Supplementary Materials for

Spatial Gene Expression Analysis Reveals Drivers of Extremely Early Lymph Node Metastasis in Breast Cancer

Satoi Nagasawa *et al.*

*Corresponding author. Email: ysuzuki@edu.k.u-tokyo.ac.jp

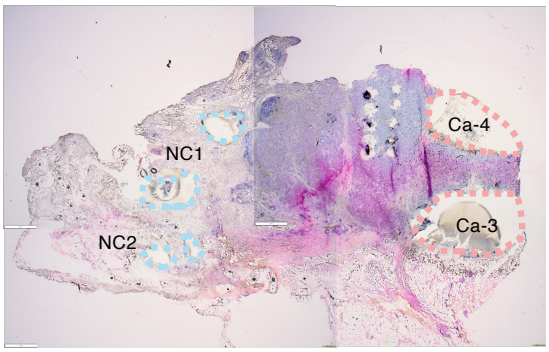
This PDF file includes:
Supplementary Figures S1 to S4

List of Supplementary Tables
Supplementary tables will be uploaded as excel files as they are large tables.

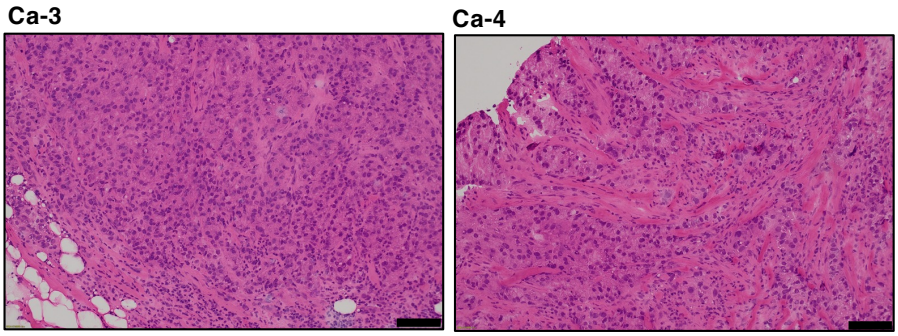
Supplementary Table S1.
Clinicopathologic information of sample
Supplementary Table S2.
General statistics RNA-seq, Enzymatic Methylation sequencing (EM-seq) and whole-genome short read sequencing (WGS)
Supplementary Table S3.
Custom gene list of Xenium
Supplementary Table S4.
Result of gene enrichement analysis across Xenium clusters
Supplementary Table S5.
Differentially expressed genes across Xenium clusters
Supplementary Table S6.
Differentially expressed genes of subsets within cancer cell of public single cell data.

Supplementary Figure 1

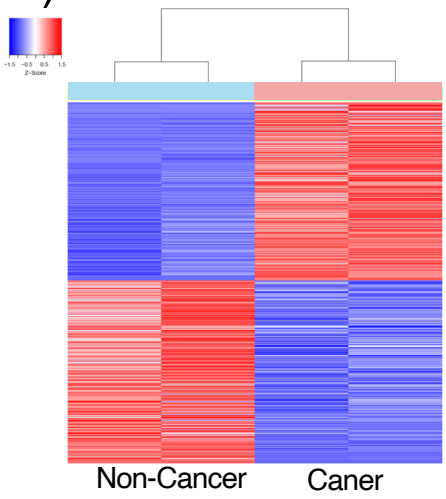
S1a)



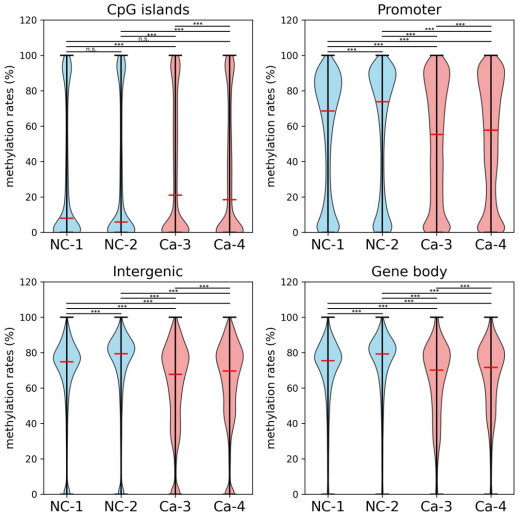
S1b)



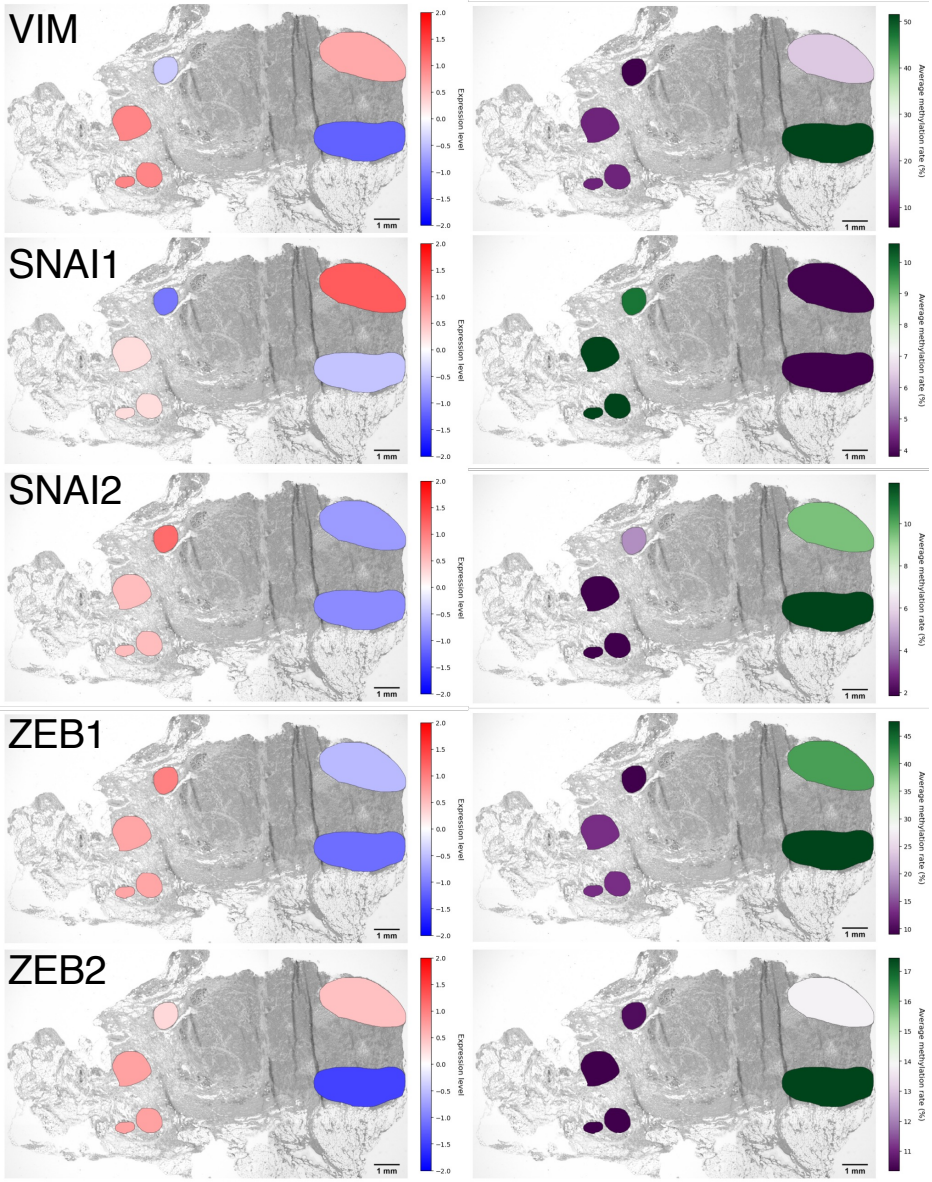
S1c)



S1d)



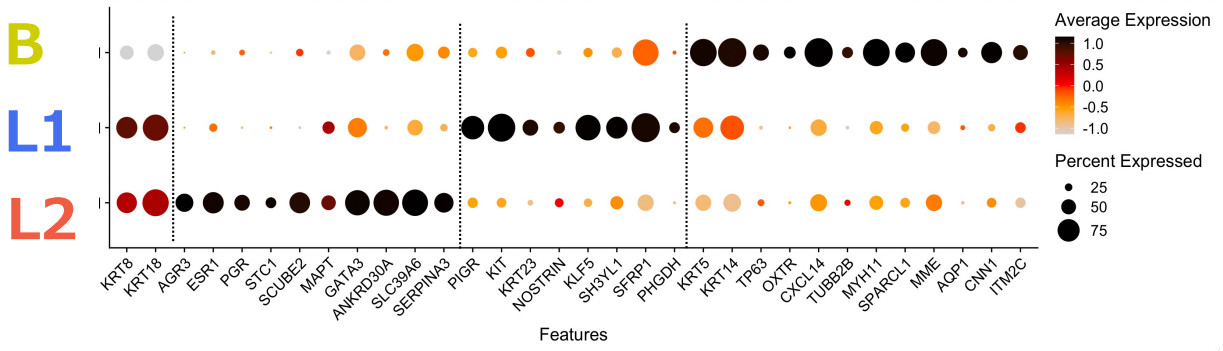
S1e)



Supplementary Figure l 1.

(**a**) HE staining image after macrodissection. (**b**) Enlarged HE staining image of the cancerous region (Ca-3 and Ca-4). Scale bar: 100 μm. (**c**) Heatmap showing DEGs between noncancerous and cancerous regions. Of the 4,584 detected genes, 2,248 were significantly upregulated in cancer regions. (**d**) DNA methylation patterns across genomic features in each region. Methylation levels were analyzed across CpG islands, promoters, gene bodies, and intergenic regions. Promoters were defined as 1 kb upstream to 500 bp downstream of the transcription start site. Intergenic regions excluded all gene-annotated intervals (exons, introns, and promoters) based on the reference GTF annotation. P-values were calculated using the two-sided Wilcoxon rank-sum test with Bonferroni correction. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$; n.s., not significant ($P > 0.05$). (**e**) Visualization of EMT-related gene expression (left) and promoter methylation (right) in each region using our custom viewer. Color scales represent z-scores (normalized from TPM) for gene expression and methylation rates for promoter regions.

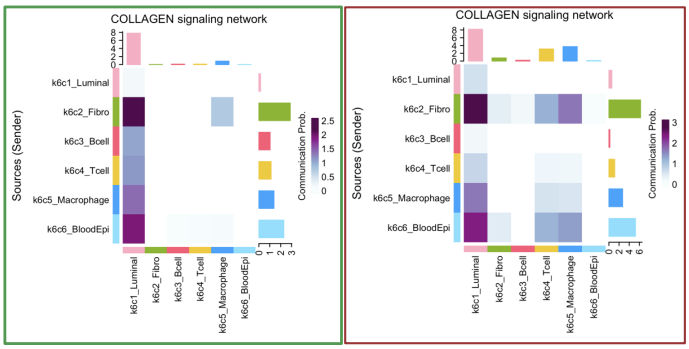
Supplementary Figure 2

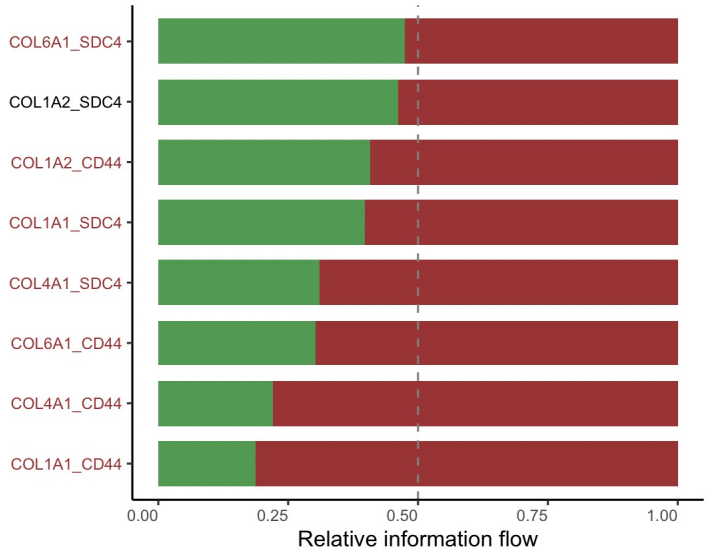


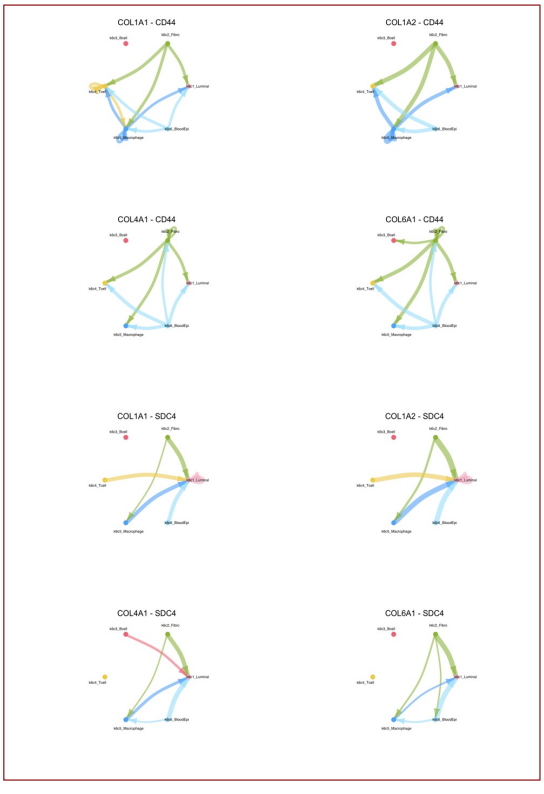
Supplementary Figure. 2 l Gene expression in clusters comprising normal lobules.
Dot plot showing expression of selected DEGs (x-axis) across normal lobule subsets (y-axis). Dot size represents the percentage of cells expressing the gene; color indicates mean expression relative to other subsets.

Supplementary Figure 3

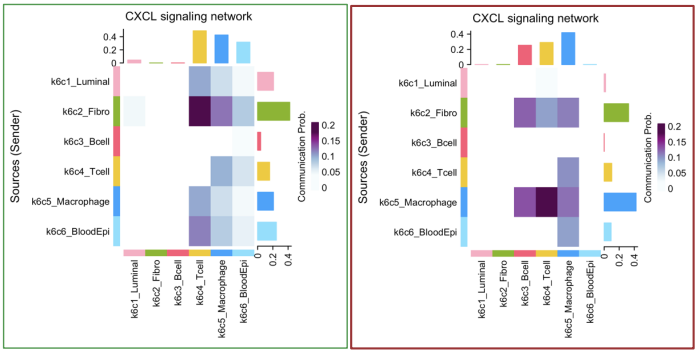
COLLAGEN signal

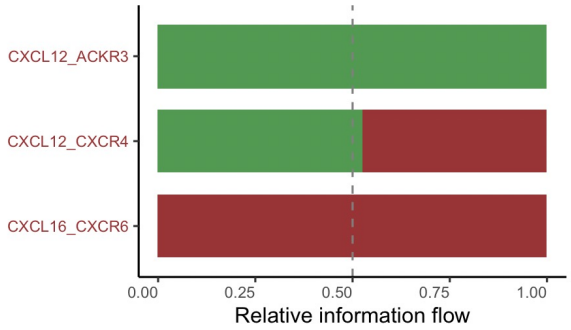


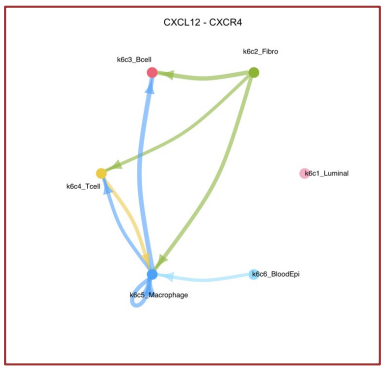


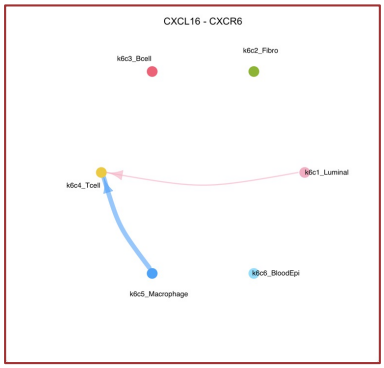


CXCL signal





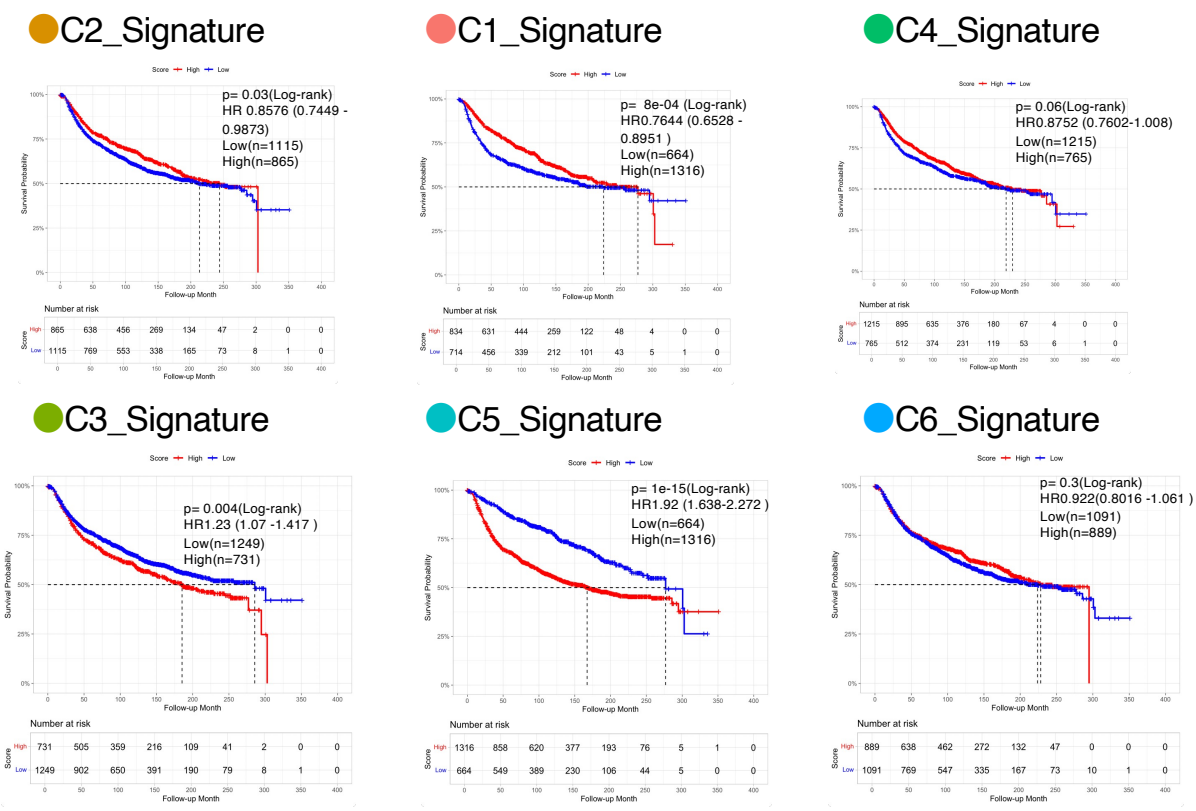




Supplementary Figure. 3 | CellChat analysis of TME in metastatic lymph nodes and the TdL.

Heatmaps of intercellular interactions via COLLAGEN and CXCL signaling in the TdL (green) and metastatic lymph nodes (dark red). Cell types are color-coded, as in Fig. 4a. Square color represents communication probabilities; blank squares denote zero probability. Bar chart ranks significant signaling pathways based on overall information flow differences between metastatic lymph nodes and the TdL. Circle plot depicts inferred COLLAGEN and CXCL signaling networks in metastatic lymph nodes and the TdL. Arrows and edges represent signaling directionality; circle color indicates cell type; circle size is proportional to cell count; edge thickness reflects interaction strength.

Supplementary Figure 4



Supplementary Figure/ 4 | Survival analysis of the METABRIC cohort based on Xenium signatures. Kaplan–Meier survival plots showing disease-free-survival for patients with high vs. low signature scores based on UCell scoring of the top 50 DEGs from Xenium subclusters (C1–C6). Cutoff values for high and low scores were determined using ROC analysis. P-values were calculated using the log-rank test. Adjusted HRs were derived from a multivariable Cox proportional hazard model using the low-score group as the reference.