A Bioinformatic Analysis of *CDH1* and its Corresponding Protein E-cadherin in Breast Carcinoma

Alexander Kim

William Fremd High School

alexk747air@gmail.com

Abstract

The *CDH1* gene and its corresponding protein, E-cadherin, are essential in cell-cell adhesion and signaling pathways, such as Wnt/β-catenin and PI3K/Akt. They are mildly dysregulated in breast cancer. Here, a bioinformatic analysis of genomic and proteomic data on breast cancer patients was performed. Initial bioinformatic analysis was performed with web-based portals such as cBioPortal and STRING db. Then, genomic and proteomic data were compared using DepMap and UALCAN. A statistical analysis to determine significant differences between expression data was performed with R software. A mildly significant statistical difference was determined between *CDH1* and *EGFR* gene expression in breast invasive ductal carcinoma (non-overlapping 95% CI), and statistically significant differences in *CDH1* protein expression were identified between primary tumor tissue compared to normal tissue, stage 2 and stage 3 breast cancer compared to normal tissue, and luminal and HER2-positive breast cancer compared to normal tissue. This study provides further evidence to support *CDH1* as a tumor suppressor gene in early breast cancer stages and suggests that *CDH1* may play a more oncogenic role in late-stage breast cancer. However, further validation is warranted.

Introduction

Cancer is an omnipresent pathology caused by the uncontrolled proliferation of abnormal cells. Cancer-inducing aberrations can include both intrinsic and extrinsic factors[1], but regardless of the cause, most forms of cancer exhibit the "hallmarks of cancer": evasion of growth suppressors, avoidance of immune destruction, enablement of replicative immortality, tumor-promoting inflammation, activation of invasion and metastasis, inducement of angiogenesis, genome instability, and mutation, resistance to apoptosis, deregulation of cellular energetics, and the sustentation of proliferative signaling[2].

Breast cancer, specifically, affects mostly women, and the rate of breast cancer incidence has risen in the United States from 2010-2019[3]. Investigating the genetic mutations that lead to the formation of breast cancer is crucial for improved prognosis. E-cadherin, or epithelial cadherin, is a calcium-dependent transmembrane glycoprotein involved in cell-cell adhesion. Mutations in the gene *CDH1*, and its corresponding protein, E-cadherin, have been commonly implicated in

the literature in breast cancer[4–8]. *CDH1* has also been implicated along with other oncologically significant genes, notably *PIK3CA*[9] and *EGFR*[10].

Loss of *CDH1* function may lead to carcinogenesis through multiple pathways (Figure 1). E-cadherin regulates cell-cell adhesion[11], and its loss of function due to alteration can lead to cellular migration, a hallmark of cancer. E-cadherin is also a key protein in the epithelial-to-mesenchymal transition (EMT) process[12]. Upregulation of N-cadherin (expressed by *CDH2*) and downregulation of E-cadherin are commonly implicated in EMT. Additionally, E-cadherin plays a role in the canonical WNT/β-catenin pathway, where it produces the β-catenin proteins essential to the pathway. Overexpression of the WNT signaling pathway has been attributed to cancer development by allowing for uncontrolled cell proliferation[13]. Finally, E-cadherin is an important member of the PI3K/Akt signaling pathway, where it regulates *PTEN* protein levels[14], inhibiting apoptosis.

This study was aimed to bioinformatically analyze *CDH1* through genomic and proteomic datasets to clarify its role in breast cancer.

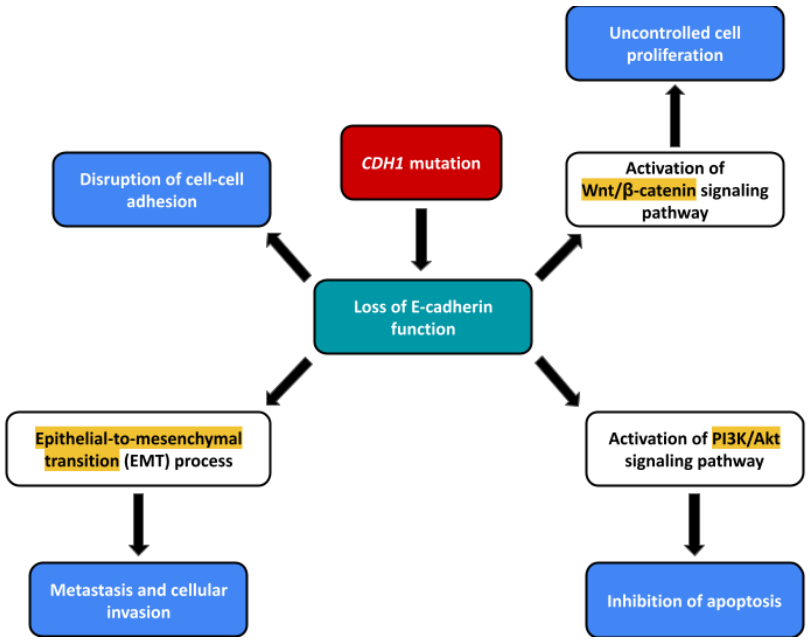

Figure 1: Summary of CDH1 pathways that may lead to carcinogenesis. E-cadherin plays a role in cell-cell adhesion, the epithelial-to-mesenchymal transition process, the WNT/β-catenin signaling pathway, and the PI3K/Akt signaling pathway.

Results

Truncating mutations of *CDH1* were most common in breast cancer

To assess *CDH1* mutation details in breast cancer patients, genomic data from the TCGA (The Cancer Genome Atlas) Firehose Legacy and METABRIC (Molecular Taxonomy of Breast

Cancer International Consortium) datasets on cBioPortal were used to analyze mutations of *CDH1* between normal and tumor tissues. The results showed that out of the total 963 samples in the TCGA dataset, *CDH1* was altered in 13% of cases ($n = 129$), and of the total 2051 samples in the METABRIC dataset, *CDH1* was altered in 11% of cases ($n = 228$). For the TCGA dataset, 18 alterations were missense mutations, 83 were truncating mutations, 10 were splice mutations, and 1 was an inframe mutation (Figure 1a). For the METABRIC dataset, 26 alterations were missense mutations, 154 were truncating mutations, 18 were splice mutations, and 6 were inframe mutations (Figure 1c). Furthermore, the OncoKB™ knowledge base categorized alterations with likely oncogenic effects as 'driver' genes: for the TCGA dataset, these 'driver' genes comprised 95 total alterations, including 2 missense mutations, all 83 truncating mutations, and 10 splice mutations, and for the METABRIC dataset, these 'driver' genes comprised 181 total alterations, including 9 missense mutations, all 154 truncating mutations, and 18 splice mutations. Thus, every truncating mutation in the TCGA dataset was identified as a likely 'driver' gene with oncogenic effects, which was confirmed by observations from the METABRIC dataset.

A modal value for the mutations in both the TCGA and METABRIC datasets was observed in the protein change Q23*, with a relative frequency of 4.65% ($n = 6$) among all alterations in the TCGA dataset and a relative frequency of 6.86% ($n = 14$) among all alterations in the METABRIC dataset. This protein change comprised a nonsense mutation at position 68772218 on chromosome 16 and was rated 'likely oncogenic' by the OncoKB™ knowledge base.

The following alterations were the next most frequent in the TCGA dataset, with a relative frequency of 3.10% ($n = 4$): R63* (position 68835596), P127Afs*41 (position 68835780 to 68835781), and G278*/G278R/G278V/X278_splice (positions 68844244, 68844244, 68845587, and 68844245, respectively). In the METABRIC dataset, the only other significant alteration was the protein change Q610*, with a relative frequency of 3.92% ($n = 8$). The frequency summaries of these mutations in the TCGA and METABRIC datasets are given by the lollipop plots in Figure 1b and Figure 1d, respectively.

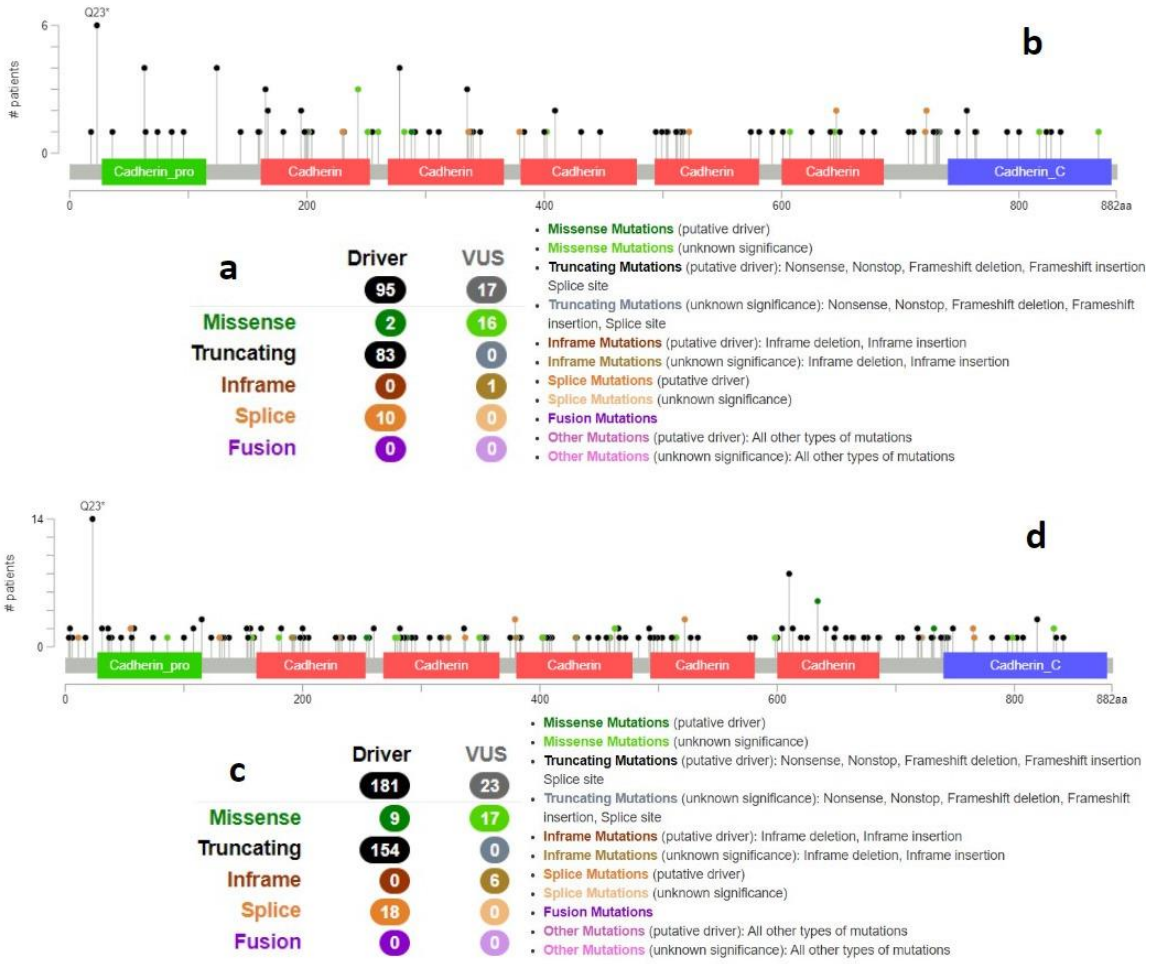

Figure 2: CDH1 mutation summaries from cBioPortal (https://www.cbioportal.org, accessed on 1 September 2022)

(a) Total mutation survey from TCGA dataset. There were 112 total mutations in the samples from this dataset, and the majority of mutations were truncating mutations.

(b) Lollipop plot of mutations from TCGA dataset. Q23 was identified as the most common alteration from this dataset.*

(c) Total mutation survey from METABRIC dataset. There were 204 total mutations in the samples from this dataset, and the majority of mutations were again truncating mutations.

(d) Lollipop plot of mutations from METABRIC dataset. Q23 was again identified as the most common alteration from this dataset.*

CDH1 **is most commonly altered in breast invasive lobular carcinoma**

It was observed that *CDH1* was most commonly altered in breast invasive lobular carcinoma based on the TCGA dataset, which was confirmed by observations from the METABRIC dataset.

For samples of breast invasive lobular carcinoma in the TCGA dataset (Figure 2a), a gene alteration was observed in 56.4% of cases ($n = 97$), with 51.74% contributed by mutations ($n = 89$), 2.91% by deep deletions ($n = 5$), and 1.74% by multiple alterations ($n = 3$). For breast mixed ductal and lobular carcinoma, a mutation was observed in 11.11% of cases ($n = 3$), and all gene alterations were mutations. Finally, for breast invasive ductal carcinoma, a gene alteration was observed in 3.86% of cases ($n = 28$), with 2.07% contributed by mutations ($n = 15$), 0.55% contributed by amplification ($n = 4$), and 1.24% contributed by deep deletions ($n = 9$).

Data from the METABRIC dataset showed the same trends (Figure 2b). A gene alteration was observed in 53.66% of cases of breast invasive lobular carcinoma ($n = 88$), with 51.22% contributed by mutations ($n = 84$), 1.83% by deep deletions ($n = 3$), and 0.61% by multiple alterations ($n = 1$). For breast mixed ductal and lobular carcinoma, a gene alteration was observed in 15.64% of cases ($n = 33$), with 13.27% contributed by mutations ($n = 28$), 1.9% by deep deletions ($n = 4$), and 0.47% by multiple alterations ($n = 1$). Finally, for breast invasive ductal carcinoma, a gene alteration was observed in 6.61% of cases ($n = 105$), with 5.16% contributed by mutations ($n = 82$), 0.25% by amplification ($n = 4$), and 1.2% by deep deletions ($n = 19$).

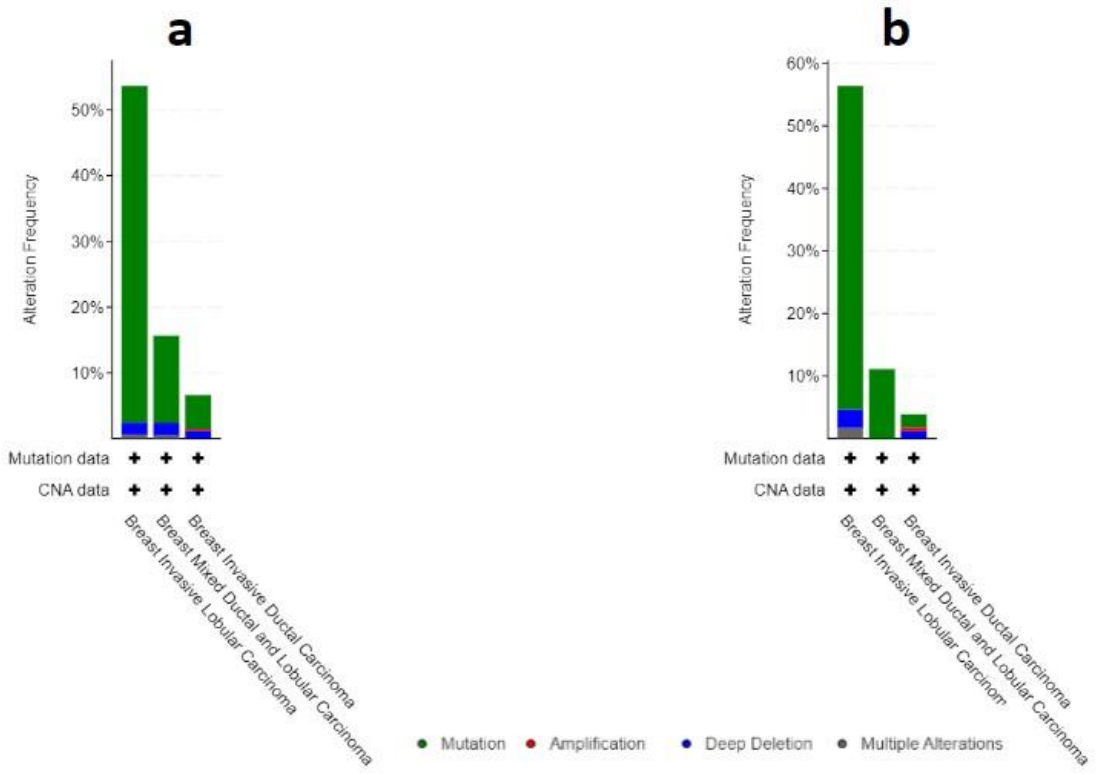

Figure 3: Alteration summary by subtype from cBioPortal (https://www.cbioportal.org, accessed on 1 September 2022). Invasive lobular breast carcinoma had the highest percentage of mutations in both TCGA (a) and METABRIC (b) datasets.

Protein interaction networks show connection between *CDH1*, *EGFR*, *CTNNA1*, *CTNNB1*, and *CTNND1*

To analyze *CDH1*'s protein interactions, STRING database was utilized.

EGFR was identified as a predicted functional partner to *CDH1* with a combined score of 0.999. Protein-protein interaction between *CDH1* and *EGFR* were detected by affinity chromatography technology assays, ubiquitin reconstruction assays, proximity labeling assays, proximity litigation assays, and two-hybrid assays with 'high' detection confidence. A tandem affinity purification assay also detected protein-protein interactions between *CDH1* and *EGFR* with 'medium' detection confidence. *CDH1* and *EGFR* interact together in epithelial-to-mesenchymal transitions (EMT). *CDH1* is known to inhibit *EGFR* as a tumor suppressor, which is an inducer of EMT. Thus, partial downregulation of *CDH1* due to alterations may lead to an upregulation of *EGFR* and a partial induction of EMT.

Finally, *CTNNA1*, *CTNNB1*, and *CTNND1* were also identified as predicted functional partners to *CDH1* with combined scores of 0.999. Interaction with *CTNNA1* was detected by a pull-down assay with 'very high' confidence and was detected with 'high' confidence by proximity labeling assay, affinity chromatography assay, and anti-tag coimmunoprecipitation assay. Interaction with *CTNNB1* was detected with 'very high' confidence by affinity chromatography assay, pull-down assay, anti-bait coimmunoprecipitation assay, and x-ray crystallography assay. Finally, interaction with *CTNND1* was detected with 'very high' confidence by anti-tag coimmunoprecipitation assay, pull-down assay, proximity labeling assay, and affinity chromatography assay.

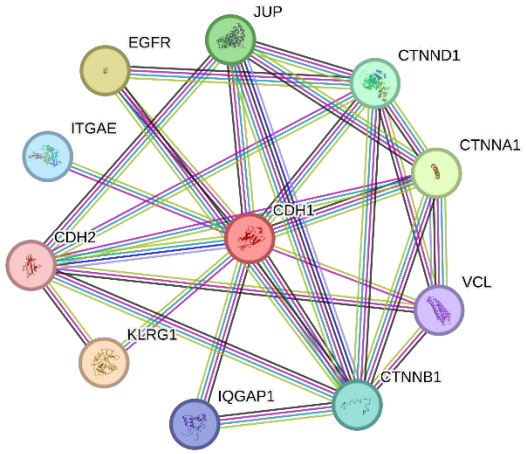

Figure 4: Protein-protein interaction network for CDH1 from STRING portal (https://string-db.org, accessed on 1 September 2022). EGFR, CTNNA1, CTNNB1, and CTNND1 are included as predicted functional partners with combined scores of 0.999 out of 1.000.

CDH1 appears to be relatively more highly expressed than *EGFR* in certain breast cancer subtypes

The DepMap Portal was used to compare the expression data of *CDH1* and *EGFR*, which were identified as predicted functional partners as previously mentioned (Figure 3).

For the disease subtype breast invasive ductal carcinoma, the average gene expression of *CDH1* on the log2(TPM+1) (log2-transformed Transcripts Per Million) scale was found to be 5.404 with a standard deviation of 2.527 (Figure 3a). The average gene expression of *EGFR* in the same disease subtype was found to be 2.932 with a standard deviation of 2.092 (Figure 3b). The 95% confidence intervals for the expression levels were 5.404 ± 0.92 for *CDH1* and 2.932 ± 0.761 for *EGFR*. Since there is no overlap between 95% confidence intervals, there is most likely a significant statistical difference in the expression of *CDH1* and *EGFR*. Based on known protein inhibition interactions between *CDH1* and *EGFR* in breast invasive ductal carcinoma, this observation is reasonable.

For the disease subtype breast invasive lobular carcinoma, the average gene expression of *CDH1* on the log2(TPM+1) scale was found to be 5.365 with a standard deviation of 3.239 (Figure 3a). The average gene expression of *EGFR* in the same disease subtype was found to be 2.805 with a standard deviation of 2.297 (Figure 3b). The 95% confidence intervals for the expression levels were 5.365 ± 2.24 for *CDH1* and 2.805 ± 1.59 for *EGFR*. Since there is an overlap between 95% confidence intervals, there is most likely not a significant statistical difference in the expression of *CDH1* and *EGFR*. For breast invasive lobular carcinoma, in which *CDH1* is most commonly altered, *CDH1*'s ability to inhibit *EGFR* appears to be repressed.

For the disease subtype invasive breast carcinoma (mixed), the average gene expression of *CDH1* on the log2(TPM+1) scale was found to be 5.092 with a standard deviation of 2.715 (Figure 3a). The average gene expression of *EGFR* in the same disease subtype was found to be 3.713 with a standard deviation of 2.578 (Figure 3b). The 95% confidence intervals for the expression levels were 5.092 ± 1.16 for *CDH1* and 3.713 ± 1.1 for *EGFR*. Since there is an overlap between 95% confidence intervals, there is most likely not a significant statistical difference in the expression of *CDH1* and *EGFR*. For breast invasive carcinoma (mixed), in which *CDH1* is second most commonly altered, *CDH1*'s ability to inhibit *EGFR* similarly appears to be repressed.

The disease subtypes breast ductal carcinoma *in situ,* breast invasive carcinoma NOS (Not Otherwise Specified), breast neoplasm NOS, and immortalized breast cell data points were excluded as they did not have a high enough sample size to run a reliable statistical analysis.

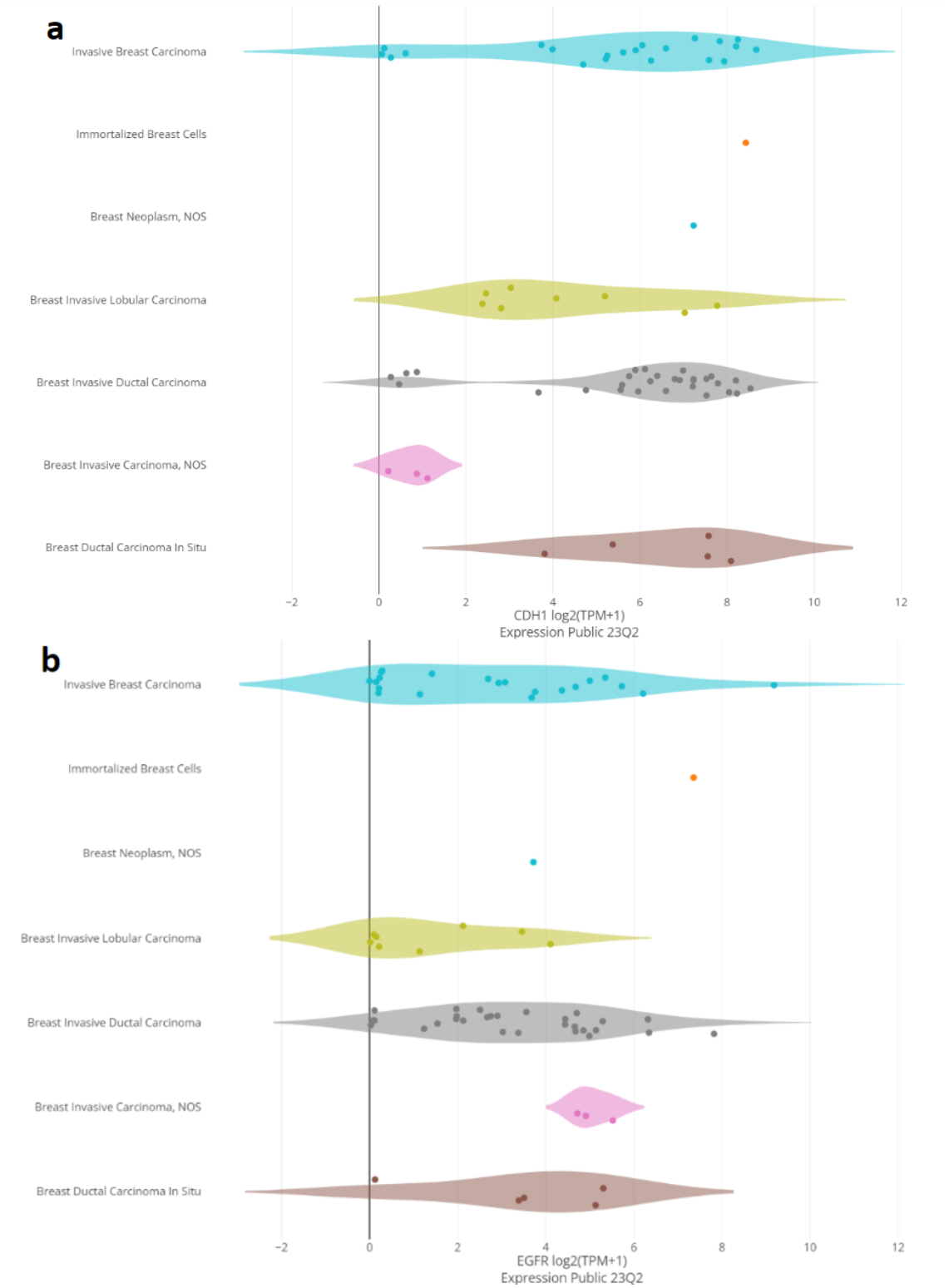

Figure 5: Plot comparing the expression data of CDH1 (a) and EGFR (b) on a log2(TPM+1) scale from DepMap Portal (https://depmap.org/portal, accessed on 1 September 2022). The difference in protein expression between CDH1 and EGFR was found to be statistically

significant using 95% CI calculated for breast invasive ductal carcinoma, with CDH1 having greater protein expression.

Protein expression of *CDH1* may be slightly elevated in late-stage luminal and HER2-positive breast cancer

The protein expression of *CDH1* in breast cancer vs. normal tissue was analyzed using the UALCAN tool.

When comparing the protein expression of *CDH1* between normal and primary tumor tissues, it was observed that on a z-score normalized distribution, the median of the primary tumors ($n = 125$) was 0.035, which was greater than the median of the normal tissue ($n = 18$), which was -0.659 (Figure 4a). The p-value was determined to be $1.199673 \cdot 10^{-2}$ using a student's t-test considering unequal variance. Thus, there appears to be a statistically significant difference in protein expression between normal and primary breast cancer tumor tissues.

The protein expression of *CDH1* was then compared between different breast cancer subclasses. On a z-score normalized distribution, the median of luminal breast cancer ($n = 64$) was highest, at 0.052, followed by HER2-positive breast cancer ($n = 10$), at 0.004, followed by triple-negative breast cancer ($n = 16$), at -0.196, finally followed by normal tissue ($n = 18$), at -0.659 (Figure 4b). Statistically significant differences between normal and altered tissues were observed for the luminal subtype (p-value = $2.258715 \cdot 10^{-2}$) and the HER2-positive subtype (p-value = $1.129510 \cdot 10^{-2}$).

Finally, the protein expression of *CDH1* was compared between different cancer stages. On a z-score normalized distribution, the median of stage 3 breast cancer ($n = 32$) was highest, at 0.185, followed by stage 1 breast cancer ($n = 4$), at 0.141, followed by stage 2 breast cancer ($n = 74$), at -0.001, finally followed by normal tissue ($n = 18$), at -0.659 (Figure 4c). Statistically significant differences between normal and altered tissues were observed for stage 3 cancer (p-value = $2.832929 \cdot 10^{-2}$) and stage 2 cancer (p-value = $3.567584 \cdot 10^{-2}$). Thus, there appears to be a positive correlation between *CDH1* protein expression and breast cancer stage.

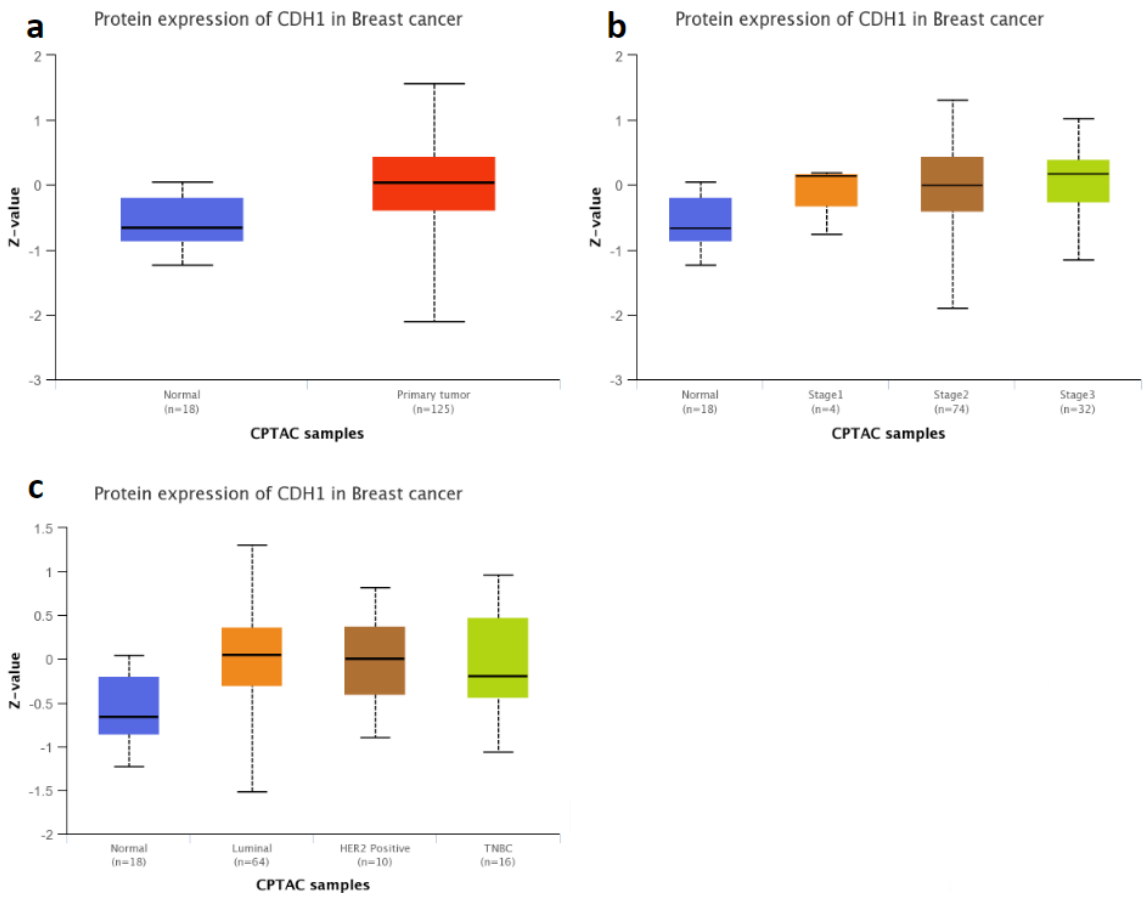

Figure 6: Boxplots of CDH1 protein expression on z-score normalized scales from UALCAN (http://ualcan.path.uab.edu/index.html, accessed on 1 September, 2023).

In (a), protein expression is compared between normal and primary tumor tissue. The difference was found to be statistically significant, with the expression being higher in primary tumor tissue.

In (b), protein expression is compared between cancer stages. The difference between normal tissue and stage 2 breast cancer and the difference between normal tissue and stage 3 breast cancer were found the be statistically significant, with the expression being higher in both stage 2 and stage 3 breast cancer than in normal tissue.

In (c), the expression is compared between breast cancer subtypes. The difference between normal tissue and luminal breast cancer and the difference between normal tissue and HER2-positive breast cancer were found to be statistically significant, with the expression being higher in both luminal and HER2-positive breast cancer than in normal tissue.

Discussion

CDH1 is a classical cadherin specialized in cell-cell adhesion. It is essential in epithelial cells as an intercellular adhesion molecule and in cancer, and its mutation in cancers has been known to be involved particularly in the epithelial-to-mesenchymal transition[15]. Under this process,

epithelial cells generally lose intercellular adhesion, gain cellular motility, and lose epithelial differentiation. Because loss or downregulation of E-cadherin results in these hallmarks of cancer, *CDH1* has often been regarded as a tumor suppressor gene.

In this study, *CDH1*'s role in breast cancer was analyzed using proteomic and genomic tools. From cBioPortal, *CDH1* was identified to be altered in just 13% of patients in the TCGA Firehose Legacy dataset. Truncating (nonsense) mutations were the most common *CDH1* alteration in these patients, with Q23*, an early nonsense mutation, being the most common. Along with other common nonsense mutations early in the *CDH1* gene, these truncations in *CDH1* as a tumor suppressor may account for some cases of carcinogenesis. These results were confirmed with the METABRIC dataset on cBioPortal.

It was then sought to determine the subtype in which *CDH1* is most relatively frequently altered. Again, using cBioPortal, it was observed that *CDH1* had the greatest percentage of alteration in breast invasive lobular carcinoma. Over half of patients with the disease subtype were identified to sustain one or more mutations in the *CDH1* gene. These results were confirmed with the METABRIC dataset on cBioPortal.

E-cadherin is also an important part of the WNT/β-catenin pathway. To analyze its interactions with other proteins in the pathway, as well as other possible proteins, STRING was used to map protein-protein interactions with E-cadherin. E-cadherin was identified as a predicted functional partner to N-cadherin (*CDH2*), an interaction important in the epithelial-to-mesenchymal process, where E-cadherin is downregulated, and N-cadherin is upregulated. E-cadherin was also identified as a functional partner to *CTNNA1*, a cadherin-connecting protein and a putative tumor suppressor gene, *CTNNB1*, a protein essential in the production of β-catenin and an oncogene, and *CTNND1* (p120-catenin), a protein crucial in maintaining and stabilizing E-cadherin and an oncogene.

Furthermore, E-cadherin was identified as a functional partner to *EGFR*, a receptor tyrosine kinase essential in the PI3K/Akt pathway but also with ties to the WNT/β-catenin pathway. E-cadherin is known to inhibit *EGFR*, a putative oncogene, and loss of E-cadherin may lead to increased *EGFR* signaling and cell proliferation.

To further examine the relationship between *CDH1* and *EGFR,* the DepMap portal was used to compare their gene expression levels in different breast cancer subtypes. Statistically significant differences between gene expression of *CDH1* and *EGFR* were observed in breast invasive ductal carcinoma. *CDH1* gene expression was significantly higher than *CDH1*. In breast invasive ductal carcinoma, *CDH1*'s expression was increased, perhaps to compensate for loss of function due to alteration. Since *CDH1* inhibits *EGFR*, increased *CDH1* levels justify the lower levels of *EGFR* expression.

Finally, protein expression of *CDH1* between normal vs. primary tumor, normal vs. different cancer stages, and normal vs. breast cancer subtypes was compared. Interestingly, protein expression was significantly higher in primary tumor tissue compared to normal tissue, higher in stage 2 and stage 3 breast cancer compared to normal tissue, and higher in luminal and HER2-positive breast cancer compared to normal tissue. One possibility for this phenomenon is that cells attempt to compensate for the loss of *CDH1* function by increasing its production (which is

ultimately unsuccessful as alterations will render produced E-cadherin dysfunctional). Another possibility is that alterations in DNA methylation patterns, such as the hypomethylation of the *CDH1* promoter region, may lead to overexpression.

While *CDH1* is often regarded as a tumor suppressor gene, its overexpression in later cancer stages, as shown by the observations, may potentially promote tumor progression, as has previously been investigated by Xie *et al.*[16].

CDH1 in breast cancer was analyzed through genomics and proteomics tools; however, limitations, especially based on sample size, remain. The study provides further evidence of the role of *CDH1* as a tumor suppressor gene in breast cancer and its interactions with other important biomarkers, such as *EGFR*. However, further validation is warranted, especially on the possible oncogenic role of *CDH1* in late-stage breast cancer.

Methods

Patient Datasets

The Cancer Genome Atlas (TCGA) Firehose Legacy dataset is an invasive breast cancer patient dataset of 1,108 carcinomas from 1,101 patients gathered by the Broad Institute of MIT and Harvard Firehose initiative[17]. It was used to download raw clinical information. The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) is a primary breast cancer dataset of 1,980 samples from patients of the METABRIC trial[18]. It was also used to download raw clinical information and tumor pathology. MSK (Memorial Sloan Kettering) datasets, developed by the Memorial Sloan Kettering Cancer Center, were also used for mutation and genomic information[19].

Gene and Mutation Analysis Using cBioPortal

cBioPortal for Cancer Genomics (http://www.cbioportal.org/, accessed on 12 July 2023) is an open-access, open-source tool for the visualization and analysis of genomic, mutation, and clinical data[20,21]. In this study, cBioPortal was used to analyze *CDH1* mutation frequency and type in breast cancer samples by querying the *CDH1* gene.

Protein-Protein Interaction Analysis Using STRING

STRING (https://string-db.org/cgi/input.pl, accessed on 7 August 2023) is a database of known and predicted protein-protein interactions determined by multiple assays[22–33]. In this study, STRING was used to analyze *CDH1* protein interactions and to construct a protein-protein interaction (PPI) network of *CDH1*.

Gene Expression Analysis Using DepMap

DepMap (https://depmap.org/portal/, accessed on 21 August 2023) is a cancer-dependency database of cancer-gene interactions and dependencies determined using CRISPR and RNAi[34,35]. In this study, DepMap was used to compare gene expression of *CDH1* and *EGFR* between cancer subtypes.

Protein Expression Analysis Using UALCAN

UALCAN (https://ualcan.path.uab.edu/index.html, accessed on 28 August 2023) is a comprehensive resource for analyzing proteomic cancer data[36,37]. In this study, UALCAN was used to compare protein expression of *CDH1* in normal tissue vs. primary tumor, normal tissue vs. breast cancer stages 1, 2, and 3, and normal tissue vs. luminal, HER2-positive, and triple-negative breast cancer.

Statistical Analysis

Statistical analysis was performed using R (v.4.3.1). Pearson or Spearman correlation tests were performed to test for correlations, as appropriate. A student's t-test considering unequal variance and a comparison of 95% confidence intervals was used to determine statistically significant differences. Differences with a $p < 0.05$ were considered statistically significant.

Acknowledgements

This study was guided by Dr. Begum Akman and Dr. Taner Tuncer.

References

1. P. Anand, A. B. Kunnumakara, C. Sundaram, K. B. Harikumar, S. T. Tharakan, O. S. Lai, B. Sung & B. B. Aggarwal. Cancer is a Preventable Disease that Requires Major Lifestyle Changes. *Pharm Res* **25**, 2097–2116 (2008).
2. D. Hanahan & R. A. Weinberg. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
3. A. N. Giaquinto, H. Sung, K. D. Miller, J. L. Kramer, L. A. Newman, A. Minihan, A. Jemal & R. L. Siegel. Breast Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians* **72**, 524–541 (2022).

4. G. Berx, A. M. Cleton-Jansen, F. Nollet, W. J. de Leeuw, M. van de Vijver, C. Cornelisse & F. van Roy. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J* **14**, 6107–6115 (1995).
5. E. M. Howard, S. K. Lau, R. H. Lyles, G. G. Birdsong, J. N. Umbreit & R. Kochhar. Expression of e-cadherin in high-risk breast cancer. *J Cancer Res Clin Oncol* **131**, 14–18 (2005).
6. A. M. Mendonsa, T.-Y. Na & B. M. Gumbiner. E-cadherin in contact inhibition and cancer. *Oncogene* **37**, 4769–4780 (2018).
7. R. Moll, M. Mitze, U. H. Frixen & W. Birchmeier. Differential loss of E-cadherin expression in infiltrating ductal and lobular breast carcinomas. *Am J Pathol* **143**, 1731–1742 (1993).
8. M Takeichi. Cadherins in cancer: implications for invasion and metastasis. *Current Opinion in Cell Biology* **5**, 806–811 (1993).
9. Y. An, J. R. Adams, D. P. Hollern, A. Zhao, S. G. Chang, M. S. Gams, P. E. D. Chung, X. He, R. Jangra, J. S. Shah, J. Yang, L. A. Beck, N. Raghuram, K. J. Kozma, A. J. Loch, W. Wang, C. Fan, S. J. Done, E. Zacksenhaus, C. J. Guidos, C. M. Perou & S. E. Egan. Cdh1 and Pik3ca Mutations Cooperate to Induce Immune-Related Invasive Lobular Carcinoma of the Breast. *Cell Reports* **25**, 702-714.e6 (2018).
10. M. Ramírez Moreno & N. A. Bulgakova. The Cross-Talk Between EGFR and E-Cadherin. *Front Cell Dev Biol* **9**, 828673 (2022).
11. M. Conacci-Sorrell, J. Zhurinsky & A. Ben-Ze'ev. The cadherin-catenin adhesion system in signaling and cancer. *J Clin Invest* **109**, 987–991 (2002).
12. C.-Y. Loh, J. Y. Chai, T. F. Tang, W. F. Wong, G. Sethi, M. K. Shanmugam, P. P. Chong & C. Y. Looi. The E-Cadherin and N-Cadherin Switch in Epithelial-to-Mesenchymal Transition: Signaling, Therapeutic Implications, and Challenges. *Cells* **8**, 1118 (2019).

13. P. Polakis. Wnt Signaling in Cancer. *Cold Spring Harb Perspect Biol* **4**, a008052 (2012).
14. M. V. Fournier, J. Fata, K. Martin, P. Yaswen & M. J. Bissell. Interaction of E-cadherin and PTEN regulates morphogenesis and growth arrest in human mammary epithelial cells. *Cancer Res* **69**, 4545–4552 (2009).
15. N. Pećina-Šlaus. Tumor suppressor gene E-cadherin and its role in normal and malignant cells. *Cancer Cell Int* **3**, 17 (2003).
16. D. Xie, Y. Chen, X. Wan, J. Li, Q. Pei, Y. Luo, J. Liu & T. Ye. The Potential Role of CDH1 as an Oncogene Combined With Related miRNAs and Their Diagnostic Value in Breast Cancer. *Front Endocrinol (Lausanne)* **13**, 916469 (2022).
17. K. Tomczak, P. Czerwińska & M. Wiznerowicz. Review
The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* **2015**, 68–77 (2015).
18. V. Leoncikas, H. Wu, L. T. Ward, A. M. Kierzek & N. J. Plant. Generation of 2,000 breast cancer metabolic landscapes reveals a poor prognosis group with active serotonin production. *Sci Rep* **6**, 19771 (2016).
19. D. T. Cheng, T. N. Mitchell, A. Zehir, R. H. Shah, R. Benayed, A. Syed, R. Chandramohan, Z. Y. Liu, H. H. Won, S. N. Scott, A. R. Brannon, C. O'Reilly, J. Sadowska, J. Casanova, A. Yannes, J. F. Hechtman, J. Yao, W. Song, D. S. Ross, A. Oultache, S. Dogan, L. Borsu, M. Hameed, K. Nafa, M. E. Arcila, M. Ladanyi & M. F. Berger. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *The Journal of Molecular Diagnostics* **17**, 251–264 (2015).
20. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander & N.

Schultz. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* **2**, 401–404 (2012).

21. J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander & N. Schultz. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci Signal* **6**, pl1 (2013).
22. A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering & L. J. Jensen. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**, D808–D815 (2013).
23. L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork & C. von Mering. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412–D416 (2009).
24. B. Snel, G. Lehmann, P. Bork & M. A. Huynen. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* **28**, 3442–3444 (2000).
25. D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, P. Bork, L. J. Jensen & C. von Mering. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638–D646 (2022).
26. D. Szklarczyk, A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen & C. von Mering. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605–D612 (2020).

27. D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen & C. von Mering. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607–D613 (2019).
28. D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen & C. von Mering. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res* **45**, D362–D368 (2017).
29. D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen & C. von Mering. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–D452 (2015).
30. D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen & C. von Mering. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* **39**, D561–D568 (2011).
31. C. von Mering, L. J. Jensen, M. Kuhn, S. Chaffron, T. Doerks, B. Krüger, B. Snel & P. Bork. STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35**, D358–D362 (2007).
32. C. von Mering, L. J. Jensen, B. Snel, S. D. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M. A. Huynen & P. Bork. STRING: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Res* **33**, D433–D437 (2005).
33. C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork & B. Snel. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* **31**, 258–261 (2003).

34. A. Tsherniak, F. Vazquez, P. G. Montgomery, B. A. Weir, G. Kryukov, G. S. Cowley, S. Gill, W. F. Harrington, S. Pantel, J. M. Krill-Burger, R. M. Meyers, L. Ali, A. Goodale, Y. Lee, G. Jiang, J. Hsiao, W. F. J. Gerath, S. Howell, E. Merkel, M. Ghandi, L. A. Garraway, D. E. Root, T. R. Golub, J. S. Boehm & W. C. Hahn. Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).

35. R. M. Meyers, J. G. Bryan, J. M. McFarland, B. A. Weir, A. E. Sizemore, H. Xu, N. V. Dharia, P. G. Montgomery, G. S. Cowley, S. Pantel, A. Goodale, Y. Lee, L. D. Ali, G. Jiang, R. Lubonja, W. F. Harrington, M. Strickland, T. Wu, D. C. Hawes, V. A. Zhivich, M. R. Wyatt, Z. Kalani, J. J. Chang, M. Okamoto, K. Stegmaier, T. R. Golub, J. S. Boehm, F. Vazquez, D. E. Root, W. C. Hahn & A. Tsherniak. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779–1784 (2017).

36. D. S. Chandrashekar, B. Bashel, S. A. H. Balasubramanya, C. J. Creighton, I. Ponce-Rodriguez, B. V. S. K. Chakravarthi & S. Varambally. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* **19**, 649–658 (2017).

37. D. S. Chandrashekar, S. K. Karthikeyan, P. K. Korla, H. Patel, A. R. Shovon, M. Athar, G. J. Netto, Z. S. Qin, S. Kumar, U. Manne, C. J. Crieghton & S. Varambally. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia* **25**, 18–27 (2022).