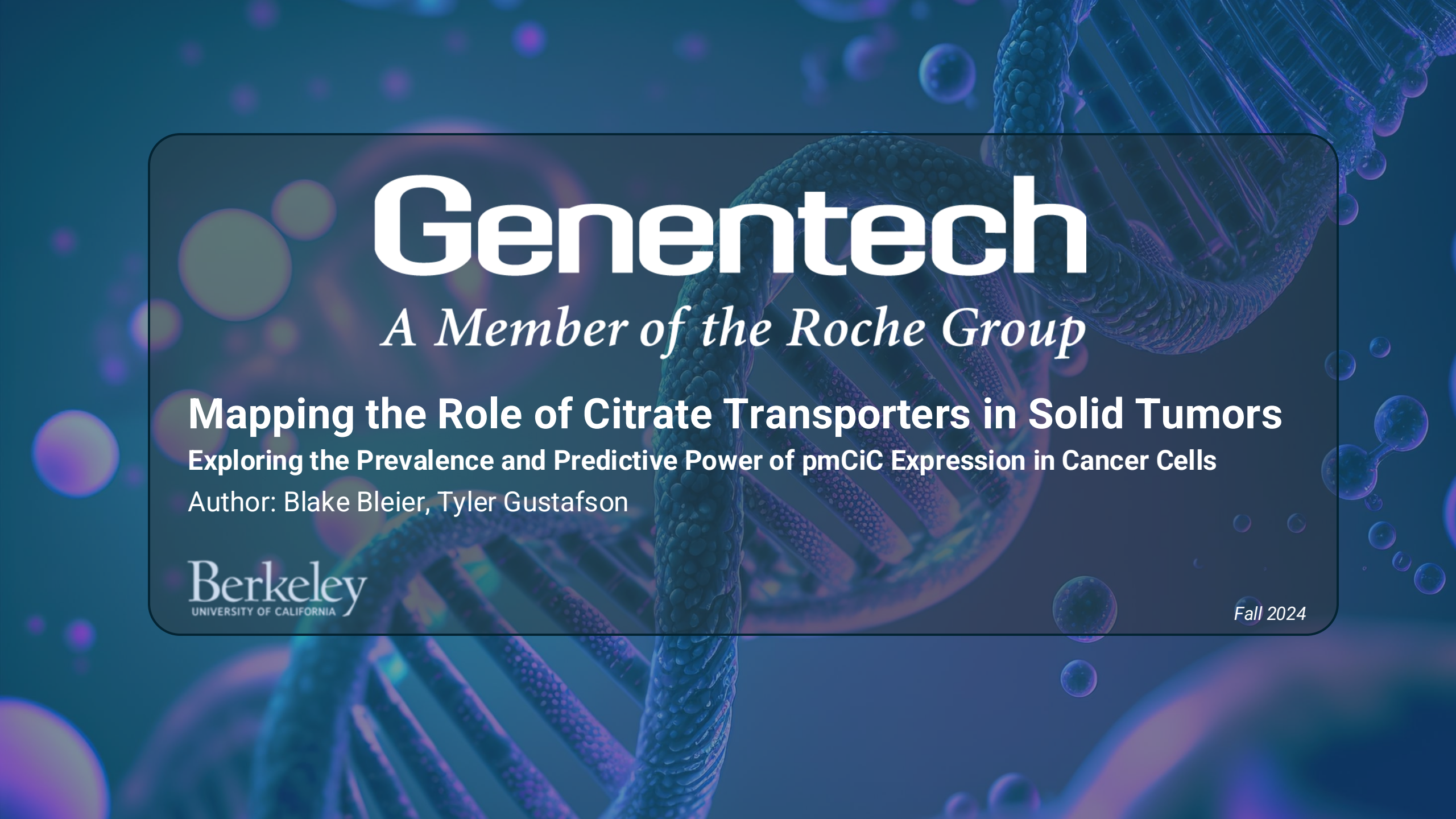


Mapping the Role of Citrate Transporters in Solid Tumors

Exploring the Prevalence and Predictive Power of pmCiC Expression in Cancer Cells

Author: Blake Bleier, Tyler Gustafson

Berkeley
UNIVERSITY OF CALIFORNIA

Fall 2024

Genentech
A Member of the Roche Group

Citrate Uptake and Cancer Progression Process (Simplified)

1 **Cancer cells have a rapid growth requirement:**

Cancer cells need to grow and divide quickly, requiring large amounts of energy and building materials like lipids.

2 **To support this growth, cancer cells look for alternative energy sources (citrate) from external sources:**

External citrate is imported into the cell to be converted into acetyl-CoA, a key component for lipid synthesis.

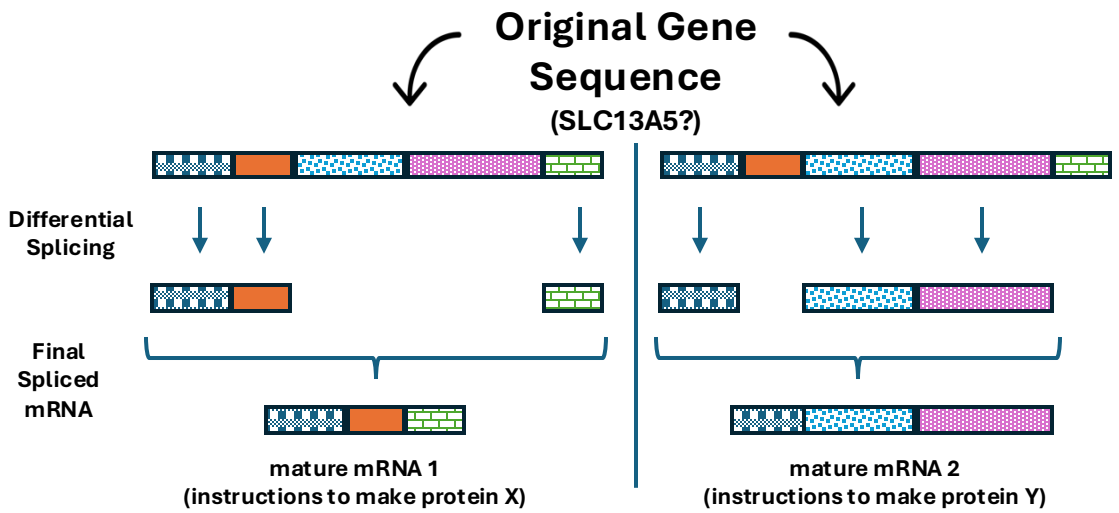
3 **To increase citrate uptake, cancer cells alter the differential expression of pmCiC transporter, making it overactive:**

Differential splicing may produce an overactive version of the pmCiC transporter, which enhances its ability to import citrate, fueling rapid growth.

4 **Ultimately, this citrate uptake promotes the cancer progression:**

The abundance of citrate allows for faster cell growth and division, contributing to the aggressiveness and progression of cancer.`

Overview of Differential Splicing

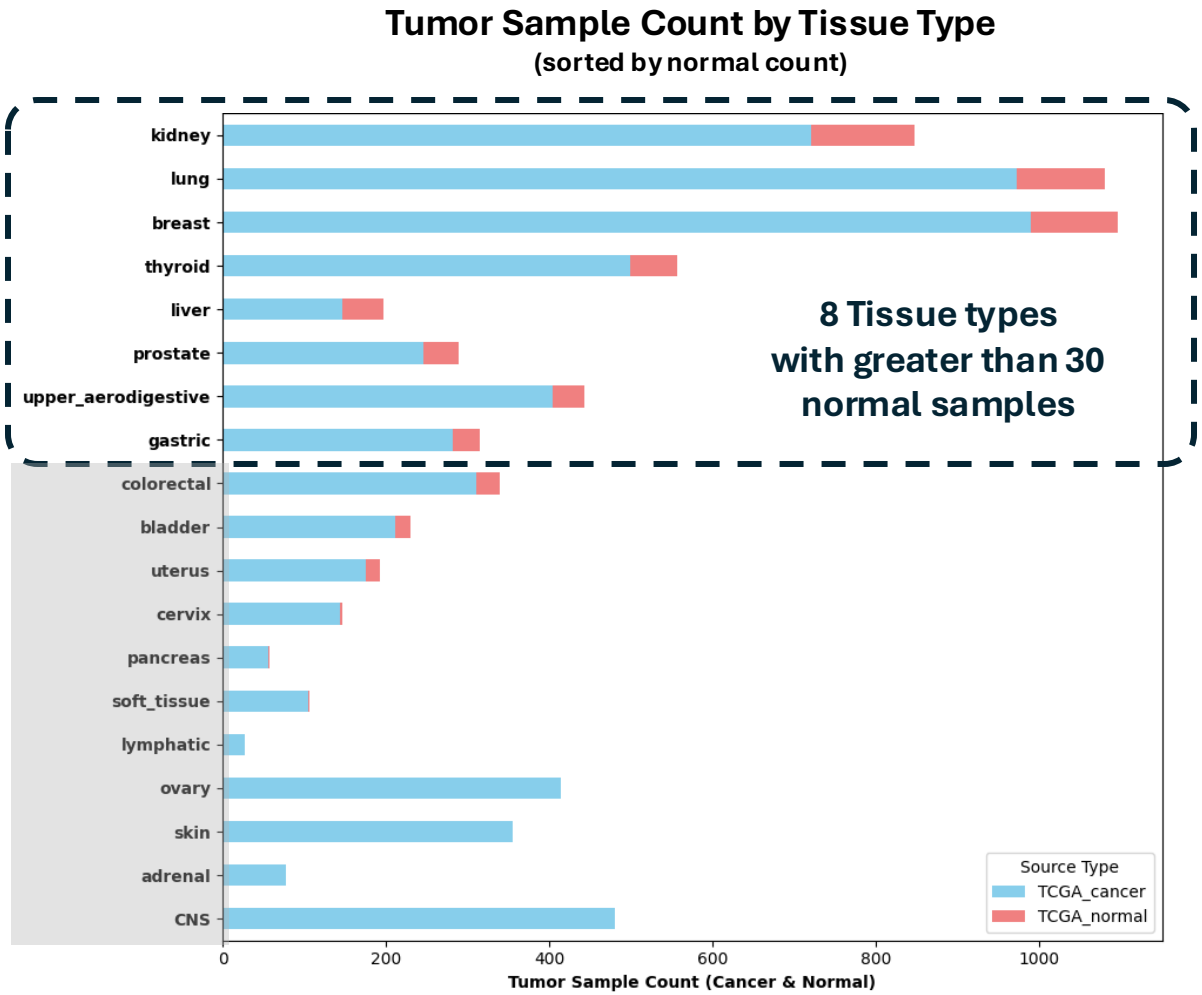


We see that our original gene can be expressed in multiple forms of mRNA through differential splicing, leading to the production of different protein variants with distinct functions.

The goal of this project is to identify the prevalence of the overactive, differentially expressed pmCiC in various cancer types

Data Ingestion Focus: Leveraging Sufficient Sample Sizes for Reliable Isoform Expression Comparisons

Genentech
A Member of the Roche Group



Focus on these select tissue groups, as they have a sufficient number of normal samples (greater than 30), allowing for a reasonable expectation of detecting differences between cancerous and normal samples across isoforms.

This will help us to:

- **Gain clearer insights into isoform expression patterns,** by comparing distributions between cancerous and normal tissues.
- **Use statistical tests like Mann-Whitney U and Wilcoxon rank-sum** to robustly assess significant differences in isoform expression.
- **Enhance the reliability of our findings** by ensuring sufficient sample size for robust hypothesis testing and exploratory data analysis

Placeholder

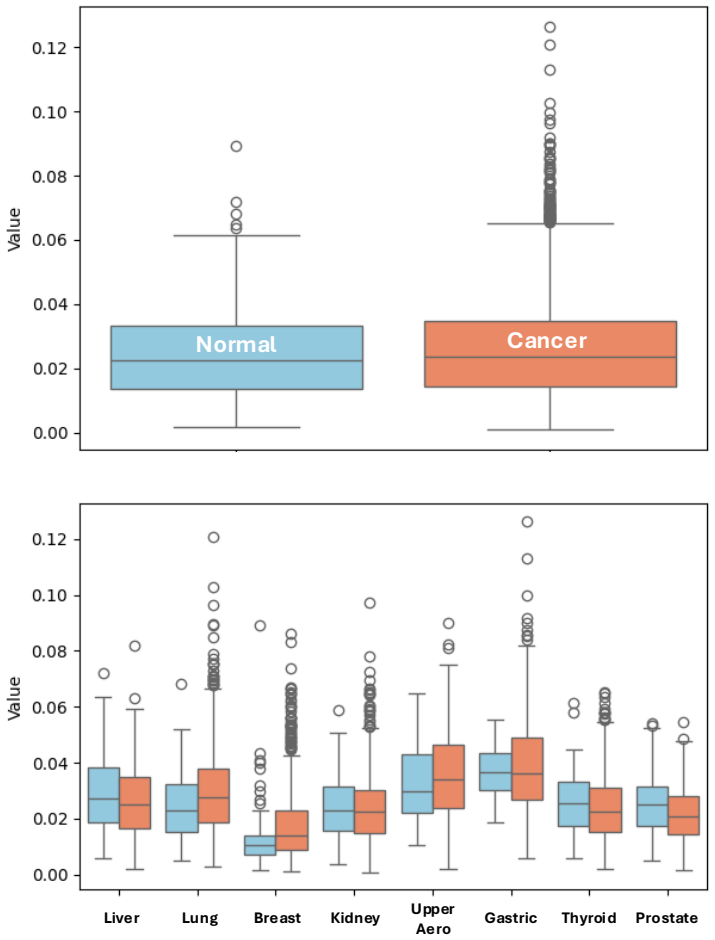
Comparative Analysis: A high-level examination of the distributions of plasma and mitochondrial mCiC, along with its precursor, indicating variations in expression levels across normal and cancer tumor samples.

Genentech
A Member of the Roche Group

Plasma Proportion

(NM_001256534.2)

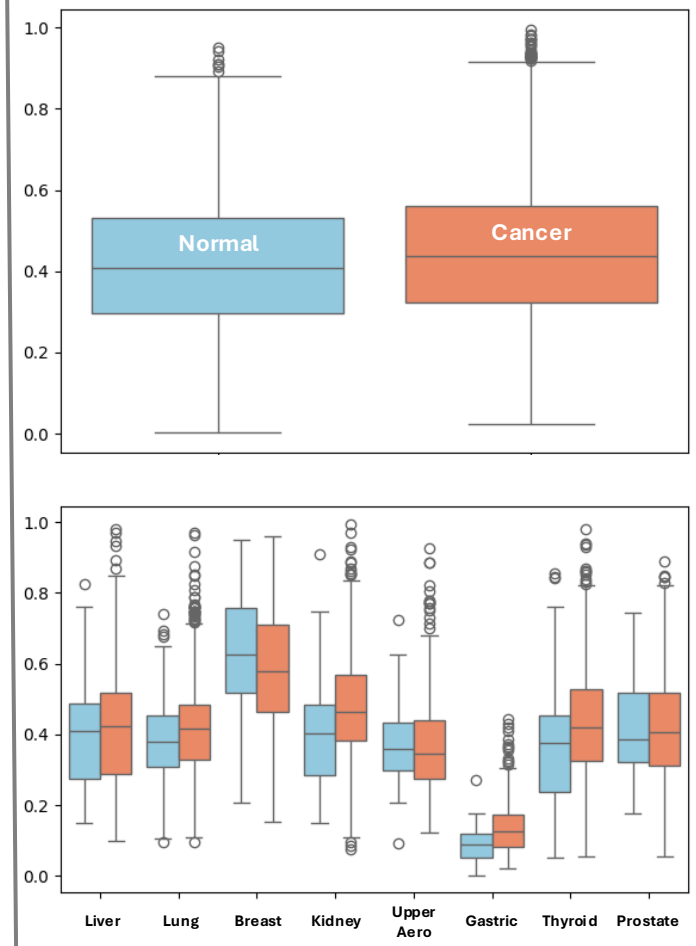
Cancer Generally Higher Expression



mCiC Proportion

(NM_001287387.2)

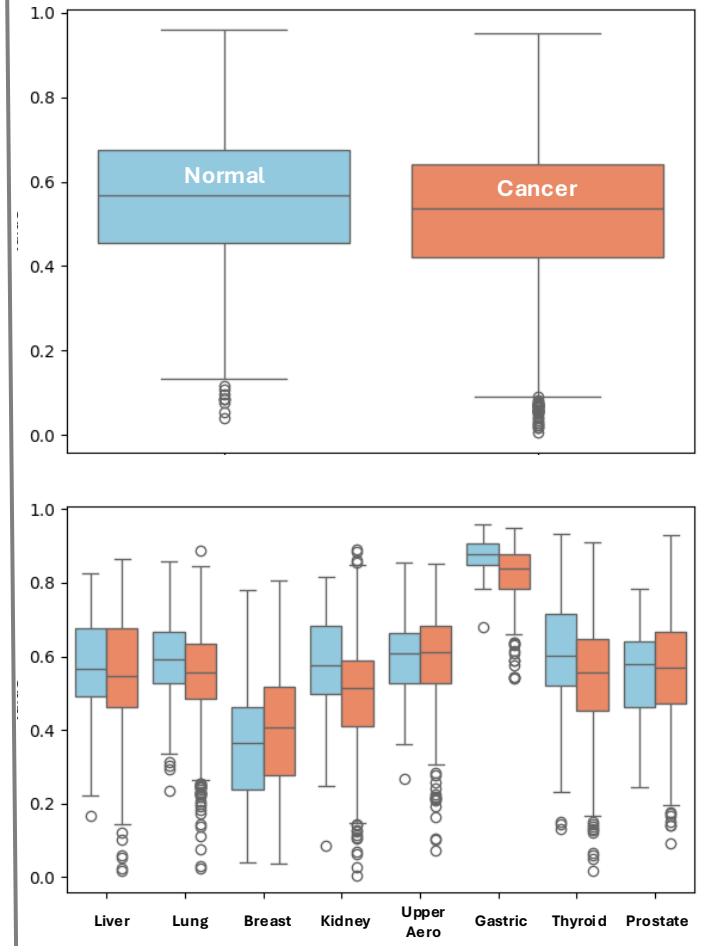
Cancer Generally Higher Expression



Precursor Proportion

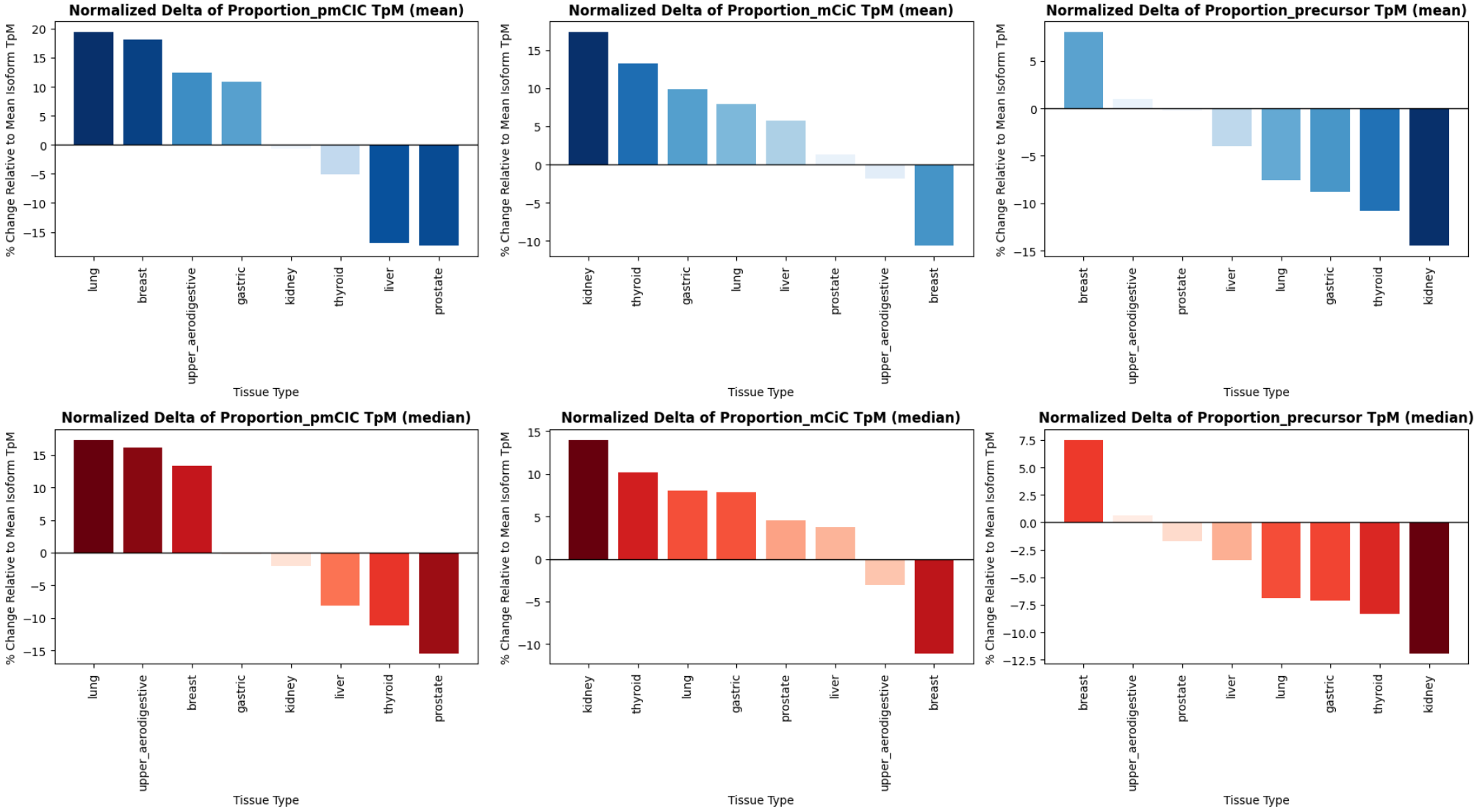
(NM_005984.5)

Normal Generally Higher Expression



Comparative Analysis: On this double click, we observe a consistent trend where the plasma isoform and mCiC show a majority increase, while the precursor shows a decrease.

Genentech
A Member of the Roche Group

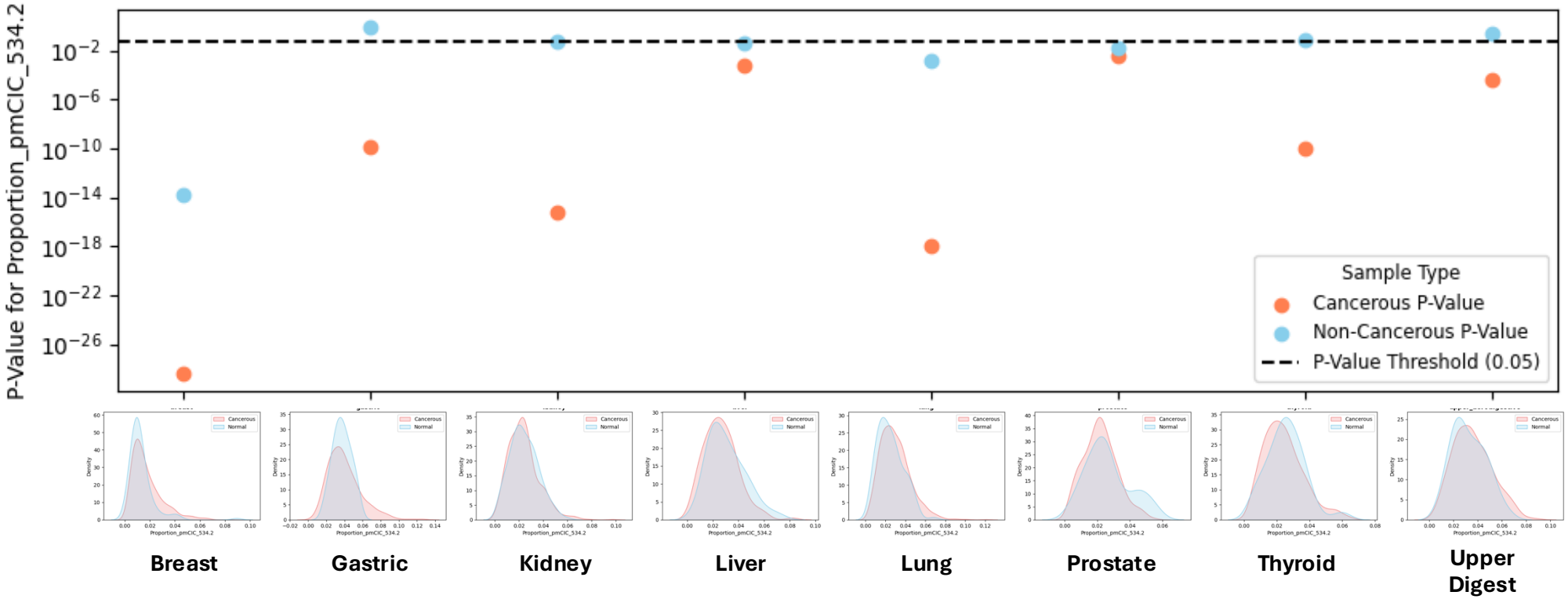


Statistical Approach (1/2): Before conducting our tests, it's important to assess the distribution and evaluate its normality to determine if it follows a Gaussian (normal) distribution. We can do this with a Shapiro-Wilk test.

Genentech
A Member of the Roche Group

Shapiro-Wilk P-Values Plasma Proportion (NM_001256534.2)

Remember the null hypothesis for this test is that the distribution is normally distributed



The plasma isoform distributions show a very low likelihood of normality, which is consistent with the other isoforms. As a result, it may be more appropriate to avoid using statistical tests that assume normality (i.e. parametric tests)

Statistical Approach (2/2): Now that we know the distributions are not normal, we can decide on our tests. We can proceed with a three-step approach to explore our hypothesis.

Genentech
A Member of the Roche Group

1. Are isoform expression levels significantly different between cancerous and normal samples?

(Step 1)

- **What?** We want to compare the expression levels of isoforms between cancerous and non-cancerous samples.
- **Implication?** This will give us a general understanding of whether there's a significant difference in expression between these two groups.
- **How?** Mann-Whitney U (Wilcoxon Rank-Sum)

2. Do different cancer types show variations in isoform expression?

(Step 2)

- **What?** Investigate if there is variation in isoform expression across different cancer types.
- **Implication?** A significant result suggests that isoform expression differs between cancer types, but it won't tell us exactly which types differ.
- **How?** Kruskal-Wallis Test

3. Does isoform expression differ between cancerous and non-cancerous samples in each cancer type?

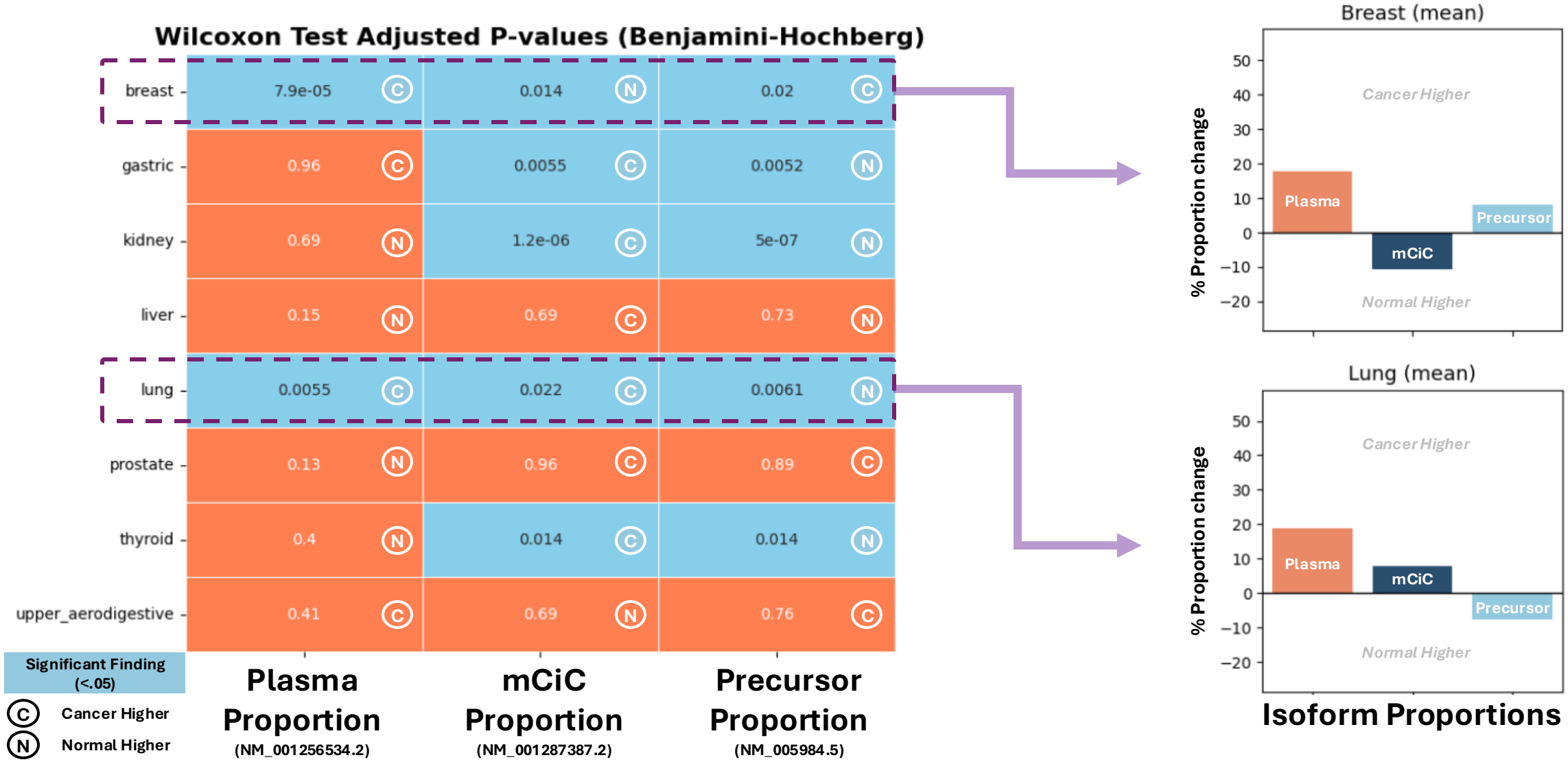
(Step 3)

- **What?** Perform pairwise comparisons between cancerous and non-cancerous samples for each specific cancer type.
- **Implication?** This helps us understand the specific differences in isoform expression for each cancer type.
- **How?** Mann-Whitney U (Wilcoxon Rank-Sum) for pairwise comparisons

Note: For all tests, we must account for multiple comparisons and apply a correction, such as the Benjamini-Hochberg method

Jumping ahead to the reveal… We see that our isoforms that show statistically different expression levels across the isoforms are breast and lung tissue types.

Genentech
A Member of the Roche Group



Double clicking into Breast and Lung … Next, we'll dive into tissue-specific subgroups to explore spikes in pmCiC proportions and key phenotypic data..

Genentech
A Member of the Roche Group

Phenotypic subgroups that showed elevated pmCiC proportions within their category

	Phenotypical Category	Breast	Lung
Subtype	*Tumor Stage*	**N/A** Not Significant Kruskal-Wallis Test: P-Value = 0.4952	**Stage 2 and 3** Not Significant Kuskal-Wallis Test: P-Value = 0.34773
Subtype	*DSS Binned*	**8+ Years** Not Significant Wilcoxon Ranked Sum Test: P-Value = 0.4952	**0-3 years** Mixed Significance (Potentially 0-3 years) Kuskal-Wallis Test: P-Value = 0.05801
Subtype	*Selected*	**Basal & Normal** Mixed Significance Pairwise (Appears Both Valid) Kuskal-Wallis Test: P-Value = 0.00000001	**LUAD.6, Classical, Primitive** **561 (58%) missing Values*
Subtype	*Immune*	**C3.Inflammatory &C6.TGF-Beta Dominated** Mixed Significance Pairwise (Appears Valid C3) Kuskal-Wallis Test: P-Value = 0.0001	**C1 (Wound Healing)** Significant Wilcoxon Ranked Sum Test: P-Value = 5.526e-08
Subtype	*mRNA*	Basal Normal	Classical Primitive

Duplicative info /
High Missing Value Counts

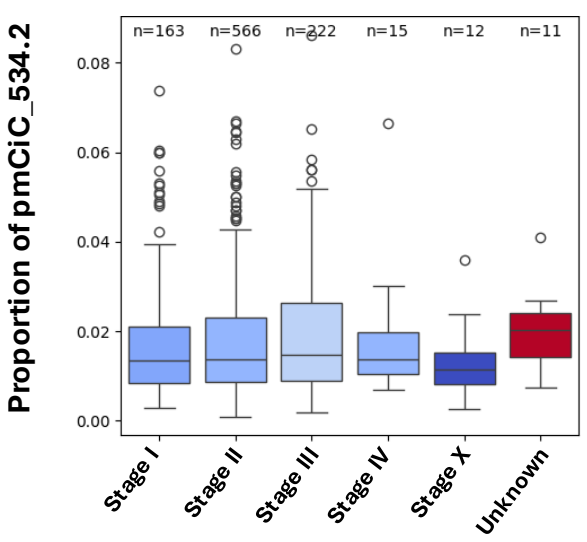
Looking at breast tissue type data (1/2)... Here we will conduct the appropriate statistical tests with corrections as needed for the phenotype groups that showed potentially elevated expressed relationships

Genentech
A Member of the Roche Group

Breast Phenotypic Data

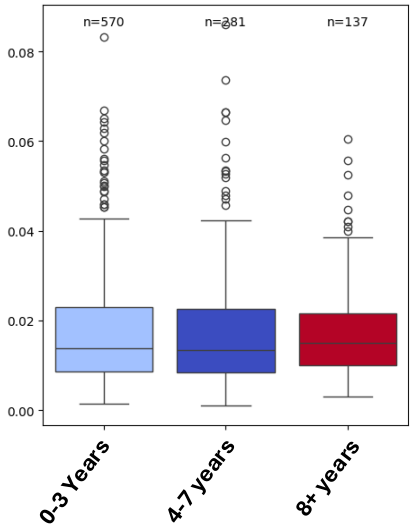
Tumor Stage (Grouped)

Not Statistically Significant
(Kruskal-Wallis + Pairwise)



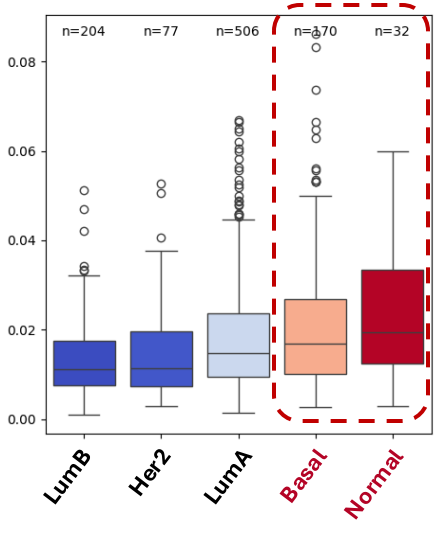
DSS (Binned)

Not Statistically Significant
(Wilcoxon Ranked Sum)



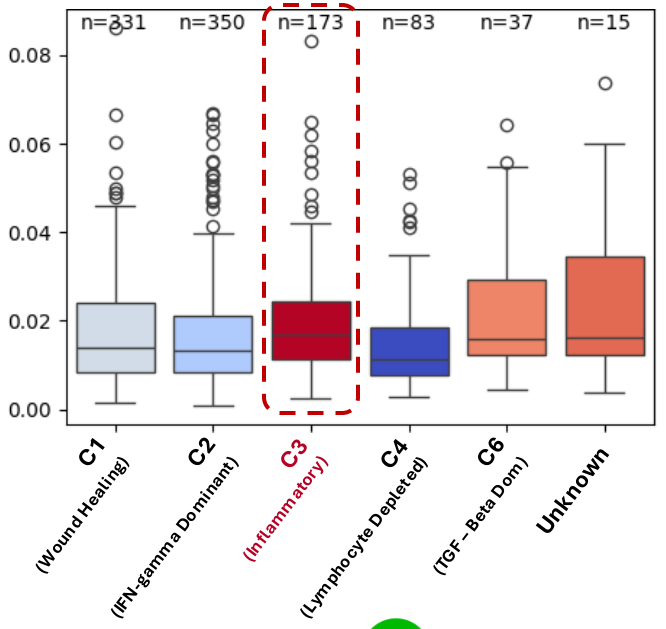
Subtype - Selected

Statistically Significant
(Kruskal-Wallis + Pairwise)



Subtype - Immune

Statistically Significant
(Kruskal-Wallis + Pairwise)



Both subtype-selected and subtype-immune demonstrated statistically significant elevation amongst the highlighted categories

Looking at breast tissue type data (2/2)... for both subtype selected and immune we needed to do additional pairwise testing to determine the significant categories

Genentech
A Member of the Roche Group

Breast Phenotypic Data

Subtype – Immune

Pairwise Double Click (P-Values)

Number of pairwise comparisons (Bonferroni dividing by): 15

	Wound Healing (C1)	IFN-gamma Dominant (C2)	Inflammatory (C3) (High Expression)	Lymphocyte Depleted (C4)	TGF-beta Dominant (C6) (High Expression)	Unknown
Wound Healing (C1)	-	1.000000	**0.014793***	1.000000	1.000000	1.000000
IFN-gamma Dominant (C2)	1.000000	-	**0.001802****	1.000000	0.529852	0.924975
Inflammatory (C3)	0.014793*	0.001802**	-	0.003725**	1.000000	1.000000
Lymphocyte Depleted (C4)	1.000000	1.000000	**0.003725****	-	0.182472	0.390205
TGF-beta Dominant (C6)	1.000000	0.529852	1.000000 (Also elevated)	0.182472	-	1.000000
Unknown	1.000000	0.924975	1.000000 (Also elevated)	0.390205	1.000000	-

* Statistically Significant ($p < 0.05$)
** Highly Significant ($p < 0.01$)
*** Very Highly Significant ($p < 0.001$)

Subtype – Selected

Pairwise Double Click (P-Values)

Number of pairwise comparisons (Bonferroni dividing by): 10

	Basal (High Expression)	Her2 (Mid Expression)	LumA (Lower Expression)	LumB (Lower Expression)	Normal (High Expression)
Basal	-	0.002403**	0.862225	0.000003***	1.000000 (Also elevated)
Her2	**0.002403****	-	0.039695*	1.000000	**0.013366***
LumA	0.862225	0.039695*	-	0.000038***	0.769803
LumB	**0.000003*****	1.000000	0.000038***	-	**0.002055****
Normal	1.000000 (Also elevated)	0.013366*	0.769803	0.002055**	-

* Statistically Significant ($p < 0.05$)
** Highly Significant ($p < 0.01$)
*** Very Highly Significant ($p < 0.001$)

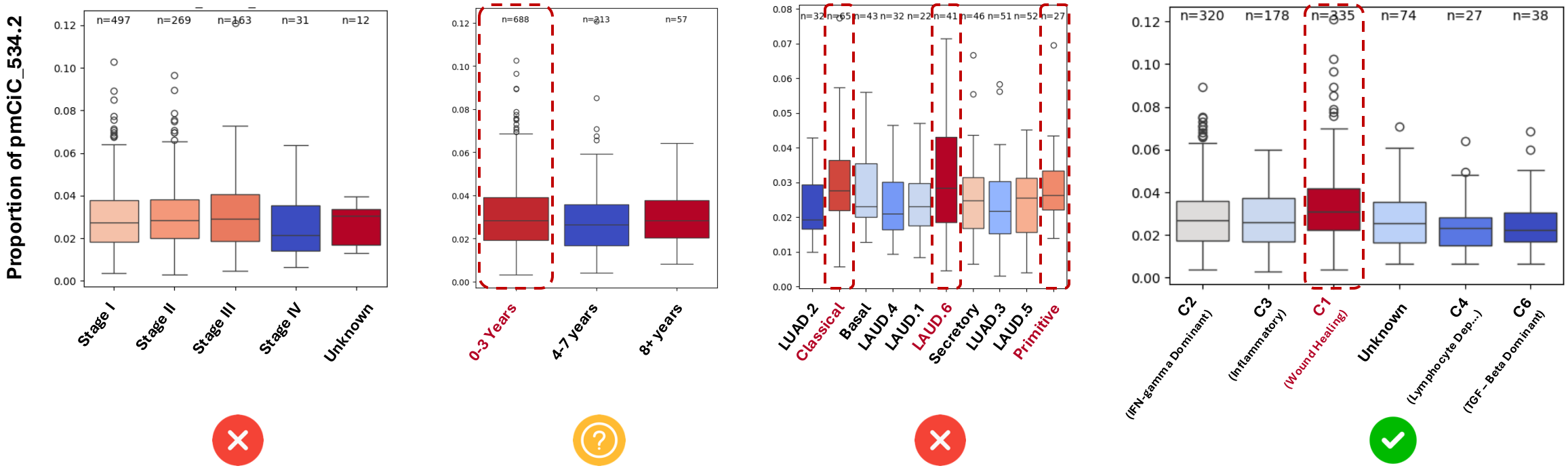
We observe that C3, Basal, and Normal exhibit statistically significant elevation compared to the other categories.

Looking at lung tissue type data (1/2)... Here we will conduct the appropriate statistical tests with corrections as needed for the phenotype groups that showed potentially elevated expressed relationships

Genentech
A Member of the Roche Group

Lung Phenotypic Data

Tumor Stage (Grouped)	DSS (Binned)	Subtype - Selected	Subtype - Immune
Not Statistically Significant (Kruskal-Wallis)	**TBD Statistical Significance** (Kruskal-Wallis)	***561 (58%) missing values** (Not tested)	**Statistically Significant** (Wilcoxon Ranked Sum)



Both C1 and potentially DSS_Binned 0-3 years demonstrated statistically significant elevation amongst the highlighted categories

Looking at lung tissue type data (2/2)... for DSS binned we needed to do additional pairwise testing to determine the significant categories

Genentech
A Member of the Roche Group

Lung Phenotypic Data

DSS Binned

Pairwise Double Click (P-Values)

Number of pairwise comparisons (Bonferroni dividing by): 3

	0-3 years (Elevated)	4-7 years (Low)	8+ years (Elevated)
0-3 years	-	0.053577	1.000000 (Also Elevated)
4-7 years	**0.053577** (Significant with no correction)	-	0.671054
8+ years	1.000000 (Also Elevated)	0.671054	-

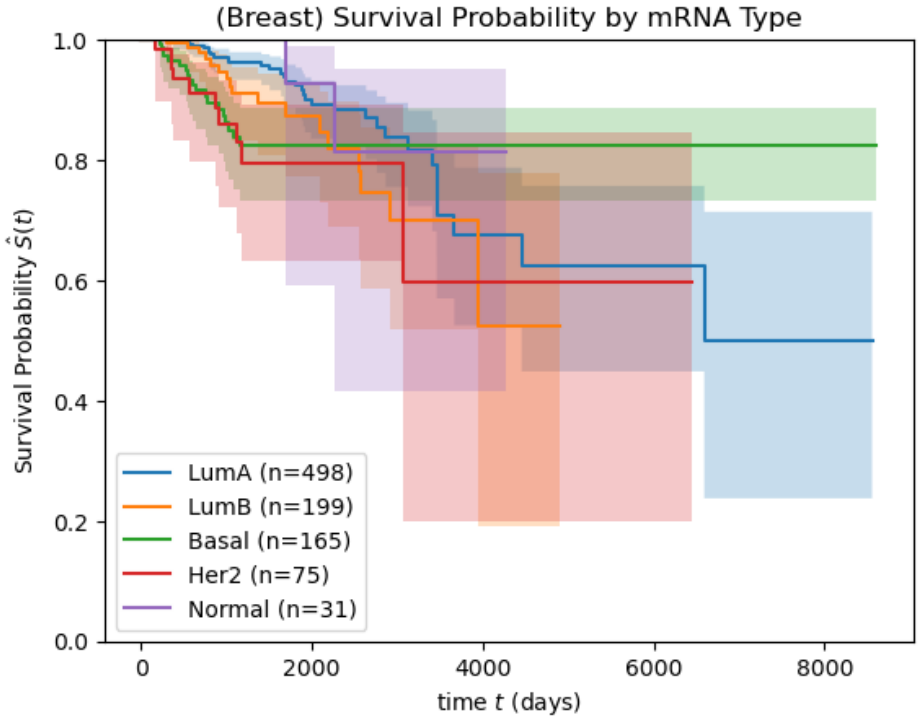
* Statistically Significant ($p < 0.05$)
** Highly Significant ($p < 0.01$)
*** Very Highly Significant ($p < 0.001$)

We observe that 0-3 years may exhibit statistically significant elevation compared to the other categories (depending on the Bonferroni correction)

Survival Analysis… Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Genentech
A Member of the Roche Group

Breast Phenotypic Data



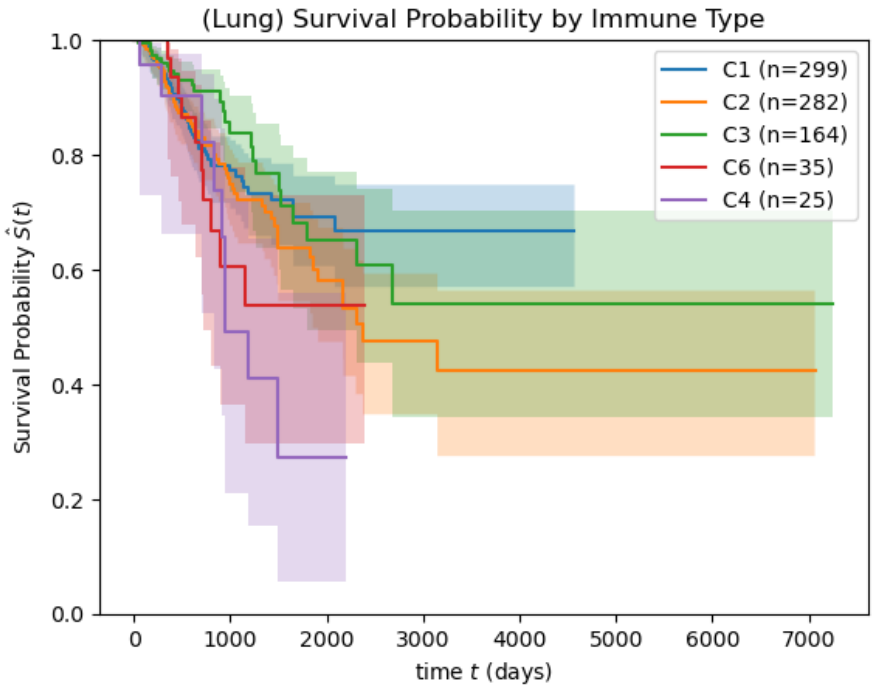
Visual

Highest to lowest pmCIC ratio:
Normal > Basal > LumA > Her2 > LumB
Longest to shortest survival:
Normal ~ Basal > LumA ~ Her2 ~ LumB

Stat Sig

Mantel-Cox (log-rank) significant w/ Bonferroni:
LumA vs Her2 (p=0.018)

Lung Phenotypic Data



Visual

Highest to lowest pmCIC ratio:
C1 > C2 ~ C3 > C4 ~ C6
Longest to shortest survival:
C1 > C2 ~ C3 ~ C6 > C4

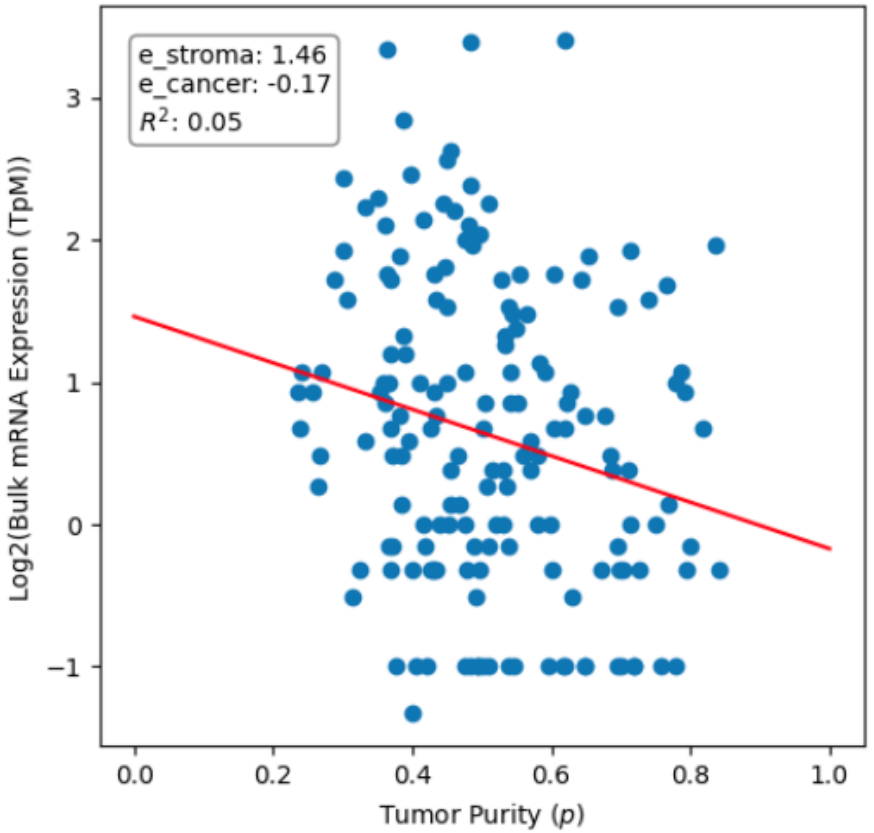
Stat Sig

Mantel-Cox (log-rank) significant w/ Bonferroni:
C3 vs C4 (p=0.029)

Using regression to break down gene expression by cell type, with *e_stroma* showing expected expression in stromal cells and *e_cancer* for cancer cells, helping us see how expression changes with tumor purity

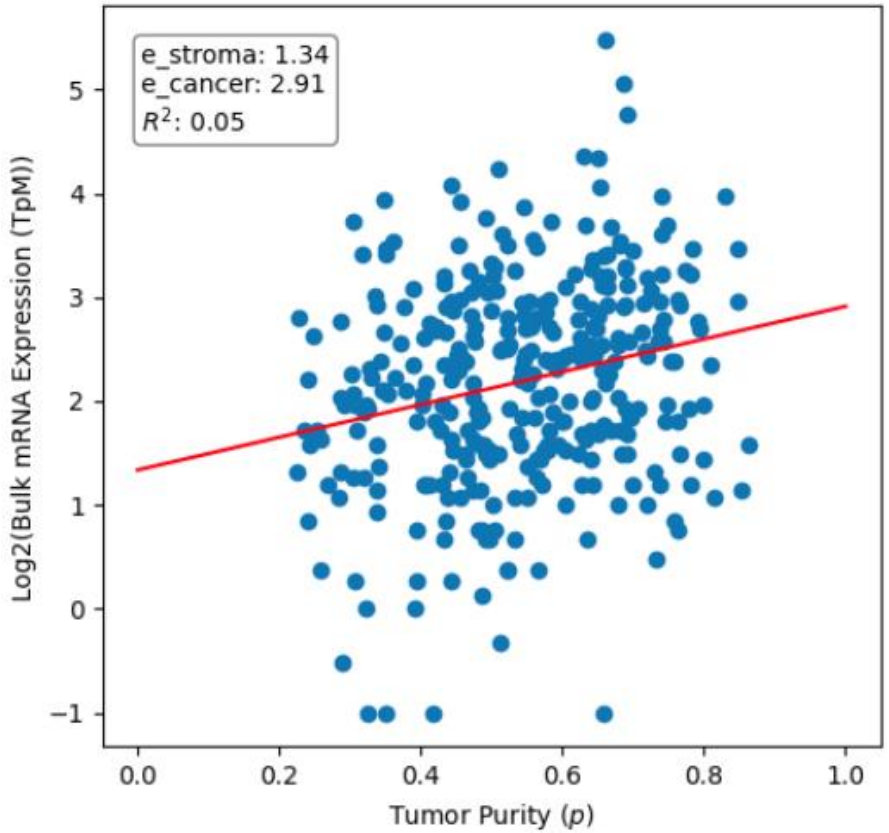
Genentech
A Member of the Roche Group

Breast | C3 - Inflammatory



Breast (C3 - Inflammatory): Although *e_stroma* (1.46) is higher than *e_cancer* (-0.17) as anticipated, the low *R*2=0.05 highlights a weak association between purity and expression, pointing to other influential factors in inflammatory breast tissue.

Lung | C1 – Wound Healing



Lung (C1 - Wound Healing): With a low *R*2=0.05, *e_cancer* (2.91) unexpectedly exceeds *e_stroma* (1.34), suggesting that factors beyond purity may drive gene expression in wound healing within lung tissue