specific methylation at the IGF2 locus in human hepatocellular carcinoma. *The Journal of Pathology*, 201(3), 473–479. <https://doi.org/10.1002/path.1477>
- [6] Yao, X., Hu, J.-F., Daniels, M., Shiran, H., Zhou, X., Yan, H., Lu, H., Zeng, Z., Wang, Q., Li, T., & Hoffman, A. R. (2003). A methylated oligonucleotide inhibits IGF2 expression and enhances survival in a model of hepatocellular carcinoma. *The Journal of Clinical Investigation*, 111(2), 265–273. <https://doi.org/10.1172/JCI15109>
- [7] van Dijk, M. A., Rodenburg, R. J. T., Holthuisen, P., & Sussenbach, J. S. (1992). The liver-specific promoter of the human insulin-like growth factor II gene is activated by CCAAT/enhancer binding protein (C/EBP). *Nucleic Acids Research*, 20(12), 3099–3104. <https://doi.org/10.1093/nar/20.12.3099>
- [8] Rodenburg, R. J. T., Holthuisen, P. E., & Sussenbach, J. S. (1997). A functional Sp1 binding site is essential for the activity of the adult liver-specific human insulin-like growth factor II promoter. *Molecular Endocrinology*, 11(2), 237–250. <https://doi.org/10.1210/mend.11.2.9888>
- [9] Li, T., Hu, J.-F., Qiu, X., Ling, J., Chen, H., Wang, S., Hou, A., Vu, T. H., & Hoffman, A. R. (2008). CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Molecular and Cellular Biology*, 28(20), 6473–6482. <https://doi.org/10.1128/MCB.00204-08>
- [10] Radhakrishnan, V. K., Ravichandran, K., Eke, C., Ortiz-Vicil, A., Tan, Q., De León, M., & De León, D. D. (2020). Methylation of a newly identified region of the INS-IGF2 gene determines IGF2 expression in breast cancer tumors and in breast cancer cells. *Oncotarget*, 11(44), 3904–3920. <https://doi.org/10.18632/oncotarget.27655>

- [11] Hlady, R. A., Tiedemann, R. L., Puszyk, W., Zendejas, I., Roberts, L. R., Choi, J.-H., Liu, C., & Robertson, K. D. (2014). Epigenetic signatures of alcohol abuse and hepatitis infection during human hepatocarcinogenesis. *Oncotarget*, 5(19), 9425–9443. <https://doi.org/10.18632/oncotarget.2444>
- [12] Nishihara, R., Wang, M., Qian, Z. R., Baba, Y., Yamauchi, M., Mima, K., Sukawa, Y., Kim, S. A., Inamura, K., Zhang, X., Wu, K., Giovannucci, E. L., Chan, A. T., Fuchs, C. S., Ogino, S., & Schernhammer, E. S. (2014). Alcohol, one-carbon nutrient intake, and risk of colorectal cancer according to tumor methylation level of IGF2 differentially methylated region. *The American Journal of Clinical Nutrition*, 100(6), 1479–1488. <https://doi.org/10.3945/ajcn.114.095539>