Regulatory Landscape of ANGPT1: Non-Coding SNPs and Epigenetic Features

Parampreet Kaur[1], Rajinder Kaur[2], and Dimple Chopra[3]
[1]*Research Scholar, Punjabi University, Patiala*
[2]*Professor, Punjabi University, Patiala*
[3]*Professor, Government Medical College, Patiala*

Abstract- Genetic variations within regulatory regions can profoundly influence gene expression, with single nucleotide polymorphisms (SNPs) and insertion–deletions (INDELs) in non-coding sequences serving as critical modulators of transcriptional and post-transcriptional regulation. Angiopoietin-1 (ANGPT1), a vascular growth factor essential for angiogenesis and vascular homeostasis, represents a prime candidate for such regulatory interrogation. In this study, an in silico approach was employed to analyze promoter sites, CpG islands, polyadenylation signals, and non-coding SNPs in the ANGPT1 gene, with a particular focus on their potential regulatory functions. Promoter analysis revealed multiple highly probable transcription start sites, accompanied by both TATA-driven and TATA-less promoter architectures. CpG island predictions identified three GC-rich regions likely to contribute to epigenetic regulation through DNA methylation. Polyadenylation site prediction uncovered several high-confidence cleavage signals, suggesting alternative polyadenylation as a regulatory mechanism. Functional annotation of non-coding SNPs revealed a small subset with high regulatory potential, supported by transcription factor binding and DNase evidence, along with SNPs and INDELs affecting microRNA (miRNA) target sites in the 3′ UTR. Collectively, these findings highlight a complex regulatory landscape of ANGPT1, where non-coding variation and epigenetic features may fine-tune gene expression, potentially influencing vascular and inflammatory disease susceptibility.

1. INTRODUCTION

The comprehensive study of genetic variation has become indispensable for unraveling the complexities of human health and disease. Among these variations, single nucleotide polymorphisms (SNPs) are the most abundant form, occurring as single base substitutions within the genome. The human genome is estimated to harbor 3–5 million SNPs per individual, making them the most common source of genetic variability [1]. Although many SNPs are functionally neutral, others exert significant influence on gene function. Variants located in coding regions may alter amino acid sequences, thereby affecting protein structure and stability [2]. Equally critical are non-coding variants that shape gene regulation by modulating transcription factor binding, CpG methylation, chromatin accessibility, RNA splicing, or post-transcriptional regulation via microRNAs (miRNAs) [3,4].

Angiopoietin-1 (ANGPT1), a ligand for the endothelial receptor tyrosine kinase TIE-2, plays a central role in vascular development and homeostasis [5]. It promotes endothelial cell migration, vascular sprouting, and stabilization while simultaneously functioning as an anti-inflammatory factor through antagonism of VEGF and NF-κB signaling [6,7]. Dysregulation of ANGPT1 has been implicated in cardiovascular disorders, cancer, chronic inflammatory diseases, and psoriasis [8–10]. Given its centrality to vascular biology, regulatory variation within ANGPT1 may contribute significantly to disease susceptibility.

In this study, we employed a bioinformatics-driven approach to investigate regulatory sequences and non-coding SNPs in ANGPT1. Promoter elements, CpG islands, polyadenylation sites, and non-coding SNPs (including miRNA target site variants) were systematically analyzed. By focusing on regulatory features and non-coding variation, this work underscores the contribution of non-coding polymorphisms to ANGPT1 regulation and provides insights into their potential functional relevance in vascular and inflammatory pathologies.

2. METHODOLOGY

2.1 Data Collection

SNP data for the ANGPT1 gene were retrieved from dbSNP. Gene FASTA sequences were obtained from the NCBI Gene Database, while the protein FASTA sequence (NCBI Accession: AAI52420) was collected from the NCBI Protein Database. Additionally, structural data for ANGPT1 were

screened from the Protein Data Bank (PDB), although the scope of this study was restricted to regulatory sequence analysis and non-coding SNP annotation.

2.2 Regulatory Sequence Analysis

To investigate the regulatory elements of the ANGPT1 gene, multiple computational approaches were employed. Promoter prediction was carried out using Promoter 2.0 (DTU HealthTech) to identify potential RNA polymerase II transcription start sites (TSS) within the gene sequence [11]. In addition, the TSSG tool (Softberry) was used to predict RNA polymerase III promoters, along with associated TATA boxes and putative transcription factor binding sites (TFBS) [12]. To assess CpG-rich regulatory regions, CpGFinder (Softberry) was applied, enabling the identification of CpG islands that may play a role in epigenetic regulation [13]. Finally, polyadenylation signal prediction was conducted using the PolyAH program (Softberry), which detects cleavage and polyadenylation sites in mRNA precursors, thereby providing insights into transcript processing and stability [14].

2.3 Non-Coding SNP Analysis

To evaluate the functional consequences of non-coding variants within the ANGPT1 gene, two complementary approaches were applied. RegulomeDB was used to annotate and classify non-coding SNPs based on multiple lines of regulatory evidence, including transcription factor binding, DNase I hypersensitivity, and motif conservation, thereby providing insights into their potential role in transcriptional regulation [15]. In parallel, variants located within microRNA (miRNA) target sites were examined using the PolymiRTS Database 3.0, which identifies SNPs and INDELs that may alter post-transcriptional regulation by predicting the gain or loss of miRNA binding interactions [16].

3. RESULTS

3.1 Promoter Site Prediction in ANGPT1 (Promoter 2.0 Analysis)

Promoter site analysis using the **Promoter 2.0** tool identified several putative transcription start sites (TSS) across the ANGPT1 gene sequence. Prediction scores varied, but sites with scores **≥ 1.0** were considered **highly likely promoter regions**, reflecting a strong probability of transcription initiation.

A total of **54 highly likely promoter regions** were identified, distributed throughout the ANGPT1 sequence (Table 1). The strongest signals were observed at **positions 116,200 (score: 1.341), 121,000 (score: 1.304), 38500 (score: 1.214), 176,700 (score: 1.214), and 150,800 (score: 1.222)**. The presence of multiple promoter sites suggests that **ANGPT1 may employ alternative promoters**, enabling tissue-specific or context-dependent expression.

The clustering of promoter regions in both upstream and internal locations indicates a **complex transcriptional control architecture**, consistent with the gene's diverse roles in angiogenesis and inflammatory modulation. Experimental validation (e.g., promoter-reporter assays or ChIP) would be required to confirm their functionality.

Table 1. Predicted Highly Likely Promoter Sites in the ANGPT1 Gene (Promoter 2.0 Analysis)

Position	Score	Likelihood
2300	1.1	Highly likely prediction
9700	1.132	Highly likely prediction
10700	1.08	Highly likely prediction
16400	1.117	Highly likely prediction
22300	1.098	Highly likely prediction
27900	1.099	Highly likely prediction
29000	1.203	Highly likely prediction
31100	1.166	Highly likely prediction
35600	1.079	Highly likely prediction
38500	1.214	Highly likely prediction
48800	1.273	Highly likely prediction
53000	1.078	Highly likely prediction
55900	1.041	Highly likely prediction
59900	1.13	Highly likely prediction
63300	1.196	Highly likely prediction
81100	1.09	Highly likely prediction
83100	1.22	Highly likely prediction
83900	1.067	Highly likely prediction
85600	1.151	Highly likely prediction
90300	1.089	Highly likely prediction
98000	1.18	Highly likely prediction
100100	1.18	Highly likely prediction
101200	1.043	Highly likely prediction
101900	1.077	Highly likely prediction
103400	1.185	Highly likely prediction
106100	1.049	Highly likely prediction
108400	1.168	Highly likely prediction
116200	1.341	Highly likely prediction
117700	1.246	Highly likely prediction
121000	1.304	Highly likely prediction
126300	1.21	Highly likely prediction
129200	1.088	Highly likely prediction
134900	1.071	Highly likely prediction
141500	1.168	Highly likely prediction
145800	1.197	Highly likely prediction
150800	1.222	Highly likely prediction
153400	1.1	Highly likely prediction
155800	1.115	Highly likely prediction

156600	1.093	Highly likely prediction
160400	1.163	Highly likely prediction
162800	1.182	Highly likely prediction
165100	1.036	Highly likely prediction
173300	1.166	Highly likely prediction
176700	1.214	Highly likely prediction
179700	1.14	Highly likely prediction
185000	1.141	Highly likely prediction
201300	1.189	Highly likely prediction
204400	1.071	Highly likely prediction
211800	1.231	Highly likely prediction
214900	1.144	Highly likely prediction
217300	1.135	Highly likely prediction
221100	1.06	Highly likely prediction
223800	1.221	Highly likely prediction
228500	1.077	Highly likely prediction
235800	1.143	Highly likely prediction
243000	1.085	Highly likely prediction

3.2 CpG Island Predictions

CpGFinder detected three CpG islands within ANGPT1:

- 161,275–161,535 bp (261 bp, GC 62.1%, Obs/Exp CpG: 0.845),
- 185,860–186,127 bp (268 bp, GC 59.0%, Obs/Exp CpG: 0.659),
- 216,145–216,370 bp (226 bp, GC 55.3%, Obs/Exp CpG: 0.825). These regions show strong enrichment for CpG dinucleotides, suggesting potential regulation via DNA methylation. Aberrant methylation here may alter ANGPT1 expression, impacting angiogenesis and inflammatory responses.

3.3 Polyadenylation Site Predictions

The POLYAH tool predicted 333 polyadenylation sites, of which 86 had LDF ≥5.0. The most prominent cleavage signal was observed at 94786 bp (LDF: 8.77), followed by 247220 bp (LDF: 7.84) and 85317 bp (LDF: 7.65). These results indicate the presence of alternative polyadenylation (APA) sites, allowing generation of multiple transcript isoforms with distinct 3′ UTRs, potentially modulating mRNA stability, localization, and translational efficiency.

3.4 Non-Coding SNP Functional Annotation in ANGPT1

RegulomeDB analysis identified a total of **21 non-coding SNPs** in the **ANGPT1** gene with regulatory relevance, distributed across ranks **2a, 2b, and 3a**. Among these, **one variant (4.8%)** was classified as **Rank 2a**, representing the strongest regulatory evidence with transcription factor binding, motif conservation, DNase hypersensitivity, and footprinting support. **Six SNPs (28.6%)** were ranked **2b**, supported by transcription factor binding and chromatin accessibility data, while the majority, **14 SNPs (66.6%)**, fell under **Rank 3a**, indicating moderate functional potential with evidence of TF binding and motif matching.

This distribution highlights that while most ANGPT1 non-coding SNPs show moderate levels of regulatory evidence (Rank 3a), the single Rank

Table 2. Non-Coding SNPs in ANGPT1 with High to Moderate Regulatory Potential (RegulomeDB Ranks 2a–3a) 2a SNP may represent a particularly important functional variant, potentially exerting significant influence on transcriptional regulation of ANGPT1.

S. No.	dbSNP ID	Rank	Score
1	rs1405473465	2a	0.91
2	rs1276102060	2b	0.62
3	rs1210615750	2b	0.85
4	rs1263735498	2b	0.64
5	rs1293158164	2b	0.74
6	rs1291547870	2b	0.64
7	rs1340477853	2b	1.00
8	rs1246333917	3a	0.49
9	rs1304673984	3a	1.00
10	rs1282789617	3a	0.68
11	rs1195240882	3a	0.99
12	rs1343604931	3a	0.55
13	rs1324367683	3a	0.68
14	rs1300016403	3a	0.50
15	rs1295251462	3a	0.85
16	rs1308639547	3a	0.49
17	rs1429897689	3a	0.86
18	rs1469711828	3a	0.78
19	rs1363470263	3a	0.57
20	rs1402073355	3a	0.81
21	rs1249420599	3a	0.99

3.5 Non-Coding SNPs in miRNA Target Sites

PolymiRTS analysis identified 13 representative SNPs/INDELs in the ANGPT1 3′ UTR affecting miRNA binding:

- Variants such as rs191233649 (T→N) and rs2507799 (G→C) disrupted multiple conserved miRNA binding sites, leading to potential loss of post-transcriptional repression.
- INDEL rs45614542 (CATTT insertion) created a novel binding site for hsa-miR-494-3p, possibly reducing ANGPT1 expression.
- SNP rs45517437 (C→T) disrupted binding sites for four different miRNAs, making it a high-impact regulatory variant.
- INDELs like rs112136102 (CATTT deletion) showed dual effects, causing both gain and loss of binding sites depending on the miRNA considered.

Overall, most variants were predicted to disrupt existing miRNA interactions, potentially elevating ANGPT1 expression and enhancing angiogenic and anti-inflammatory functions, while a minority created novel sites that could suppress ANGPT1.

Table 3. Predicted ANGPT1 miRNA target site polymorphisms (PolymiRTS Database 3.0)

	Variant Type	Reference / Derived Allele	Affected miRNA(s)	Predicted Effect on miRNA Binding	Context+ Score Change	Functional Class
rs191233649	SNP	T → N	hsa-miR-3185, hsa-miR-4511, hsa-miR-3156-5p, hsa-miR-4699-5p, hsa-miR-7161-3p	Loss of binding sites	-0.087 to -0.227	Target site alteration
rs2507799	SNP	G → C	hsa-miR-10a-3p, hsa-miR-3673, hsa-miR-6505-5p	Loss of binding	-0.109 to -0.319	Target site alteration
rs112136102	INDEL	Deletion (CATTT)	hsa-miR-1252-3p, hsa-miR-3646	Gain/Loss of binding	-0.045 to -0.157	Target site disruption
rs45614542	INDEL	Insertion (CATTT)	hsa-miR-494-3p	Gain of binding site	+0.011	Novel site creation
rs34247877	INDEL	Deletion (CATTT)	hsa-miR-3646, hsa-miR-3662	Loss of binding	-0.099 to -0.153	Target site alteration
rs41516446	INDEL	Deletion	hsa-miR-186-5p, hsa-miR-6507-5p	Loss of binding	-0.015 to -0.087	Target site alteration
rs35519577	INDEL	Deletion	hsa-miR-186-5p, hsa-miR-6507-5p	Loss of binding	-0.015 to -0.017	Target site alteration
rs45517437	SNP	C → T	hsa-miR-4643, hsa-miR-466, hsa-miR-4789-3p, hsa-miR-6507-5p	Multiple site disruption	-0.089 to -0.415	Target site alteration
rs182655901	SNP	T → C	hsa-miR-493-5p, hsa-miR-5580-3p	Loss of binding	-0.038 to -0.189	Target site alteration
rs185174572	SNP	T → C	hsa-miR-545-5p, hsa-miR-144-5p, hsa-miR-3935, hsa-miR-936	Loss of binding	-0.171 to -0.233	Target site alteration
rs191161582	SNP	C → A	hsa-miR-299-5p, hsa-miR-29b-1-5p, hsa-miR-582-3p, hsa-miR-452-5p, hsa-miR-4676-3p	Loss of binding	-0.154 to -0.233	Target site alteration
rs148992517	SNP	T → C	hsa-miR-3653	Loss of binding	-0.136	Target site alteration
rs143696955	SNP	C → T	hsa-miR-20b-3p, hsa-miR-3137, hsa-miR-8065	Loss of binding	-0.106 to -0.265	Target site alteration

4. DISCUSSION

This study provides a comprehensive in silico evaluation of ANGPT1 regulatory architecture, integrating promoter, CpG, polyadenylation, and non-coding SNP analyses. The presence of multiple promoters, CpG islands, and strong polyA signals reflects a highly dynamic transcriptional and post-transcriptional regulatory system. Importantly, functional SNPs identified in regulatory and miRNA target regions may fine-tune ANGPT1 expression, influencing angiogenesis and inflammation.

Given ANGPT1's established roles in cardiovascular and inflammatory diseases, these findings have potential implications for disease susceptibility and therapeutic intervention. Future work should validate these predictions experimentally, using promoter-reporter assays, ChIP-seq for TF binding, and functional miRNA assays to confirm allele-specific regulation.

5. CONCLUSION

The regulatory sequence and non-coding SNP analysis of ANGPT1 highlights a complex network of transcriptional, epigenetic, and post-transcriptional controls. A small subset of high-confidence SNPs and INDELs may play disproportionate roles in shaping ANGPT1 expression, providing candidate variants for further investigation in vascular and inflammatory diseases.

REFERENCE

[1] Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of

human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001;409(6822):928–933. doi:10.1038/35057149

[2] Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet*. 2006;7:61–80. doi:10.1146/annurev.genom.7.080505.115630

[3] Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol*. 2012;30(11):1095–1106. doi:10.1038/nbt.2422

[4] Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res*. 2014;42(Database issue):D86–91. doi:10.1093/nar/gkt1028

[5] Davis S, Aldrich TH, Jones PF, Acheson A, Compton DL, Jain V, et al. Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning. *Cell*. 1996;87(7):1161–1169. doi:10.1016/S0092-8674(00)81812-7

[6] Fiedler U, Krissl T, Koidl S, Weiss C, Koblizek T, Deutsch U, et al. Angiopoietin-1 and angiopoietin-2 share the same binding domains in the Tie-2 receptor involving the first Ig-like loop and the epidermal growth factor-like repeats. *J Biol Chem*. 2003;278(3):1721–1727. doi:10.1074/jbc.M207130200

[7] Brindle NP, Saharinen P, Alitalo K. Signaling and functions of angiopoietin-1 in vascular protection. *Circ Res*. 2006;98(8):1014–1023. doi:10.1161/01.RES.0000218275.54089.12

[8] Huang H, Bhat A, Woodnutt G, Lappe R. Targeting the ANGPT-TIE2 pathway in malignancy. *Nat Rev Cancer*. 2010;10(8):575–585. doi:10.1038/nrc2894

[9] Zhang L, Yang N, Park JW, Katsaros D, Fracchioli S, Cao G, et al. Tumor-derived vascular endothelial growth factor up-regulates angiopoietin-2 in host endothelium and destabilizes host vasculature in ovarian cancer. *Cancer Res*. 2003;63(12):3409–3415.

[10] Bielenberg DR, Zetter BR. The contribution of angiogenesis to the process of metastasis. *Cancer J*. 2015;21(4):267–273. doi:10.1097/PPO.0000000000000138

[11] Knudsen S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics*. 1999;15(5):356–361. doi:10.1093/bioinformatics/15.5.356 (Tool: Promoter 2.0, DTU HealthTech, https://services.healthtech.dtu.dk/service.php?Promoter-2.0)

[12] Solovyev VV, Shahmuradov IA. PromH: Promoters identification using orthologous genomic sequences. *Nucleic Acids Res*. 2003;31(13):3540–3545. doi:10.1093/nar/gkg535 (Tool: TSSG, Softberry, http://www.softberry.com/berry.phtml?topic=tssg&group=programs&subgroup=promoter)

[13] Ioshikhes IP, Zhang MQ, Majors J. Prediction of cis-acting elements by clustering analysis. *Nucleic Acids Res*. 2000;28(2):311–315. doi:10.1093/nar/28.2.311 (Tool: CpGFinder, Softberry, http://www.softberry.com/berry.phtml?topic=cpgfinder&group=programs&subgroup=promoter)

[14] Salamov AA, Solovyev VV. Recognition of 3′-processing sites of human mRNA precursors. *J Mol Biol*. 1997;268(1):24–31. doi:10.1006/jmbi.1997.0924 (Tool: PolyAH, Softberry, http://www.softberry.com/berry.phtml?topic=polyah&group=programs&subgroup=promoter)

[15] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790–1797. doi:10.1101/gr.137323.112 (Tool: RegulomeDB, https://regulomedb.org/)

[16] Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res*. 2014;42(Database issue): D86–91. doi:10.1093/nar/gkt1028 (Tool: PolymiRTS Database 3.0, https://compbio.uthsc.edu/miRSNP/)