Kaiso (ZBTB33) subcellular partitioning functionally links LC3A/B, the tumor microenvironment, and breast cancer survival

Sandeep K. Singhal, Jung S. Byun, Samson Park, Tingfen Yan, Ryan Yancey, Ambar Caban, Sara Gil Hernandez, Stephen M. Hewitt, Heike Boisvert, Stephanie Hennek, Mark Bobrow, Md Shakir Uddin Ahmed, Jason White, Clayton Yates, Andrew Aukerman, Rami Vanguri, Rohan Bareja, Romina Lenci, Paula Lucia Farré, Adriana De Sierra, Anna María Nápoles, Nasreen Vohra & Kevin Gardner

The use of digital pathology for the histomorphologic profiling of pathological specimens is expanding the precision and specificity of quantitative tissue analysis at an unprecedented scale; thus, enabling the discovery of new and functionally relevant histological features of both predictive and prognostic significance. In this study, we apply quantitative automated image processing and computational methods to profile the subcellular distribution of the multi-functional transcriptional regulator, Kaiso (ZBTB33), in the tumors of a large racially diverse breast cancer cohort from a designated health disparities region in the United States. Multiplex multivariate analysis of the association of Kaiso’s subcellular distribution with other breast cancer biomarkers reveals novel functional and predictive linkages between Kaiso and the autophagy-related proteins, LC3A/B, that are associated with features of the tumor immune microenvironment, survival, and race. These findings identify effective modalities of Kaiso biomarker assessment and uncover unanticipated insights into Kaiso’s role in breast cancer progression.
last year in the United States there were over 260,000 new cases of invasive breast cancer and, by the year’s end, over 40,000 women diagnosed with breast cancer died from their disease. Breast cancer is a very heterogeneous disease with various subtypes of diagnostic and prognostic significance that differ in distribution by both age and race. A major tool in the diagnosis, management, and prevention of breast cancer is the identification and characterization of biomarkers that will predict survival, and guide treatment decisions with higher precision. Traditional visual assessment of histopathological images in combination with antibody-based biomarker profiling has been the standard of practice. However, the recent application of algorithms for digital image analysis to augment image characterization and quantitative assessment has broadened the utility, application, and accuracy of antibody-based characterization, thus providing new insight into the functional roles of specific protein biomarkers, and spanning a new age of diagnostic, prognostic, and therapeutic innovation. Here we apply multiple algorithms for digital image analysis to perform an antibody-based assessment of patient breast cancer tissues that quantitatively profiles the subcellular distribution of Kaiso (ZBTB33), a functional protein biomarker previously shown to be elevated in cancers of the breast, prostate, pancreas, and colon.

Kaiso was originally identified as a transcription factor and member of the BTB/POZ (Broad complex, Tramtrak, Bric a Brac/POXvirus zinc finger) subfamily of zinc finger proteins. It contains a bimodal DNA-binding domain that recognizes both sequence-specific consensus sites and methylated CpG nucleotides through 3 C-terminal DNA-binding Kruppel-like CH2 zinc fingers. In its nuclear capacity, it has been shown to drive transcriptional programs that increase breast cancer growth and metastasis. Subsequent studies revealed that it binds to a variety of sequence motifs throughout the nucleus, the majority of which are unmethylated and located at loci with open chromatin. Because of its role in transcriptional regulation, Kaiso has been predominantly investigated as a nuclear protein and multiple studies have shown that its nuclear accumulation is dynamically regulated by association with p120 (CTNND1), a catenin family member that regulates membrane-bound E-cadherin cell adhesion assemblies. Although several studies have noted that the cytoplasmic localization of Kaiso is a common feature in human tissues and tumors, the major emphasis placed on Kaiso, as a prognostic biomarker, has focused on its nuclear accumulation. In fact, the nuclear accumulation of Kaiso has been shown, in multiple studies, to be a predominant feature of more aggressive forms of breast cancer, including triple-negative breast cancer. Moreover, in some studies, nuclear Kaiso was shown to be more predictive of poor survival in women of African heritage diagnosed with TNBC. Nonetheless, there have been multiple reported observations of cytoplasmic or “non-nuclear” accumulation of Kaiso, implying potential functional roles for Kaiso outside of the nucleus. These include interaction with centrosomes, assembly with RhoH and p120 at actin-containing cell protrusions, and regulation of Kaiso subcellular distribution by the EGF-19,18,19.

In this study, we leverage the analytic precision of automated image analysis algorithms to quantitatively profile the compartment-specific (nuclear versus cytoplasmic) distribution of Kaiso in a racially diverse cohort of breast cancer patients residing in a health disparities region of rural East North Carolina. These findings reveal that both nuclear and cytoplasmic Kaiso are associated with breast cancer outcome and each are independent predictors of overall breast cancer survival. Furthermore, compartment-specific profiling of Kaiso with multiple prognostic breast cancer biomarkers reveal new functional correlations that link the specific subcellular distribution of Kaiso with (1) the autophagy-related factor LC3A/B; (2) cellular phenotypes within the tumor immune microenvironment; and (3) overall breast cancer survival.

Results

The subcellular distribution of Kaiso is differentially correlated with breast cancer subtype and overall survival from breast cancer. The protein expression of Kaiso was spatially profiled by automated analysis of immunohistochemically stained tissue microarrays (TMAs) containing 555 tumors from a cohort of racially diverse breast cancer patients residing in a designated health disparities catchment area of East North Carolina (median follow-up 8.5 years). By this analysis, staining intensity in the nucleus or cytoplasm of each cell in the annotated tumor regions is assigned one of four scores: from negative staining (0) to weak staining (+1), moderate staining (2+) or strong staining (3+). The percent of cells in the annotated regions demonstrating one of the 4 different intensities is then aggregated to derive an H-score (H-score = 3 × [%3+] + 2 × [%2+] + 1 × [%1+] + 0 × [%0+]), thus generating a continuous score from 0 to 300. The subcellular distribution of the staining intensity differs across many of the tumors, revealing various patterns of enrichment in the nucleus, cytoplasm, or both.

Notably, the cell segmentation algorithm-based quantitative profiling of Kaiso subcellular distribution reveals distinct differences in the cytoplasmic versus nuclear Kaiso patterns of distribution when examined in rank order (Fig. 1b, left). This Kaiso-specific difference in subcellular distributions is contrasted by the significant similarity in the nuclear and cytoplasmic distribution of a typical nuclear antigen that shuttles between the cytoplasm and nucleus like the androgen receptor (AR) (Fig. 1b, right), as well as the lower correlation between nuclear and cytoplasmic Kaiso, compared to AR (Supplementary Figure 2). These stark differences suggest that the mechanisms governing the nuclear versus cytoplasmic localization of Kaiso are far more dynamic and complex than the androgen receptor whose intensity distribution in the nucleus reflects a more passive ligand-mediated nuclear-to-cytoplasmic distribution.

Unlike nuclear Kaiso, the levels of cytoplasmic Kaiso in this cohort were significantly different based on hormone receptor (ER) status and breast cancer subtype, where the cytoplasmic levels of Kaiso were distinctly higher in the subtypes of breast cancer known to be more aggressive, including triple-negative breast cancer (TNBC), human epidermal growth factor receptor 2 positive (HER2+), and Luminal B (LumB). Although the levels of nuclear Kaiso have been previously reported to be higher in patients of African heritage, we did not detect such differences in the current cohort although a minor trend for preferential distribution of high levels of cytoplasmic Kaiso in patients of African, compared to European descent, was observed. Interestingly, although the mRNA levels of Kaiso are highly predictive of poor breast cancer survival, as demonstrated in publicly available data sets (Fig. 1e); a direct comparison of available mRNA and either nuclear Kaiso, cytoplasmic Kaiso, or their combined total (Fig. 1f).

Nuclear Kaiso, cytoplasmic Kaiso, and their combined score, defined as total Kaiso, are highly correlated with poor breast cancer survival (Fig. 2a–c). However, cytoplasmic Kaiso appears to be significantly more predictive with a hazard ratio (HR) of 16.29 (confidence interval (CI): 7.6–34.8; p-value 5.3E–13) for cytoplasmic Kaiso compared to HR: 2.83 (CI: 2.02–3.9; p-value 6.1E–11) for nuclear Kaiso, and HR: 7.86 (CI: 5.0–12.22; p-value 1.7E–18) for total Kaiso (Fig. 2a). These differences are also in...
good agreement with a non-digitally assisted expert pathologist’s assessment (Supplementary Fig. 4). In the multivariate setting, both nuclear and cytoplasmic Kaiso are independent predictors of overall breast cancer survival (Table 1). Finally, both nuclear and cytoplasmic Kaiso can stratify patients with lymph node-positive (high risk of recurrence) or lymph node-negative breast cancer (low risk of recurrence) (Fig. 2b). However, in each case, although both nuclear Kaiso and cytoplasmic Kaiso are independent predictors of survival, cytoplasmic Kaiso is consistently more predictive of poor overall survival compared to nuclear Kaiso (Fig. 2a–c). Interestingly, total Kaiso appears to be the best predictor of overall breast cancer survival as demonstrated in two independent publicly available breast cancer cohorts. 88.  

Cytoplasmic Kaiso reveals a high correlation with TNBC and the autophagy-related protein LC3A/B. The high correlation between cytoplasmic Kaiso expression and more aggressive subtypes of breast cancer (Fig. 1c) implies that similar comparisons, between cytoplasmic Kaiso and other biomarkers that stratify aggressive forms of cancer, will provide deeper insights into the role of cytoplasmic Kaiso in poor breast cancer survival. To accomplish this goal, quantitative profiles comparing the relative enrichment of biomarkers recently implicated to be strongly associated with breast cancer progression 21–29, were analyzed together by robust unsupervised hierarchical clustering (Fig. 3a). Biomarkers investigated included, estrogen-receptor ER 30, the ER pioneer proteins FOXA1 31–33 and GATA3 34–36, HER2 membrane expression 37,38, the protein-membrane adhesion molecule and tumor suppressor E-cadherin 39–41, epithelial growth factor receptor EGFR 37,38, and the autophagy-related factor LC3A/B (MAP1L3A and MAP1L3B), an autophagy-related biomarker recently implicated to be strongly associated with breast cancer progression 21–29. A heatmap of clinical and pathological features of patients including survival, tumor subtype, and ER status is provided underneath for a direct comparison of biomarker expression with patient characteristics. Notably, patient clustering by these biomarkers in combination with cytoplasmic Kaiso expression identifies multiple breast cancer groups (A1–C2) with distinct survival differences (Fig. 3b and Supplementary Fig. 5). Of particular interest is the clustering of cytoplasmic Kaiso with the autophagy marker LC3A/B (Fig. 3a), which, in combination with the other biomarkers, stratifies TNBC patients into 3 different survival subgroups (B1, B2, and B3). The LC3A/B family of proteins has a major role in a variety of autophagy-related phagocytic and secretory processes including autophagy, phagocytosis, conventional secretion of cytokines, extracellular...
Fig. 2 Both nuclear Kaiso and cytoplasmic Kaiso are predictive of poor breast cancer survival. a Analysis of the association between subcellular Kaiso distribution and survival demonstrates that high cytoplasmic Kaiso is much more predictive of poor survival compared to nuclear Kaiso. Nuclear, cytoplasmic, and total Kaiso H-scores predict survival in both high risk (lymph node-positive) and low risk (lymph node-negative) breast cancer patients, where total Kaiso score is most predictive of survival in both low and high-risk breast cancer patients. (HR could not be calculated for cytoplasmic Kaiso in low-risk patients because no deaths were recorded in that risk group). NHW non-Hispanic white; NHB non-Hispanic black. Optimized cut-offs were determined by the method of maximally selected rank statistics (Supplementary Fig. 3 and Supplementary Table 2).

Table 1 Univariate and multivariate analysis of the hazard ratio for overall survival associated with patient demographics, subtype, nuclear Kaiso, cytoplasmic Kaiso, and cytoplasmic LC3A/B expression in patient breast cancer samples.

| Univariate analysis | HR (95% CI for HR) | p-value | Multivariate analysis | HR (95% CI for HR) | p-value |
|---------------------|--------------------|---------|----------------------|--------------------|---------|
| AGE                 | 1.02 (1.01-1.03)  |         | AGE                  | 1.02 (1.00-1.05)  | 5.28E-02|
| RACE (Black)        | 1.18 (0.89-1.56)  | 0.91    | RACE (Black)         | 0.85 (0.58-1.26)  | 4.25E-01|
| Menopause status    | 0.76 (0.55-1.12)  | 0.34    | Menopause status     | 0.89 (0.47-1.70)  | 7.23E-01|
| BMI                 | 0.10 (0.98-1.00)  |         | BMI                  | 0.98 (0.96-1.00)  | 1.00E-01|
| ER status           | 1.76 (1.32-2.35)  | 1.30E-04| ER status            | 1.01 (0.49-2.10)  | 8.99E-01|
| Nuclear Kaiso       | 1.011 (1.01-1.10) | 2.00E-09| Nuclear Kaiso        | 1.01 (1.01-1.01)  | 2.98E-07|
| Cytoplasmic Kaiso   | 1.011 (1.02-1.10) | 3.96E-10| Cytoplasmic Kaiso    | 1.01 (1.01-1.01)  | 2.98E-07|
| LC3AB               | 1.00 (1.00-1.01)  | 5.90E-05| LC3AB                | 1.00 (1.00-1.00)  | 6.06E-01|
| Node (positive)     | 2.30 (1.69-3.13)  | 1.30E-07| Node (positive)      | 2.64 (1.79-3.89)  | 8.78E-07|
| Subtype (LumA)      |                     |         | Subtype (LumA)       |                     |         |
| LumB                | 1.29 (0.87-1.93)  | 2.06E-01| LumB                 | 1.31 (0.77-2.21)  | 3.19E-01|
| HER2+               | 1.54 (0.91-2.62)  | 1.10E-01| HER2+                | 1.83 (0.76-4.43)  | 1.81E-01|
| TNBC                | 2.50 (1.77-3.53)  | 2.25E-07| TNBC                 | 2.25 (1.10-5.08)  | 5.20E-02|

Referent for Subtype is Luminal A subtype. Referent for menopause status is pre-menopause. Referent for Race is White. Referent for Node status is positive. Significant p-values (<0.05) are in bold.
release of lysozymes, extracellular vesicle (EV) production, unconventional protein secretion, and LC3-dependent EV loading and secretion (LDELS). In each case, lipid conjugated LC3 has a major role in the loading of different cargoes (organelles, protein, and nucleic acid) into membrane-bound structures destined for secretion or degradation. Collectively these processes are referred to as secretory autophagy. Furthermore, each of these processes has the potential to influence tumor microenvironment.

Remarkably, the clustering of low cytoplasmic Kaiso and low LC3A/B, in the context of the other biomarkers, identifies a class of TNBC patients that show favorable survival (compare B1 & B2, to B3) (Fig. 3b). Thus, low levels of cytoplasmic Kaiso combined with low levels of the autophagy-related factor LC3A/B predict favorable survival in breast cancer patients with TNBC, implicating a significant role for this biomarker in tumor progression and survival. Representative LC3A/B IHC staining of patient tumors reveals a heterogeneous staining pattern in the
cytoplasm with a mixture of diffuse and punctate cytoplasmic staining typical of LC3A/B\(^+\) (Fig. 3c). Quantitative analysis of the cytoplasmic staining shows that cytoplasmic LC3A/B is predictive of survival (HR: 2.5, CI: 1.77–3.68; \(p\)-value 5.9e–07) (Fig. 3d and Supplementary Fig. 3). Similar to cytoplasmic Kaiso, LC3A/B is (1) preferentially expressed in more aggressive forms of breast cancer including LumB, HER2, and TNBC (Fig. 3e); and (2) preferentially elevated in ER+ compared to ER− tumors with a trend toward higher expression in women of African compared to European ancestry (Fig. 3f). Analysis of the correlation between IHC-based protein expression and RNA expression (Fig. 3g and Supplementary Data 1) reveals very little correlation between MAP1LC3B mRNA and Kaiso protein. However, there is a modest correlation between MAP1LC3B mRNA and LC3A/B with the highest correlation between ZBTB33, MAP1LC3B mRNA, and LC3A/B protein (Fig. 3g). By univariate survival analysis, LC3A/B expression is significantly predictive of poor survival but loses all significance in the multivariate setting, where only nuclear Kaiso, cytoplasmic Kaiso, and lymph node status are significant independent predictors of outcome (Table 1). This suggests that the survival predictive value of LC3A/B is closely associated with Kaiso protein expression.

Kaiso is required for the functional activation of LC3A/B.

Given the close correlation between Kaiso and the LC3A/B autophagy-related proteins, we sought to further establish this relationship by employing a “bedside-to-bench” approach utilizing the human triple-negative breast cancer cell line, MDA-MB-231 depleted of Kaiso by RNA interference (RNAi) (Fig. 4a and Supplementary Fig. 6). The analysis of genes that are significantly differentially expressed (\(p\)-values < 0.001) in WT versus Kaiso-depleted cells reveals a large overlap (142 of 518) with a list of autophagy-associated genes compiled from the MSigDbase and the Autophagy Database\(^{1,2}\) (Fig. 4a) (also see Supplementary Data 2). This overlap is supported further by GSEA revealing a large and substantial enrichment in six (6) different autophagy-related gene sets (Fig. 4b) (also see Supplementary Data 3).

A hallmark feature of autophagy-related processes is the conjugation of LC3A/B to phosphatidylethanolamine (PE) followed by its incorporation into intracellular membrane vesicles within the endocytic vesicle system\(^{14,46,51}\). The lipid conjugation and membrane incorporation of LC3 protein enable it to load specific cargo into various endocytic trafficking vesicles including autophagosomes and extracellular vesicles\(^{44–46}\). A commonly used method to examine autophagocytic flux is to visualize the incorporation of GFP-labeled LC3A/B into autophagocytic membrane structures that appear as intracytoplasmic puncta (Fig. 4c)\(^{52,53}\). As demonstrated in Fig. 4c, MDA-MB-231 cells depleted of Kaiso show a significant absence (\(p<0.0001\)) of puncta both in the presence or absence of autophagosome stabilization by the lysosome inhibitor chloroquine (CQ) (Fig. 4d). LC3A/B maturation can be followed by immunoblot analysis through the detection of changes in LC3A/B phosphatidylethanolamine (PE) lipid conjugation by the demonstration of a change in mobility, where the lipid conjugated LC3A/B (LC3ii) migrates with faster mobility than the unconjugated form (LC3i) (Fig. 4e). This difference in LC3 conjugation can be amplified by the inhibition of LC3 lysosomal degradation by chloroquine (CQ) (Fig. 4e, f). As shown in Fig. 4f, compared to the non-targeting short hairpin, the addition of 3 different RNAi short hairpins targeting Kaiso, results in a significant reduction of GFP-LC3 conjugation. Similar results are observed with the ER+ cell line MCF-7 (Supplementary Fig. 7). Finally, Kaiso and LC3A/B immunofluorescence show significant colocalization in both ER+ (MCF-7) and TNBC (MDA-MB-231) cell lines in both the cytoplasm and nucleus (Fig. 4g, left and right, respectively), with quantitative profiling in Fig. 4h (also see Supplementary Fig. 8).

Kaiso and LC3A/B show significant colocalization in patient tumors.

Consistent with the observations in breast cancer cell lines, there is also significant cytoplasmic colocalization of Kaiso and LC3A/B in patient tumors with substantial variation by subtype (Fig. 5a, b) Thus, consistent with the co-enrichment of cytoplasmic Kaiso with LC3A/B in patient samples (Fig. 3a) and their colocalization in breast cancer cell lines (Fig. 4g), both proteins also show a substantial colocalization in patient tumors that varies by subtype (Fig. 5a, b).

Patients stratified by nuclear and cytoplasmic Kaiso are variably enriched in cellular stress and immune response pathways and differentially predict the overall outcome based on genetic ancestry. A comparison of the gene expression enrichment patterns of patients stratified by nuclear versus cytoplasmic Kaiso shows concordant (red) and discordant (blue) expression of genes linked to cellular differentiation, metabolism, immune modulation, and cell-microenvironment interactions (Fig. 6a, b). This observation is corroborated by GSEA profiles revealing opposing enrichment for inflammatory response pathways in nuclear versus cytoplasmic Kaiso (Fig. 6a, b, right) with notable negative enrichment for allograft rejection pathways in patients over-expressing cytoplasmic Kaiso (also see Supplementary Data 4 and 5).

As discussed earlier, nuclear Kaiso expression has been linked previously to racial differences in breast cancer outcome where nuclear Kaiso was found to be higher and more predictive of poor outcome in women of African heritage (non-Hispanic black (NHB)) diagnosed with TNBC compared to their European (non-Hispanic white (NHW)) counterparts\(^{8,14,15}\). Most notably this distinction seemed to be greater depending on the degree of African ancestry\(^{15}\). To profile the degree of racial admixture in our study cohort, ancestral informative markers (AIMs) were extracted from patient (\(N=131\)) tumor RNA-seq data (23% of the study cohort)\(^{54,55}\). Each patient in this group was then assigned a percent ancestry based on five genetic populations (African, European, East Asian, South Asian, and Admixed Native American) Fig. (6c). Among the 69 patients that self-identified as NHB, all but 1 had >50% African ancestry. Of the 62 patients that self-identified as NHW, 3 patients showed greater than 80% admixed Native American ancestry (Fig. 6c). Notably, forest plot analysis of the Cox proportional-hazards model for overall survival, optimized by Race, reveals that cytoplasmic Kaiso is more predictive of survival in women of African ancestry (Fig. 6d), consistent with prior indications of a differential survival risk based on Kaiso and African genetic background\(^{8,14}\). Moreover, even within admixed populations, and consistent with previous reports, there is a greater survival risk associated with nuclear Kaiso in patients of African ancestry diagnosed with TNBC (Fig. 6g). However, neither LC3A/B nor Nuclear Kaiso shows significant racial differences in survival hazard in the total breast cancer cohort (Fig. 6e, f).

Cytoplasmic Kaiso levels and LC3A/B are associated with an immune-suppressed tumor microenvironment in breast cancer tissues. The immune tumor microenvironment has been found to have a broad prognostic and predictive role in breast cancer\(^{56–61}\). With respect to breast cancer racial health disparities, there is a wide consensus supporting a deterministic role for race-based genetic variation in the immune response in influencing racial survival disparities in breast cancer\(^{62–67}\). Given the observed influence of elevated levels of cytoplasmic and nuclear Kaiso on immune regulatory pathways (Fig. 6a, b), and the extensive
potential role for secretory autophagy in the immune response, we sought to define the linkage between LC3A/B, the subcellular distribution of Kaiso, and immune properties of the tumor microenvironment (Fig. 7). This analysis was conducted through a nearest-neighbor analysis of TMA tissues containing for pan-cytokeratin (tumor), CD8 (killer T-cells), CD68 (macrophages), and PD-L1 (immune checkpoint regulator) (Fig. 7a, b). The x and y-coordinates (Fig. 7b) of the resultant combinations of the tumor and immune phenotypes (Supplementary Figs. 9 and 10) were then mapped and the frequency distribution of distances between each cellular phenotype was profiled (Fig. 7c). Notably, nearest-neighbor profiling showed significant coordination between cytoplasmic Kaiso and LC3A/B, and the proximity of PD-L1-positive CD8 cells near tumor compared to significant association with nuclear Kaiso and Race (Fig. 7c). Similarly, elevated levels of both cytoplasmic Kaiso and LC3A/B were associated with increased proximity of PD-L1-positive CD68 cells near tumor compared to insignificant association with nuclear Kaiso and Race (Fig. 7c). In a similar fashion nearest-neighbor profiling of total CD8 cells in proximity to PD-L1-positive tumor showed a significant association between elevated cytoplasmic Kaiso, LC3A/B and Race compared to the insignificant association with nuclear Kaiso (Fig. 7c). Finally, proximity profiling of total CD68 cells near PD-L1-positive tumor cells similarly shows a significant association with elevated cytoplasmic Kaiso and LC3A/B compared to nuclear Kaiso and Race (Fig. 7c). Given the known immunosuppressive role of PD-L1 expression in both immune cells and tumor, these findings reveal a strong association between LC3A/B and cytoplasmic Kaiso expression and the potential for an immunosuppressive tumor microenvironment. Notably, all associations between Kaiso and LC3A/B are specific to PDL1-positive cells as the trends described above are not significant when comparing total CD8 nor CD68 cells (Supplementary Fig. 11).
Fig. 5 Kaiso and LC3A/B show extensive colocalization in both the nucleus and cytoplasm of patient tumors. a Immunofluorescent staining of Kaiso (green) and LC3A/B (red) in TNBC breast cancer (upper panel) and Luminal B breast cancer (lower panel). b Colocalization maps showing relative colocalization of Kaiso and LC3A/B in patient tumors.

Discussion
In this work, we utilized a novel application of computational digital analysis and a unique cohort of breast cancer patients to define new relationships between biomarkers based on their differential subcellular distribution to define strong prognostic markers of breast cancer survival. This “bedside-to-bench” approach not only reveals biomarkers correlated with overall breast cancer survival and an immune-suppressed tumor microenvironment but provides new functional and mechanistic insights into the cellular processes linked to the biomarkers. The functional linkage of cytoplasmic Kaiso to LC3A/B and the tumor microenvironment uncovers a new area of investigation into the role and mechanism of Kaiso in breast cancer progression. Kaiso could have multiple roles in promoting tumor progression through both modulations of transcription and autophagy-related events including (1) autophagy-mediated degradation pathways; and (2) secretory autophagy pathways dependent on LC3 conjugation. The precise mechanism of how Kaiso influences LC3 maturation will require further investigation. One possibility is that it may act as a scaffold to facilitate enzymatic lipid conjugation of LC3 proteins in the cytoplasm. Such a possibility is supported by the computational identification of a predicted LC3 interaction domain (LIR) in the C-terminal region of Kaiso adjacent to the DNA-binding zinc fingers. The possible dual role for Kaiso in autophagy and LC3-mediated secretion are not necessarily mutually exclusive since recent studies indicate that LCDELs can occur independently of autophagy based on the observation that gene deletions that impair autophagy initiation do not block LCDELs. Further observations that will require further investigation arise from the morphological studies of cell lines and tumors. Notably, immunofluorescent staining for Kaiso reveals staining on tubulin (Fig. 4g and Supplementary Fig. 8) consistent with prior reports of the association of Kaiso with centrosomes. Other observations include the accumulation of Kaiso at focal adhesion-like structures or assemblies reminiscent of actin-containing cell protrusion sites in MCF-7 (Supplementary Fig. 8). Such associations have not been previously described, but suggest a linkage between autophagy-related processes and the cycling of cell-matrix adhesion machinery. Another interesting finding is the detection of race-specific differences in the tumor environment where there is a trend to a more suppressed immune microenvironment in women of African ancestry, particularly when considering CD8 cells near PD-L1-positive tumor (Fig. 7c). The differential contribution of race in the association of nuclear and cytoplasmic Kaiso with overall breast cancer survival (Fig. 6d, g) may contribute in part to this trend.

The general findings presented in this study, including the discovery of the prognostic significance of Kaiso subcellular partitioning and its linkage to immune-suppressive features of the tumor microenvironment, highlight its potential as a predictive biomarker to guide future treatment decisions, particularly in the use of immune checkpoint inhibitors. These results provide support for future applications in prospective studies where profiles of nuclear and cytoplasmic Kaiso are evaluated in clinical trials as both a predictive and prognostic breast cancer biomarker. Furthermore, because the predictive value of nuclear and cytoplasmic Kaiso varies across racial groups, these findings further emphasize the need for the inclusion of diverse racial and ethnic groups in clinical trials.

Methods
Study population, tissue microarray construction, and analysis. Following IRB approval from East Carolina University and the National Institutes of Health intramural research program, de-identified formalin-fixed and paraffin-embedded (FFPE) tissue samples and de-identified clinical information abstracted from medical records were requisitioned and initially procured for breast cancer patients who underwent surgery for Stage 0 to Stage IV breast cancer between 2001 and 2010 at Pitt County Memorial Hospital (now Vidant Medical Center), Greenville, NC. Race, ethnicity, or “ancestry” was self-reported at the initial visit and captured in the medical record. Survival was recorded retrospectively from the medical records and the cancer registry. All patient samples and data obtained were de-identified and approved by the East Carolina University Institutional Review Board as a human subject exempt project, for which no informed consent is needed. The study was conducted in accordance with the Declaration of Helsinki. Race and/or ethnicity was self-reported at the initial visit and captured in the medical record. Survival was recorded retrospectively from the medical records and the cancer registry. The median follow-up is 8.5 years. 588 patient tumor blocks from this cohort were found suitable for use in the construction of a tissue

In this work, we utilized a novel application of computational digital analysis and a unique cohort of breast cancer patients to define new relationships between biomarkers based on their differential subcellular distribution to define strong prognostic markers of breast cancer survival. This “bedside-to-bench” approach not only reveals biomarkers correlated with overall breast cancer survival and an immune-suppressed tumor microenvironment but provides new functional and mechanistic insights into the cellular processes linked to the biomarkers. The functional linkage of cytoplasmic Kaiso to LC3A/B and the tumor microenvironment uncovers a new area of investigation into the role and mechanism of Kaiso in breast cancer progression. Kaiso could have multiple roles in promoting tumor progression through both modulations of transcription and autophagy-related events including (1) autophagy-mediated degradation pathways; and (2) secretory autophagy pathways dependent on LC3 conjugation. The precise mechanism of how Kaiso influences LC3 maturation will require further investigation. One possibility is that it may act as a scaffold to facilitate enzymatic lipid conjugation of LC3 proteins in the cytoplasm. Such a possibility is supported by the computational identification of a predicted LC3 interaction domain (LIR) in the C-terminal region of Kaiso adjacent to the DNA-binding zinc fingers. The possible dual role for Kaiso in autophagy and LC3-mediated secretion are not necessarily mutually exclusive since recent studies indicate that LCDELs can occur independently of autophagy based on the observation that gene deletions that impair autophagy initiation do not block LCDELs. Further observations that will require further investigation arise from the morphological studies of cell lines and tumors. Notably, immunofluorescent staining for Kaiso reveals staining on tubulin (Fig. 4g and Supplementary Fig. 8) consistent with prior reports of the association of Kaiso with centrosomes. Other observations include the accumulation of Kaiso at focal adhesion-like structures or assemblies reminiscent of actin-containing cell protrusion sites in MCF-7 (Supplementary Fig. 8). Such associations have not been previously described, but suggest a linkage between autophagy-related processes and the cycling of cell-matrix adhesion machinery. Another interesting finding is the detection of race-specific differences in the tumor environment where there is a trend to a more suppressed immune microenvironment in women of African ancestry, particularly when considering CD8 cells near PD-L1-positive tumor (Fig. 7c). The differential contribution of race in the association of nuclear and cytoplasmic Kaiso with overall breast cancer survival (Fig. 6d, g) may contribute in part to this trend.

The general findings presented in this study, including the discovery of the prognostic significance of Kaiso subcellular partitioning and its linkage to immune-suppressive features of the tumor microenvironment, highlight its potential as a predictive biomarker to guide future treatment decisions, particularly in the use of immune checkpoint inhibitors. These results provide support for future applications in prospective studies where profiles of nuclear and cytoplasmic Kaiso are evaluated in clinical trials as both a predictive and prognostic breast cancer biomarker. Furthermore, because the predictive value of nuclear and cytoplasmic Kaiso varies across racial groups, these findings further emphasize the need for the inclusion of diverse racial and ethnic groups in clinical trials.

Methods
Study population, tissue microarray construction, and analysis. Following IRB approval from East Carolina University and the National Institutes of Health intramural research program, de-identified formalin-fixed and paraffin-embedded (FFPE) tissue samples and de-identified clinical information abstracted from the medical records were requisitioned and initially procured for breast cancer patients who underwent surgery for Stage 0 to Stage IV breast cancer between 2001 and 2010 at Pitt County Memorial Hospital (now Vidant Medical Center), Greenville, NC. Race, ethnicity, or “ancestry” was self-reported at the initial visit and captured in the medical record. Survival was recorded retrospectively from the medical records and the cancer registry. All patient samples and data obtained were de-identified and approved by the East Carolina University Institutional Review Board as a human subject exempt project, for which no informed consent is needed. The study was conducted in accordance with the Declaration of Helsinki. Race and/or ethnicity was self-reported at the initial visit and captured in the medical record. Survival was recorded retrospectively from the medical records and the cancer registry. The median follow-up is 8.5 years. 588 patient tumor blocks from this cohort were found suitable for use in the construction of a tissue
Fig. 6 Patients stratified by nuclear and cytoplasmic Kaiso are variably enriched in cell stress and immune response pathways and differentially predict survival based on genetic ancestry. a Volcano plot of differential gene expression of patients stratified by nuclear Kaiso. (Right) Gene set enrichment analysis of patients stratified by cytoplasmic Kaiso (Supplementary Data 5 and 6). b Volcano plot of differential gene expression of patients stratified by cytoplasmic Kaiso (right) Gene set enrichment analysis of patients stratified by nuclear Kaiso. c Population-specific composition of a representative portion (N = 131) of the breast cancer cohort, by genetic ancestry-based ancestry informative markers extracted from the RNA-seq data. Indicated colors reflect the percent admixture of each genetic population. d Forest plot analysis of overall survival hazard-based optimized cut-off for cytoplasmic Kaiso for the total population, NHW patients, NHB patients, and median cut-off for the total population. e Forest plot analysis of overall survival hazard-based optimized cut-off for LC3A/B in the total population, NHW patients, NHB patients, and the median cut-off for the total population. f Forest plot analysis of overall survival hazard-based optimized cut-off for LC3A/B in the total population, NHW patients, NHB patients, and the median cut-off for the total population. g Forest plot analysis of overall survival hazard-based optimized cut-off for nuclear Kaiso in TNBC patients using optimized cut-off for NHW patients, NHB patients, and the median cut-off for the total TNBC population.

Microarray. Replicate tissue microarrays were constructed using 1 mm cores per previously described methods73,74, with a representation inclusive of 555 patients. Detailed methods for IHC, scoring, and the assignment of clinical variables are provided in the Supplementary Methods.

Patient tumor RNA-seq analysis. RNA-seq analysis was performed on RNA extracted from FFPE tissue blocks (Total N = 126; EA N = 61; AA N = 65, Other = 13). Following a review of H&E stained slides, areas for tumors with >80% nuclei were circled, and 2.5 x 2–3 mm tissue cores were extracted from the corresponding regions of FFPE tissue blocks. Cores were shipped to the Beijing Genomics Institute (BGI) (Beijing, China) where RNA was extracted and sequenced (60M paired-end reads per sample) as previously described75,76. Detailed methods are provided in the data supplement.

Immunofluorescence. Cells were grown on 22-mm glass coverslips in 6-well plates to ~80% confluence before exposure to appropriate drugs or vehicle controls. After washing with PBS, cells were fixed with 4% formaldehyde in PBS at room temperature (RT) for 10 min. Cells were permeabilized with 0.1% Triton X-100 for 10 min, then incubated with blocking medium (PBS, 0.1% Tween-20, 10% normal goat serum) for 30 min in a humidified chamber at RT. Blocked samples
were then stained with mouse anti-Kaiso (Abcam, ab12723; 1:1000) and rabbit anti-LC3A/B (Abcam, ab128025; 1:1000) in antibody diluent (PBS, 0.1% Tween-20, 1% BSA) for 1 h at RT in a humidified chamber. Anti-LC3A/B antibody recognizes both LC3A and LC3B proteins. After appropriate PBS washing, cells were stained with anti-mouse Alexa Fluor 488 (Invitrogen, A-11001; 1:2000) and anti-rabbit Alexa Fluor 594 (Invitrogen, A-11002; 1:1000) secondary antibodies in a humidity chamber in the dark for 1 h at RT. Coverslips were washed and then mounted on slides using ProLong Gold antifade reagent (Molecular Probes, P36934). MDA-MB-231 and MCF-7 used in this study were obtained from ATCC and identify was validated by STRS profiling. All cells were tested and found to be free of mycobacterial contamination.

Colocalization analysis. Z-stack images for each channel were captured at ×100 magnification with Immersol 518F (Zeiss) oil immersion using a Zeiss Axiovert 200M fluorescent microscope running AxioVision software. Slices with the clearest

---

**Fig. 7 Elevated cytoplasmic Kaiso, LC3A/B, and race are associated with an immune-suppressive tumor microenvironment.**

**a** Representative multispectral quantitative immunofluorescence (mQIF) of the tumor microenvironment of a breast cancer TMA core, stained for pan-cytokeratin (Cy7, cyan); PDL1 (Cy5, red); CD8 (FITC, green); CD68 (TRITC, carmine).

**b** Coordinate map for nearest-neighbor analysis of tumor and stromal immunophenotypes.

**c** Nearest-neighbor analysis showing the frequency distribution of immune cell proximities to tumor-associated with expression quartiles (Q1-Q4) of cytoplasmic Kaiso, LC3A/B, nuclear Kaiso, and race (white versus black).
resolution were selected for further analysis. The background was removed by subtracting the mean gray value for each channel in an area containing no cells. Colocalization analysis was performed using “TMA.” The adjusted p-value was measured according to the investigator’s instructions. Negative controls were achieved by rotating one channel at least 90° and re-running the analysis. Positive controls were achieved by running colocalization analysis on two of the same channels. All graphs are plotted as the mean Pearson’s coefficient of at least three independent experiments with error bars representing the standard error. Each MCF-7 image contains ~15–20 cells, and MDA-MB-231 images contain ~5–10 cells per image.

Multi-spectral fluorescent imaging and nearest-neighbor analysis. We used the Ultive UltiMapper I/O PD-L1 assay to collect the qmIF data. This kit uses the following antigens: CD8, CD68, PD-L1, pan-cytokeratin (panCK), and DAPI (DNA marker). The raw image data is collected at 20x. The fluorescent dye intensities are normalized to 0–255. Image analysis was performed using a commercial software package (HALO, Indica Labs) at full magnification. The TMA spots were decomposed into individual analysis regions using the TMA module, with an invalidation threshold of about 80–90% empty space. A coordinate system was established for each spot with the origin being the bottom-left corner of the square that contains the TMA core. The coordinate system is equivalent to one pixels per cell. Watershed nuclear identification was performed on the DAPI channel with a nuclear contrast threshold of 0.5, and nuclear segmentation aggressiveness of 92%. Nuclei are required to be between 10 and 250μm2 in size. A cytoplasmic region was grown from the nuclear boundary up to a radius of 4.2μm. Cells were required to be >0.25μm2. The stain intensity within the cytoplasmic region was then measured, and the positive-dye status for the antigen was defined as follows: CD8 (15) CD68 (8), panCK (10), and PD-L1 (13). Overall, we observed low background and strong signals. Phenotypes are defined using the coincidence/anti-coincidence logic of the positive-dye status. The logical combination for the main cell types is as follows: CD8 (CD8+ and PD-L1-), panCK+ and panCK-, and CD68+ and CD68-. These four cell types have 3 sub-phenotypes, inclusive and PD-L1+ (and PD-L1-) or PD-L1- (and not PD-L1). The result of the phenotyping analysis is a text file for each tissue sample consisting of entries listing information about each cell location, including the manual phenotyping result and raw staining intensities using the defined coordinate system. The cell point location was taken as the center of the rectangle which fully bounds the cell.

Statistics and reproducibility. The nearest-neighbor algorithm was implemented as follows: for a given pair of phenotypes Pi, Pj, each composed of cells (detected with thresholds on their staining intensities) with two coordinates, k in [1,2], we compute, for a given cell Ci belonging to Pi, the Euclidean distances to all cells belonging to Pj, excluding those whose distance is less than 0.05 microns to prevent cell overlaps (Supplementary Figs 9 and 10). We keep the minimum distance value among those, which we call the nearest-neighbor distance, and repeat this process for each cell in Pi to form a distribution of nearest-neighbor distances, d. Measures of central tendencies for d were recorded as well as a histogram of frequencies of d, using a bin width of 2 pixels up to 120 pixels. A counting error was assigned to each bin as being the square root of the number of entries. The normalization was chosen to be the total number of cells in the sample core, such that the integral of the histogram is equal to the density of the base phenotype being considered.

Population statistics or the average histogram shape were obtained by computing the mean value for each bin given a sub-population sample. The counting error was propagated and summed in quadrature with the standard error of the mean. The combined error is shown in the shaded band. To establish a test of statistical significance between two different histograms, we first define a test statistic as being the summed log likelihood that each bin in the distribution has the same mean between the two sub-populations. The natural log of the p-value, or “likelihood”, from a t-test between the individual bin values is taken. If the hypothesis sub-population mean is larger, this likelihood is defined to be positive, else it is negative. Schematically, large positive likelihoods represent significant upward fluctuations while large negative likelihoods represent significant downward fluctuations. These likelihoods are then summed across all bins. This forms the observed (hypothesis) statistic. The summed log-likelihood was then recomputed for 1k iterations using randomly assigned sub-populations, which have the same number of patients as the test sub-population. This forms the null distribution for the test statistic. The final p-value reported is the one-sided integral of the resulting null distribution from the observed statistic (Supplementary Fig. 10). This method overestimates the p-value since real differences in the sub-population can be double-counted when building the null distribution. However, this method treats bin-to-bin correlations correctly since it samples from real data.

Patient baseline characteristics and disease factors were summarized using descriptive statistics. Categorical variables were compared using the Fisher’s exact test. Continuous and categorical variables were compared using the Pearson χ2 test. A comparison of IFHC scoring was performed by a two-sided t-test and plotted as previously described27. Univariate and multivariate Cox proportional-hazards model was used to test the independent and combined prognostic values of proteins of interest with/without the presence of selected clinical factors. Spearman rank correlations were used to assess the correlation between protein H-score and gene expression (RPKM) values78. The significance of individual hazard ratios was estimated by Wald’s test. Optimal cut-off points for H-score were determined as previously described29 (Supplementary Fig. 3). The solid lines and histogram present data for samples with levels higher (red) or lower (blue) than the median, and the IQRs are drawn in red (higher-red or lower-blue) based on the “optimal cut-off” algorithm79. Unsupervised hierarchical clustering of IHC protein score from all breast cancer samples was performed using complete linkage and distance correlations with the number of bootstrap replications (n = 1000) using the ‘prclust’ R package80. The estimated clustering stability was measured by AU (approximately unbiased) and BP (bootstrap probability) (green) value for each cluster in a dendrogram80 (Supplementary Fig. 12). To explore the expression value together with clinical-pathological information, a heatmap was drawn where patients were arranged based on the order of the hierarchical clustering outcome.

Gene set enrichment analysis of patient RNA-Seq data. The median cut-off of protein data was used to classify patients into two groups based on H-scores (e.g., low versus high Kaiso cytoplasmic) and mRNA abundance (RNA-seq). A two-sided t-test was performed, and all available genes were ranked according to p-value (lowest to highest). The p-value ranked gene list was used for functional correlation using the GSEA software (http://software.broadinstitute.org/gsea/index.jsp).

Gene set enrichment analysis of MDA-MB-231 cell line gene expression. Four (N = 4) RNA samples each from WT MDA-MB-231 and RNAi Kaiso-depleted MDA-MB-231 cells were analyzed on the Affymetrix Human Gene 2.0 ST microarray and CEL files generated were normalized to produce gene-level expression data. The functions log2 (fold change), fold enrichment with the gene symbol and fold abundance were used to compare to the MSigDB database84. The GSEA pre-ranked method from GSEA was applied for this analysis. Human Gene 2.0 ST microarray of WT versus RNAi depleted MDA-MB-231 shows significant concordance with scrambled hairpin RNAi expression MDA-MB-231 cells compared to Kaiso-depleted cell lines analyzed on the Nanostring DGEM platform (Supplementary Fig. 6 and Supplementary Table 7).

Genetic admixture analysis. For admixture analysis, RNA-Seq reads from 136 breast cancer patients were aligned to hg19 using STAR v2.5.2b85 with subsequent variant calling completed using GATK (v3.4) HaplotypeCaller44. After variant calling, Admixture v1.3.086 was used to identify ancestry proportions based on reference populations from the 1000 Genomes Project Phase 39 super populations. The IDs with a SNP burden of ≥5% across all phase 3 1000 genomes, all INDELs, and any SNPs that were notbiallelic were removed before analysis.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. RNA-seq data are available at SRA archives https://www.ncbi.nlm.nih.gov/sra?SRP158272. Proteomic and immunofluorescent intensity (x,y) coordinate point cloud data for nearest-neighbor analysis (Figs 1–7) has been uploaded to figshare https://figshare.com/s/b65562db71721e83c8be. Additional supplemental data for Figs 1–6 are also provided as Supplementary Data 1–7. Additional clinical source data and custom program code used to generate figures will be made available upon request (contact: SK Singhal).

Received: 28 February 2020; Accepted: 29 December 2020; Published online: 01 February 2021.

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. CA Cancer J. Clin. 69, 7–34 (2019).
2. DeSantis, C. E. & Jemal, A. Re: Black-White breast cancer incidence trends: the influence of ethnicity. J. Natl Cancer Inst. 101, 199–199 (2010).
3. Braden, A. M., Stankowski, R. V., V. M., Engell, M. J. & Onitilo, A. A. Breast cancer biomarkers: risk assessment, diagnosis, prognosis, prediction of treatment efficacy and toxicity, and recurrence. Curr. Pharm. Des. 20, 4879–4898 (2014).
4. Di, L., J. et al. Genome-wide profiles of CTBP link metabolism with genome stability and epithelial reprogramming in breast cancer. Nat. Commun. 4, 1449 (2013).
5. Acs, B. et al. An open source automated tumor infiltrating lymphocyte algorithm for prognosis in melanoma. Nat. Commun. 10, 5440 (2019).
6. Byun, J. S. et al. Racial differences in the association between luminal master regulator gene expression levels and breast cancer survival. Clin. Cancer Res. 26, 1905–1914 (2020).

7. Davis, M. et al. Identification of distinct heterogenic subtypes and molecular signatures associated with African ancestry in triple negative breast cancer using quantified genetic ancestry models in admixed race populations. Cancers 12, https://doi.org/10.3390/cancers12012120 (2020).

8. Debnath, J. C. et al. FOXA1 as a breast cancer prognostic marker by global gene expression meta-analysis. Cancer Res. 65, 11259–11264 (2005).

9. Hsu, J. L. & Hung, M. C. The role of HER2, EGFR, and other receptor tyrosine kinases in breast cancer. Cancer Metastasis Rev. 35, 375–388 (2016).

10. Macrina, V. & Romano, E. Clinical updates on EGFR/HER targeted agents in early-stage breast cancer. Clin. Breast Cancer 10(Suppl 1), E38–46 (2010).

11. Corso, G. et al. Hereditary lobular breast cancer with an emphasis on E-cadherin genetic defect. J. Med. Genet. 55, 431–441 (2018).

12. Fearon, E. R. Connecting estrogen receptor function, transcriptional repression, and E-cadherin expression in breast cancer. Cancer Cell 3, 307–310 (2003).

13. Berx, G. & Van Roy, F. The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. Breast Cancer Res. 3, 289–293 (2001).

14. Kaur, J. & Debnath, J. Autophagy at the crossroads of catabolism and anabolism. Nat. Rev. Mol. Cell Biol. 16, 461–472 (2015).

15. Ponpaku, M. et al. Secretory autophagy. Curr. Opin. Cell Biol. 35, 106–116 (2015).

16. Leidal, A. M. et al. The LC3-conjugation machinery specifies the loading of RNA-binding proteins into extracellular vesicles. Nat. Cell Biol. https://doi.org/10.1038/s41556-019-0450-y (2020).

17. Leidal, A. M. & Munz, C. LC3-associated phagocytosis. Autophagy 10, 526–528 (2014).

18. Rak, J. (C3)icensing of exosomes for RNA export. Nat. Cell Biol. https://doi.org/10.1038/s41556-020-0466-3 (2020).

19. Liu, J. & Debnath, J. The evolving, multifaceted roles of autophagy in cancer. Adv. Cancer Res. 130, 1–53 (2016).

20. Bereczewska, S. & Galvan, A. Immunohistochemical detection of the autophagy markers LC3 and p62/SQSTM1 in formalin-fixed and paraffin-embedded tissue. Methods Mol. Biol. 1560, 189–194 (2017).

21. Homma, K., Suzuki, K. & Sugawara, H. The Autophagy Database: an all-inclusive information resource on autophagy that provides nourishment for research. Nucleic Acids Res. 39, D986–990 (2011).

22. Liberson, A. A description of the Molecular Signatures Database (MSigDB) web site. Methods Mol. Biol. 1150, 153–160 (2014).

23. New, J. & Thomas, S. M. Autophagy-dependent secretion: mechanism, factors secreted, and disease implications. Autophagy 15, 1682–1693 (2019).

24. Bhutia, S. K. et al. Monitoring and measuring mammalian autophagy. Methods Mol. Biol. 1854, 209–222 (2018).

25. du Toit, A., Hofmeyr, J. S., Gniadek, T. J. & Loos, R. Measuring autophagosome flux. Autophagy 14, 1060–1071 (2018).

26. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).

27. Alexander, D. H., November, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664 (2009).

28. Li, X. et al. Infiltration of CD8(+) T cells into tumor cell clusters in triple-negative breast cancer. Proc. Natl Acad. Sci. USA 116, 3678–3687 (2019).

29. Kaur, H. B. et al. Association of tumor-infiltrating T-cell density with molecular subtype, racial ancestry and clinical outcomes in prostate cancer. Mod. Pathol. 31, 1539–1552 (2018).

30. Savas, P. et al. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. Nat. Rev. Clin. Oncol. 13, 228–241 (2016).

31. Roux, C. et al. Reactive oxygen species modulate macrophage immunosuppressive phenotype through the up-regulation of PD-L1. Proc. Natl Acad. Sci. USA 116, 4326–4335 (2019).

32. Wang, Y. et al. Regulation of PD-L1: emerging routes for targeting tumor immune evasion. Front. Pharmacol. 9, 536 (2018).

33. Brochez, L. et al. CHALLENGING PD-L1 EXPRESSING CYTOTOXIC T CELLS AS A PREDICTOR FOR RESPONSE TO IMMONOTHERAPY IN MELANOMA. Front. Immunol. 10, 1480–1421 (2019).

34. Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. Nat. Rev. Genet. 15, 379–393 (2014).

35. Scharf, T. A. et al. Genetic variants in immune-related pathways and breast cancer risk in African American women in the AMBER Consortium. Cancer Epidemiol. Biomark. Prev. 27, 321–330 (2018).

36. Deschamps, M. et al. Functional importance of non-coding variants derived from archaic hominins at human innate immunity genes. Am. J. Hum. Genet. 98, 5–21 (2016).

37. Quan, L. et al. Cytokine and cytokine receptor genes of the adaptive immune response are differentially associated with breast cancer risk in African American women of African and European ancestry. Int. J. Cancer 134, 1408–1421 (2014).

38. Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. Nat. Rev. Genet. 15, 379–393 (2014).
67. Ness, R. B., Haggerty, C. L., Harger, G. & Ferrell, R. Differential distribution of allelic variants in cytokine genes among African Americans and White Americans. *Am. J. Epidemiol.* **160**, 1033–1038 (2004).

68. Lee, J., Kim, D. M. & Lee, A. Prognostic role and clinical association of tumor-infiltrating lymphocyte, programmed death ligand-1 expression with neutrophil-lymphocyte ratio in locally advanced triple-negative breast cancer. *Cancer Res. Treat.* **51**, 649–663 (2019).

69. Park, I. H. et al. Prognostic implications of tumor-infiltrating lymphocytes in association with programmed death ligand 1 expression in early-stage breast cancer. *Clin. Breast Cancer* **16**, 51–58 (2016).

70. Kalvari, I. et al. ILIR: A web resource for prediction of Atg8-family interacting proteins. *Autophagy* **10**, 913–925 (2014).

71. Shariati, M. N. et al. Autophagy promotes focal adhesion disassembly and cell motility of metastatic tumor cells through the direct interaction of paxillin with Lc3. *Cell Rep.* **15**, 1660–1672 (2016).

72. Jacquelot, N. et al. Predictors of responses to immune checkpoint blockade in advanced melanoma. *Nat. Commun.* **8**, 592 (2017).

73. Hewitt, S. M. The application of tissue microarrays in the validation of microarray results. *Methods Enzymol.* **410**, 409–415 (2006).

74. Khouy, T. et al. Ki-67 expression in breast cancer tissue microarrays: assessing tumor heterogeneity, concordance with full section, and scoring methods. *Am. J. Clin. Pathol.* **148**, 108–118 (2017).

75. Jia, W. et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol.* **14**, R12 (2013).

76. Peng, Z. et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* **30**, 253 https://www.nature.com/articles/nbt.2122#supplementary-information (2012).

77. Hintze, J. L. & Nelson, R. D. Violin plots: A box plot-density trace synergism. *J. Comput. Stat. Data Anal.* **43**, 105–122 (2003).

78. Myers, J. L., Well, A. & Lorch, R. F. *Research Design and Statistical Analysis* 3rd edn (Routledge, 2010).

79. Myers, J. L., Well, A. & Lorch, R. F. *Research Design and Statistical Analysis* 3rd edn (Routledge, 2010).

80. Suzuki, R. & Shimodaira, H. Pclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

81. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

82. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).

83. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

84. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

85. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

86. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

87. Aman, J. A. et al. Global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

88. Ringner, M., Fredlund, E., Hakkinen, J., Borg, A. & Staaf, J. GOBO: gene expression-based outcome for breast cancer online. *PLoS ONE* **6**, e17911 (2011).