CellGDnG: predicting cell-cell communication by combining heterogeneous ensemble deep learning and weighted geometric mean

Lihong Peng, Longlong Liu, Liangliang Huang, Zongzheng Bai, Min Chen∗, Xing Chen∗

Figures **S1-S2** in Supplementary Materials discussed the settings of the feature dimensions and project feature dimensions.

Figures **S3-S5** in Supplementary Materials provide the UpSetR maps within the other three tissues.

Figure **S6** in Supplementary Materials discussed the setting of α.

Figure **S7** in Supplementary Materials discussed the setting of the encoder layers *L*.

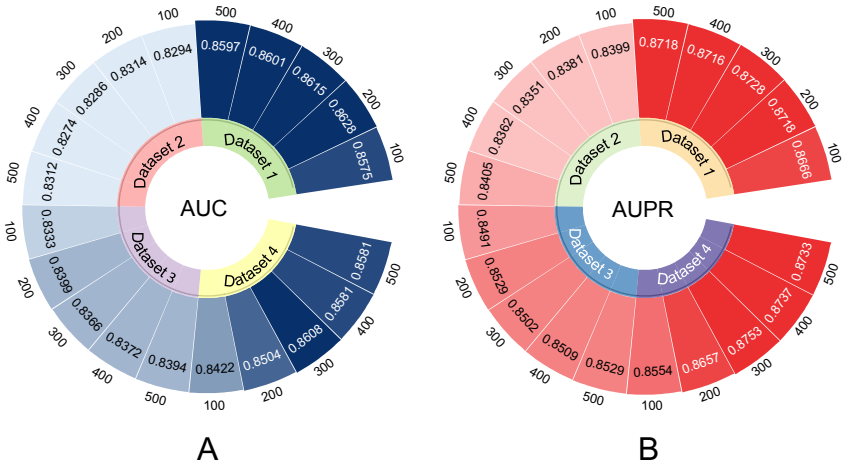
Figure **S8** in Supplementary Materials show the CCC prediction results based on different scoring methods (i.e., specific expression, expression product, and total expression).

Table **S3** in Supplementary Materials show the CCC prediction results based on different scoring methods (i.e., specific expression, expression product, and total expression).

Tables **S4-S6** in Supplementary Materials provided CCC prediction results of CellGDnG and the five baselines within CRC, HNSCC, and melanoma tissues.

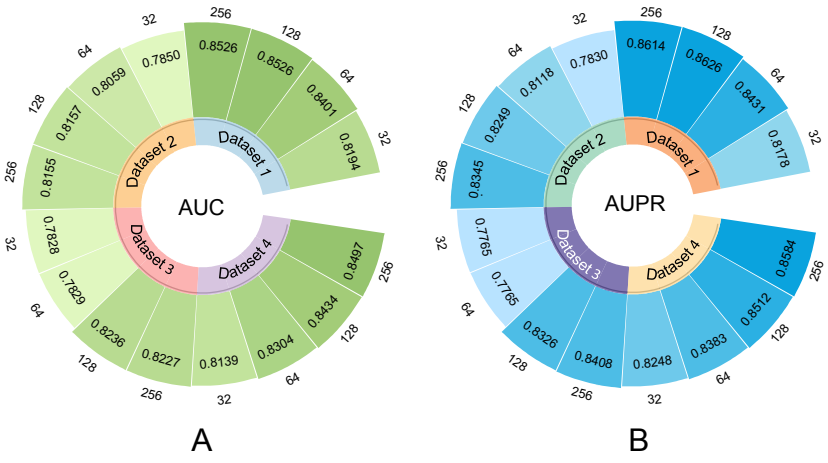
1.Effect of feature dimensions and project feature dimensions

Considering that the AUC and AUPR values can reflect the prediction performance of models more comprehensively, we compared the effect of several feature dimensions on the performance of CellGDnG upon 5-fold cross-validation on the four datasets. The results were shown in Figures S1. For dataset 1 and dataset 3, we can see that both AUC and AUPR values improve as the feature dimensionality increases, and the best results are achieved at 200-*d*. For dataset 2, the value of AUC works best in 200-*d* and 500-*d*. For dataset 4, we can see that with the increase of feature dimensions, the AUC and AUPR values have been improved significantly improved. Considering the collective impact of the dimensions on the model, a dimension that is excessively small may result in the loss of pertinent features, whereas an excessively high dimension introduces redundant information and requires extensive computational resources. In order to effectively assess the intensity of CCC using dataset 1, we ultimately established the model dimension as 300-*d*.



Figures S1: Performance comparison of feature dimension. A. AUC B. AUPR.

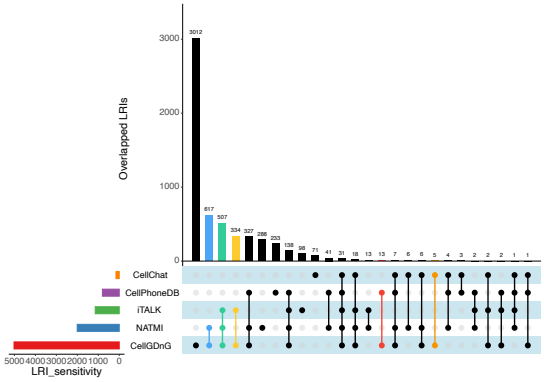
We compared the effect of several projection dimensions E on the performance of HGAE under 1-layers encoder upon 5-fold cross-validation. The results were shown in Figures S2. We can see that with the increase of projection dimensions, the AUC and AUPR values have been improved. We chose 128-dimensions as our default projection dimensions.



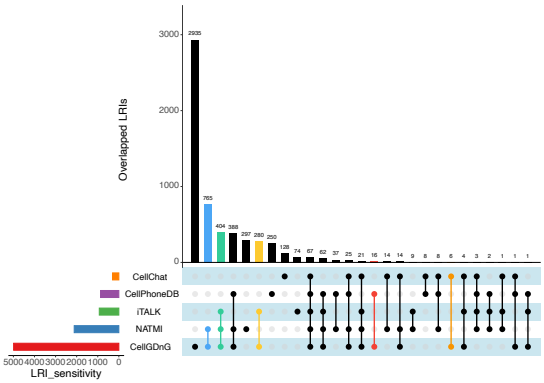
Figures S2: Performance comparison of project feature dimensions. A. AUC B. AUPR.

2. The CCC inference result validation

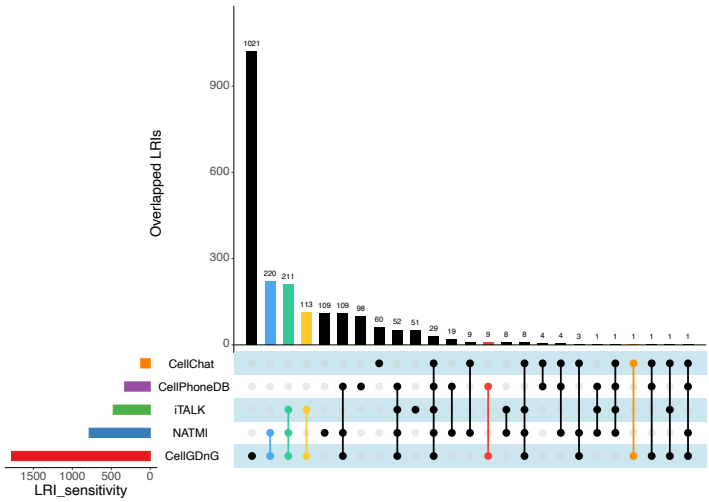
Figures S3-S5 in Supplementary Materials provide the UpSetR maps within CRC, HNSCC, and Melanoma tissues, respectively. On the left of Figures S3-S5, each bar chart depicts the LRI_*sensitivity* value counted by the five tools. On the right of Figures S3-S5, each straight line composed of multiple dots reflects the overlapped LRI number between corresponding several CCC inference tools, each histogram at the top indicates the overlapped LRI number shared by the corresponding tools. The results demonstrated that CellGDnG computed the highest LRI_*sensitivity* and displayed the most overlapped LRIs with the other four tools.



Figures S3: UpSetR map within CRC tissue.



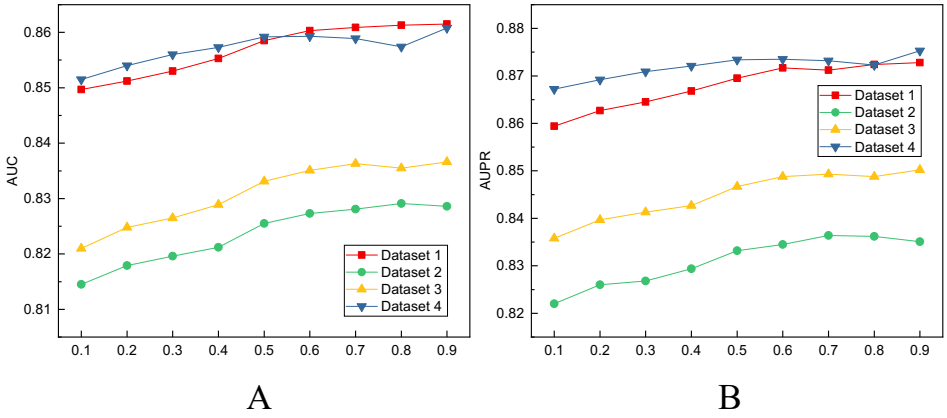
Figures S4: UpSetR map within HNSCC tissue.



Figures S5: UpSetR map within Melanoma tissue.

3. Effect of parameter α and encoder layers

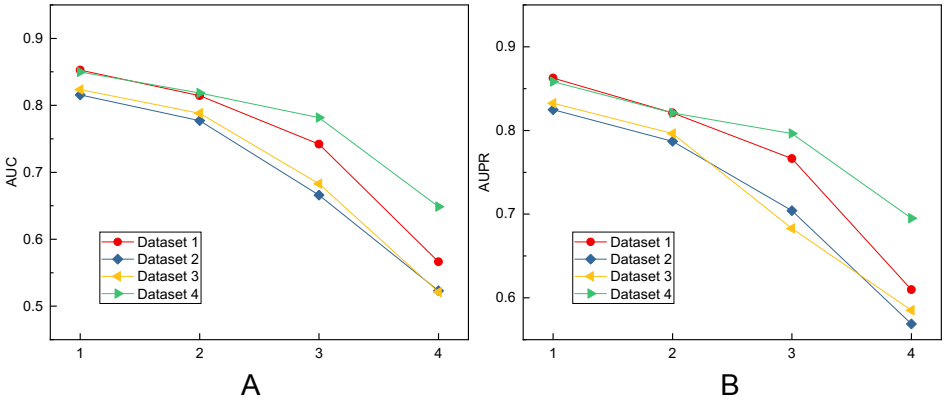
Figure S6 in Supplementary Materials discussed the setting of parameter α. The results elucidated that CellGDnG computed the best classification performance in most cases when α is set to 0.9. Consequently, we set α to 0.9.



Figures S6: Performance comparison of feature dimension. A. AUC B. AUPR.

4. Effect of encoder layers

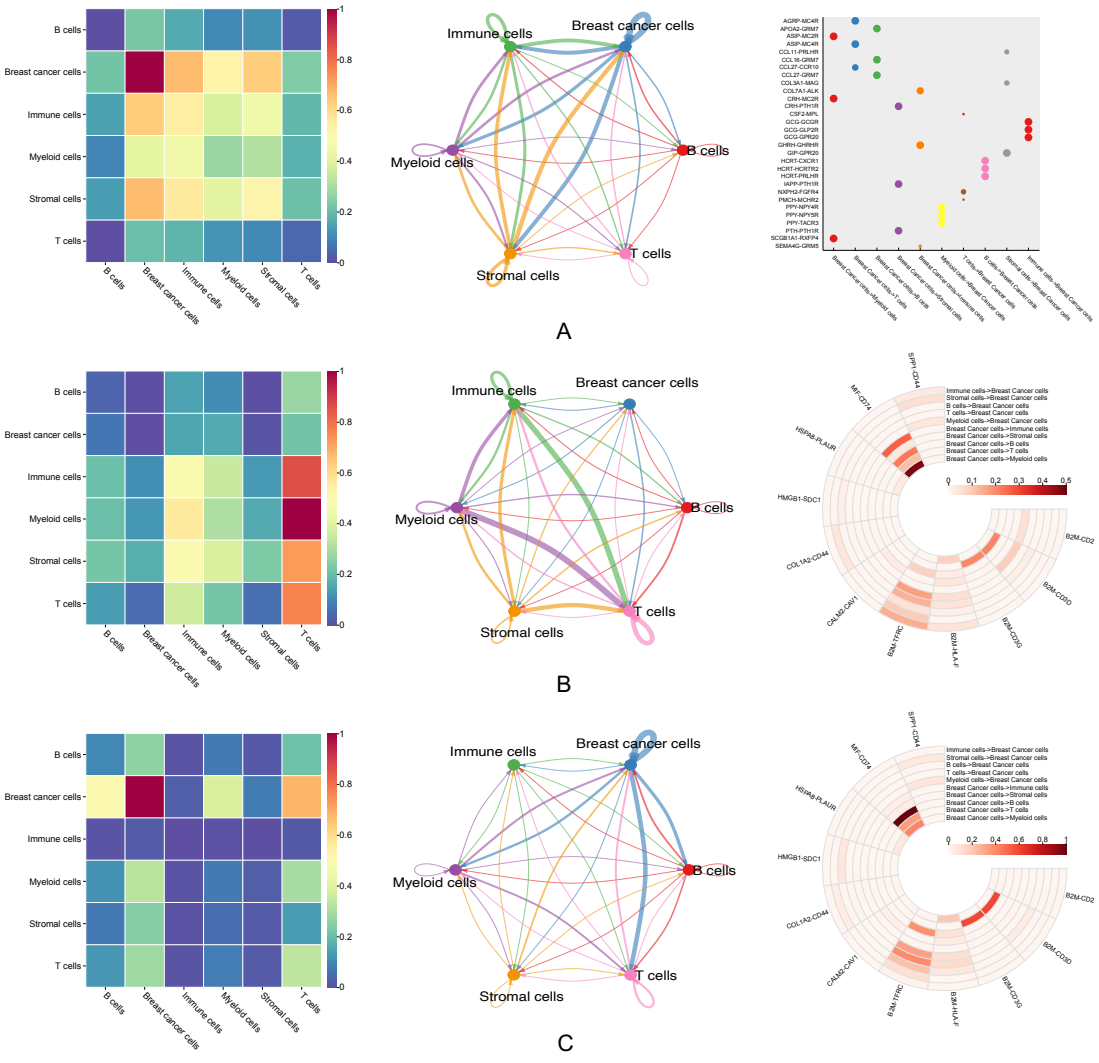
Figure S7 in Supplementary Materials discussed the setting of the encoder layers L. When L were set to 1, HGAE achieved the optimal predictive performance. However, when the encoder layer number increased, the CellGDnG performance sharply de- clined. Particularly, when the encoder layer number exceeded 4, CellGDnG were difficult to train, resulting in a sharp drop in performance. Thus, we did not consider scenarios where the encoder layer number exceeded 4.



Figures S7: Performance comparison of feature dimension. A. AUC B. AUPR.

5. CCC prediction results based on different scoring methods

Figure S8 and Table S3 in Supplementary Materials show the CCC prediction results based on different scoring methods (i.e., specific expression, expression product, and total expression). The left, middle, and right subfigures in A delineate heatmap view of CCC network, circle plot of CCC network, and bubble plot representing the top 3 LRI mediating corresponding CCC based on specific expression. The left, middle, and right subfigures in B and C delineate heatmap view of CCC network, circle plot of CCC network, and circle heatmap plot representing the top 3 LRI mediating corresponding CCC based on expression product, and total expression, respectively.



Figures S8: CCC prediction results based on different scoring methods.

Table S3: The CCC prediction results based on different scoring methods.

Function	Cell type	"calling"LRIs	"callback"LRIs	Evidence
Specific expression	Immune cells	GHRH-GHRHR, COL7A1-ALK, SEMA4G-GRM5	GCG-GPR20, GCG-GCGR, GCG-GLP2R	Ref[1-3]
	Stromal cells	IAPP-PTH1R, CRH-PTH1R, PTH-PTH1R	GIP-GPR20, COL3A1-MAG, CCL11-PRLHR	Ref [4-6]
	Myeloid cells	CRH-MC2R, ASIP-MC2R, SCGB1A1-RXFP4	PPY-NPY5R, PPY-NPY4R, PPY-TACR3	Ref [7-10]
	B cells	CCL27-GRM7, CCL16-GRM7, APOA2-GRM7	HCRT-HCRTR2, HCRT-CXCR1, HCRT-PRLHR	
	T cells	AGRP-MC4R, ASIP-MC4R, CCL27-CCR10	NXPH2-FGFR4, CSF2-MPL, PMCH-MCHR2	
Product expression	Myeloid cells	MIF-CD74, B2M-HLA-F, HSPA8-PLAUR	B2M-TFRC, B2M-HLA-F, SPP1-CD44	Ref [7-10]
	Immune cells	MIF-CD74, B2M-CD3D, B2M-CD2	B2M-TFRC, B2M-HLA-F, SPP1-CD44	Ref [1-3]
	T cells	B2M-CD3D, B2M-CD3G, MIF-CD74	B2M-TFRC, B2M-HLA-F, B2M-CD3D	Ref [11-13]
	Stromal cells	B2M-HLA-F, MIF-CD74, CALM2-CAV1	B2M-TFRC, SPP1-CD44, COL1A2-CD44	
	B cells	MIF-CD74, B2M-TFRC, B2M-HLA-F	B2M-TFRC, B2M-HLA-F, HMGB1-SDC1	
Total expression	T cells	B2M-CD3D, B2M-CD3G, MIF-CD74	B2M-TFRC, B2M-HLA-F, B2M-CD3D	Ref [11-13]
	B cells	MIF-CD74, B2M-TFRC, B2M-HLA-F	B2M-TFRC, B2M-HLA-F, HMGB1-SDC1	Ref [14-16]
	Myeloid cells	MIF-CD74, B2M-HLA-F, HSPA8-PLAUR	B2M-TFRC, B2M-HLA-F, SPP1-CD44	Ref [7-10]
	Stromal cells	B2M-HLA-F, MIF-CD74, CALM2-CAV1	B2M-TFRC, SPP1-CD44, COL1A2-CD44	
	Immune cells	MIF-CD74, B2M-CD3D, B2M-CD2	B2M-TFRC, B2M-HLA-F, SPP1-CD44	
Weighted geometric-mean joint	Myeloid cells	MIF-CD74, ASIP-MC2R, CRH-MC2R	PPY-CCR9, PPY-NPY5R, PPY-TACR3	Ref [7-10]
	Stromal cells	GHRH-PTH1R, FSHB-PTH1R, IAPP-PTH1R	GIP-GPR20, COL3A1-MAG, CCL11-PRLHR	Ref [4-6]
	Immune cells	GHRH-GHRHR, COL7A1-ALK, SEMA4G-GRM5	GCG-GPR20, GCG-GCGR, GCG-GLP2R	Ref [1-3]
	T cells	ASIP-MC4R, AGRP-MC4R, CCL27-CCR10	NXPH2-FGFR4, CSF2-MPL, PMCH-MCHR2	
	B cells	MIF-CD74, CCL27-GRM7, APOA2-GRM7	HCRT-HCRTR2, HCRT-CXCR1, HCRT-PRLHR	

6. CCC prediction results within CRC, HNSCC, and melanoma tissues.

Table S4: Comparison of CellGDnG with six tools in CRC tissue.

Ranking	CellChat	iTALK	CellPhoneDB	NATMI	CellComNet	CellEnBoost$_p$	CellEnBoost$_s$	CellEnBoost$_c$	CellGDnG
1	Macrophages	Fibroblasts	Fibroblasts	ECs	ECs	ECs	Epithelial cells	Fibroblasts	Epithelial cells
2	ECs	T cells	Epithelial cells	Macrophages	Macrophages	Macrophages	Fibroblasts	ECs	Fibroblasts
3	B cells	Epithelial cells	B cells	Fibroblasts	T cells	T cells	B cells	Epithelial cells	T cells
4	Mast cells	Macrophages	ECs	Mast cells	Fibroblasts	Fibroblasts	T cells	T cells	B cells
5	Fibroblasts	B cells	T cells	T cells	B cells	B cells	Macrophages	Mast cells	Macrophages
6	T cells	ECs	Macrophages	B cells	Epithelial cells	Epithelial cells	ECs	B cells	ECs
7	Epithelial cells	Mast cells	Mast cells	Epithelial cells	Mast cells	Mast cells	Mast cells	Macrophages	Mast cells

ECs: Endothelial cells

Table S5: Comparison of CellGDnG with six tools in HNSCC tissue.

Ranking	CellChat	iTALK	CellPhoneDB	NATMI	CellComNet	CellEnBoost$_p$	CellEnBoost$_s$	CellEnBoost$_c$	CellGDnG
1	Fibroblasts	Fibroblasts	Macrophages	Macrophages	ECs	Macrophages	Fibroblasts	Macrophages	Fibroblasts
2	ECs	ECs	ECs	ECs	Macrophages	ECs	ECs	ECs	ECs
3	Macrophages	Macrophages	Fibroblasts	Myocytes	Fibroblasts	Fibroblasts	Macrophages	Fibroblasts	Macrophages
4	T cells	Myocytes	T cells	Fibroblasts	Dendritic cells	Dendritic cells	T cells	Dendritic cells	T cells
5	Mast cells	Dendritic cells	Dendritic cells	Dendritic cells	Mast cells	Mast cells	Myocytes	Mast cells	Dendritic cells
6	Dendritic cells	Mast cells	Mast cells	Mast cells	T cells	T cells	Dendritic cells	T cells	Mast cells
7	B cells	B cells	B cells	B cells	Myocytes	Myocytes	Mast cells	Myocytes	Myocytes
8	Myocytes	T cells	Myocytes	T cells	B cells	B cells	B cells	B cells	B cells

ECs: Endothelial cells

Table S6: Comparison of CellGDnG with six tools in Melanoma tissue.

Ranking	CellChat	iTALK	CellPhoneDB	NATMI	CellComNet	CellEnBoost$_p$	CellEnBoost$_s$	CellEnBoost$_c$	CellGDnG
1	CAFs	CAFs	Macrophages	CAFs	CAFs	CAFs	CAFs	CAFs	T cells
2	ECs	Macrophages	ECs	ECs	Macrophages	Macrophages	Macrophages	Macrophages	Macrophages
3	Macrophages	ECs	CAFs	T cells	ECs	ECs	ECs	ECs	CAFs
4	T cells	NK cells	T cells	Macrophages	NK cells	NK cells	T cells	T cells	B cells
5	NK cells	T cells	B cells	B cells	T cells	T cells	B cells	NK cells	ECs
6	B cells	B cells	NK cells	NK cells	B cells	B cells	NK cells	B cells	NK cells

ECs: Endothelial cells

References

[1] W. Chung, H. H. Eum, H.-O. Lee, K.-M. Lee, H.-B. Lee, K.-T. Kim, H. S. Ryu, S. Kim, J. E. Lee, Y. H. Park et al., "Single-cell rna-seq enables comprehensive tumour and immune cell profiling in primary breast cancer," Nature communications, vol. 8, no. 1, p. 15081, 2017.

[2] D. G. DeNardo and L. M. Coussens, "Inflammation and breast cancer. balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression," Breast cancer research, vol. 9, pp. 1–10, 2007.

[3] S. Sousa, R. Brion, M. Lintunen, P. Kronqvist, J. Sandholm, J. Mönkkönen, P.-L. Kellokumpu-Lehtinen, S. Lauttia, O. Tynninen, H. Joensuu et al., "Human breast cancer cells educate macrophages toward the m2 activation status," Breast cancer research, vol. 17, no. 1, pp. 1–14, 2015.

[4] B. S. Hill, A. Sarnella, G. DAvino, and A. Zannetti, "Recruitment of stromal cells into tumour microenvironment promote the metastatic spread of breast cancer," in Seminars in cancer biology, vol. 60. Elsevier, 2020, pp. 202–213.

[5] D. R. Welch and D. R. Hurst, "Defining the hallmarks of metastasis," Cancer research, vol. 79, no. 12, pp. 3011–3027, 2019.

[6] Z. Tu and A. E. Karnoub, "Mesenchymal stem/stromal cells in breast cancer development and management," in Seminars in Cancer Biology. Elsevier, 2022.

[7] L. Racioppi, E. R. Nelson, W. Huang, D. Mukherjee, S. A. Lawrence, W. Lento, A. M. Masci, Y. Jiao, S. Park, B. York et al., "Camkk2 in myeloid cells is a key regulator of the immune-suppressive microenvironment in breast cancer," Nature Communications, vol. 10, no. 1, p. 2450, 2019.

[8] S. Kim, Y. Gao, T. Welte, H. Wang, J. Liu, M. Janghorban, K. Sheng, Y. Niu, A. Goldstein, N. Zhao et al., "Immuno-subtyping of breast cancer reveals distinct myeloid cell profiles and immunotherapy resistance mechanisms," Nature cell biology, vol. 21, no. 9, pp. 1113–1126, 2019.

[9] E. L. Kuan and S. F. Ziegler, "A tumor–myeloid cell axis, mediated via the cytokines il-1α and tslp, promotes the progression of breast cancer," Nature immunology, vol. 19, no. 4, pp. 366–374, 2018.

[10] Sawant, J. Deshane, J. Jules, C. M. Lee, B. A. Harris, X. Feng, and S. Ponnazhagan, "Myeloid-derived suppressor cells function as novel osteoclast progenitors enhancing bone loss in breast cancer," Cancer research, vol. 73, no. 2, pp. 672–682, 2013.

[11] S. B. Coffelt, K. Kersten, C. W. Doornebal, J. Weiden, K. Vrijland, C.-S. Hau, N. J. Verstegen, M. Ciampricotti, L. J. Hawinkels, J. Jonkers et al., "Il-17-producing γδ t cells and neutrophils conspire to promote breast cancer metastasis," Nature, vol. 522, no. 7556, pp. 345–348, 2015.

[12] Byrne, P. Savas, S. Sant, R. Li, B. Virassamy, S. J. Luen, P. A. Beavis, L. K. Mackay, P. J. Neeson, and S. Loi, "Tissue-resident memory t cells in breast cancer control and immunotherapy responses," Nature reviews Clinical oncology, vol. 17, no. 6, pp. 341–348, 2020.

[13] P. B. Olkhanud, B. Damdinsuren, M. Bodogai, R. E. Gress, R. Sen, K. Wejksza, E. Malchinkhuu, R. P. Wersto, and A. Biragyn, "Tumorevoked regulatory b cells promote breast cancer metastasis by converting resting cd4+ t cells to t-regulatory cells," Cancer research, vol. 71, no. 10, pp. 3505–3515, 2011.

[14] D. P. Hollern, N. Xu, A. Thennavan, C. Glodowski, S. Garcia-Recio, K. R. Mott, X. He, J. P. Garay, K. Carey-Ewend, D. Marron et al., “B cells and t follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer,” Cell, vol. 179, no. 5, pp. 1191–1206, 2019.

[15] Y. Gu, Y. Liu, L. Fu, L. Zhai, J. Zhu, Y. Han, Y. Jiang, Y. Zhang, P. Zhang, Z. Jiang et al., “Tumor-educated b cells selectively promote breast cancer lymph node metastasis by hspa4-targeting igg,” Nature medicine, vol. 25, no. 2, pp. 312–322, 2019.

[16] S. Garaud, L. Buisseret, C. Solinas, C. Gu-Trantien, A. de Wind, G. Van den Eynden, C. Naveaux, J.-N. Lodewyckx, A. Boisson, H. Duvillier et al., “Tumor-infiltrating b cells signal functional humoral immune responses in breast cancer,” JCI insight, vol. 4, no. 18, 2019.