Supplemental Results

Genomic Concordance

WES and sWGS performed on paired primary and metastatic tumors, coupled with multi-region sequencing (MRS) of multiple blocks from the same tumor, revealed that somatic coding mutations and whole-genome CN were each markedly less concordant for primary-metastatic tumor pairs than for multiple regions of the same tumor (mean Jaccard index of shared mutations: paired tumors = 0.29 ± 0.21, MRS = 0.78 ± 0.14, $P = 5.4 \times 10^{-7}$; mean R^2 of genome-wide CN: paired tumors = 0.55 ± 0.29, MRS = 0.86 ± 0.17, $P = 3.8 \times 10^{-4}$; Figure 1A). Notably, concordance metrics for mutations and CN were highly correlated ($P = 5.8 \times 10^{-5}$, Figure 1B). The extent of concordance observed between differentially preserved (*i.e.*, FFPE vs. OCT) tumor pairs did not differ significantly from the concordance observed between tumor pairs with the same preservation type, indicating that the relatively low concordance between paired tumors was independent of preservation type for both WES and sWGS.

As anticipated from their ancestral relationship, the majority of metastases were clonally related to their primary tumors of origin, as evidenced by co-clustering of 27 out of 28 primary-metastatic tumor pairs according to recurrent somatic coding mutations (Supplemental Figure 2). Metastases exhibited a greater number of coding mutations ($P = 6.9 \times 10^{-5}$) and CNAs ($P = 0.0093$) compared with the primary tumors from which they arose (Figure 1C). This was accompanied by an increase in the number of large-scale state transitions (LSTs) in metastatic tumors ($P = 3.1 \times 10^{-6}$), defined as the number of occurrences in which adjacent genomic segments of ≥10 Mbp change CN state (1).

Support for SAR DSS associations within METABRIC

Aggregating *P*-values (Fisher's method) across DSS associations within the METABRIC study for loss of *CDKN2A*, loss of *STK11*, gain of *PTK6*, and gain of *PAQR8* resulted in $P = 1.6 \times 10^{-13}$. When testing the reverse direction for each gene (*e.g.*, *CDKN2A* gain instead of loss), the aggregate *P*-value was much greater ($P = 0.078$), demonstrating that the DSS associations of these 4 SARs were specific to CN directions exhibited by metastases and, thus, not a result of increased CN burden in metastases. This conclusion was further supported by comparing aggregated *P*-values to 10,000 aggregated *P*-values generated by testing DSS associations of 4 randomly selected genes and CN directions. The true aggregate *P*-value was more significant than 99% of permuted tests ($P_{perm} = 0.009$), whereas the aggregated *P*-value from testing reverse directions was not significant ($P_{perm} = 0.83$). This provides further evidence that the DSS associations of metastasis-enriched SARs do not simply result from an increased CN burden in metastases. This was also true when the analysis was limited to ER^+ tumors. When testing the CN directions observed in metastases, the aggregate *P*-value ($P = 2.6 \times 10^{-12}$) was more significant than 97% of permuted tests ($P_{perm} = 0.027$). This was not the case when reverse directions were tested (aggregated $P = 0.33$, $P_{perm} = 0.94$).

mTOR pathway components are preferentially mutated in metastases

The mTOR Signaling pathway gene set was preferentially mutated in metastases compared with primary tumors both in our paired tumor dataset ($P_{perm} = 0.002$) and in the Lefebvre dataset compared with TCGA-BRCA ($P_{perm} = 0.052$). Mutations in the mTOR Signaling pathway within metastatic tumors in our cohort occurred in several functionally distinct modules (Supplemental Figure 5), of which the most highly mutated was the core

PI3K/AKT signaling pathway. However, primary and metastatic tumors exhibited similar frequencies of *PIK3CA* mutations, which included common (H1047R/Q/L [n = 14], E545K [n = 6], E542K/Q [n = 4]) (2), as well as rare, variants (E39K, C378W, E726K).

In contrast to mutations in *PIK3CA*, *AKT1* mutations occurred solely in metastases and included the activating mutations E17K (n = 3) and D323H (n = 1) (3, 4). Two novel *AKT2* missense mutations, S398I and S1098R, were identified in a metastasis from 1 patient. An additional patient harbored a 1 bp frameshift deletion in *PTEN* that co-occurred with focal deletion in both primary and metastatic tumors, abrogating PTEN function. Mutations specific to the PI3K-AKT Signaling gene set included components of JAK/STAT signaling (*JAK1* [n = 3], *JAK2* [n = 2], *JAK3*, *STAT5B*), as well upstream growth factors (*FGF10*, *FGF20*, *NGF*), hormones (*GH1*, *GH2*), and their respective receptors (*FGFR4*, *GHR*).

Metastases preferentially harbored mutations in mTOR modules other than PI3K/AKT, including the MAPK, WNT, AMPK, and amino acid sensing pathways. Within the MAPK module of the mTOR Signaling gene set, mutations occurred exclusively in metastases and co-occurred with CN gain in 5 out of 6 tumors, with 1 metastasis harboring both the activating *KRAS* G12V mutation and CN gain of *MYC*. Though not included in the mTOR Signaling gene set, 3 metastases (4.5%) exhibited mutations in *NF1* that were likely deleterious (nonsense and/or co-occurred with deletion or copy neutral loss of heterozygosity). Mutations in the MAPK Signaling gene set were also enriched in the Lefebvre dataset compared with TCGA-BRCA primary tumors (P_{perm} = 0.029).

Upstream of MAPK and PI3K/AKT signaling, a large proportion of primary and metastatic tumors harbored at least 1 mutation in an RTK (28.6% and 34.8%, respectively). Intriguingly, *PTK6*, which is preferentially amplified in metastases, has been reported to interact

with a number of RTKs, thereby activating AKT1 (5-8). Although the mTOR Signaling gene set does not include *PTK6* nor most RTKs, their alteration would be anticipated to up-regulate mTOR activity.

Metastatic tumors also exhibited significantly more mutations in the WNT signaling module of the mTOR Signaling gene set (Supplemental Figure 5B). Within the mTOR Signaling gene set, WNT pathway mutations were found exclusively in metastatic tumors and affected multiple WNT ligands and receptors (LRPs and FZDs), including the *LRP5* activating mutation, A65V (9).

TSC1/2 inhibition via upstream PI3K/AKT, MAPK, WNT, and AMPK signaling represents a critical node in mTOR pathway activation. Four metastatic tumors, but no primary tumors, harbored mutations in *TSC1* ($n = 3$) and *TSC2* ($n = 1$). *TSC1/2* mutations have recently been reported to be enriched in metastatic breast cancer compared with primary tumors (10), with frequencies of 6% in both our cohort and the Lefebvre et al. dataset.

Unlike upstream pathways that affect mTORC1 activity by regulating TSC1/2, the amino acid sensing module directly impacts mTORC1 activity. High concentrations of amino acids induce binding of the Rag subfamily of Ras small GTPases (RagA/B/C/D) to Raptor, a component of mTORC1 (11, 12). Rag binding, in turn, promotes localization of mTORC1 to the surface of Rheb-containing amino acid-rich lysosomes, ultimately resulting in mTORC1 activation. Mutations were identified in components of the v-ATPase complex ($n = 6$), which stimulates the ability of Ragulator to promote mTORC1-Rag binding via Raptor. Additional mutations were identified in genes that regulate mTORC1-Rag binding, including components of the GATOR1/2 complexes. Overall, 22.7% of metastases, but only 7.1% of primary tumors, harbored mutations in the amino acid sensing module ($P = 0.061$); moreover, 14 components of

this module were mutated in metastases, with only 2 being mutated in primary tumors (Supplemental Figure 6).

Multiple novel mutations were identified in metastatic tumors within subunits common to mTORC1 and mTORC2 (*MTOR* [$n = 2$], *MLST8*, *DEPTOR*), or specific to mTORC2 (*RICTOR* [$n = 2$]). Intriguingly, while the *RICTOR* mutation, K1125E, has not been reported in human cancers, acetylation of K1125 in response to elevated glucose and/or acetate levels activates mTORC2 independently of upstream PI3K/AKT signaling (13). Mutations were also found in several downstream targets of mTORC1/2 involved in cytoskeletal organization, lipid biosynthesis, autophagy, cell growth and proliferation. Particularly interesting amongst this group includes mutations in 2 metastatic tumors in *RHOA*, G14E and G17V, the latter of which is frequently found in lymphomas where it promotes proliferation via activation of PI3K and MAPK signaling (14).

Of the 36 mutations with known or putative effect, 24 (67%) mutations are predicted to result in increased mTOR activity and 12 (33%) mutations are predicted to lead to decreased mTOR activity, with mutations being consistent with activation in 17 (26%) metastases and inactivation in 6 (9%) metastases (4 metastases had both activating and inactivating mutations).

Alterations in the WNT Signaling pathway

The Wnt Signaling pathway gene set was identified as a metastasis-enriched SMP in our cohort ($P_{perm} = 0.029$) and exhibited trending enrichment between the Lefebvre and TCGA-BRCA datasets ($P_{perm} = 0.22$). Metastases harbored mutations in several genes within the WNT Signaling gene set (Supplemental Figure 13), including multiple WNT ligands (*WNT1*, *WNT2B*, *WNT11*) and receptors (*LRP5* [$n = 2$], *LRP6* [$n = 2$], *FZD1*, *FZD2*, *FZD3*, *FZD7*, *FZD10*), as well as genes involved in receptor-ligand interactions (*WIF1* [$n = 2$], *BAMBI*, *NOTUM*,

SERPINF1) and regulation of Dvl (*CSNK1E* [*n* = 2], *PRICKLE1*, *INVS*, *VANGL1*), which affect downstream mTOR, MAPK, focal adhesion, and cell cycle pathways. Notably, 1 metastasis exhibited an activating mutation in the WNT receptor, *LRP5* A65V, which abrogates the ability of Dkk1 to inhibit LRP5 and resulting pathway activity (9).

Metastases also harbored mutations in genes that regulate β-catenin (*CHD8* [*n* = 3], *APC2*, *AXIN1*, *FRAT1*, *CSNK2A1*, *CSNK2A3*), cell cycle promoting transcription factors (*SMAD4*, *LEF1*), ubiquitin-mediated proteolysis (*CUL1*, *TBL1X*, *TBL1XR1*), and calcium signaling (*PLCB1* [*n* = 2], *PLCB3* [*n* = 2], *PLCB4*, *NFATC1*, *NFATC3*, *CAMK2D*, *PPP3CB*, *PRKCB*). Similar numbers of putatively activating and inactivating mutations occurred within canonical WNT (10 GOF and 9 LOF) and calcium (4 GOF and 4 LOF) signaling sub-pathways.

Mutations in the cAMP Signaling pathway

The cAMP Signaling pathway gene set was also identified as a metastasis-enriched SMP in our cohort (P_{perm} = 0.045) and within the Lefebvre dataset compared with TCGA-BRCA (P_{perm} = 0.045). The second messenger, cyclic adenosine monophosphate (cAMP), mediates a broad array of cellular processes, including apoptosis, metabolism, differentiation, and proliferation, and exerts its effects primarily through activation of the cAMP-dependent kinase, PKA (19). Mutations in this pathway (Supplemental Figure 14) were identified in many genes whose products regulate cellular cAMP levels, including adenylyl cyclases (*ADCY10* [*n* = 2], *ADCY4*, *ADCY5*, *ADYC6*, *ADYC7*, *ADYC9*); phosphodiesterases (*PDE3A*, *PDE4D*); a membrane transporter that regulates cAMP efflux (*ABCC4* [*n* = 2]); and G-coupled protein receptors that activate (*ADRB1*, *DRD1*, *DRD5*, *TSHR*, *GLP1R*, *PTGER2*) or inhibit (*CHRM1* [*n* = 2], *CHRM2*, *GHSR*, *GNAI3*, *HTR1B*, *HTR1F*, *SSTR1*, *SSTR5*) adenylyl cyclases.

Downstream of PKA, mutations were identified in several sub-pathways, including PI3K/AKT, MAPK, Rac GEFs, and genes involved in the regulation of actin (described above). Additional mutations were identified in genes involved in RAP1 signaling (*RAPGEF4* [*n* = 3], *PLCE1*, *VAV1*, *VAV2*), CREB binding (*PPP1CB* [*n* = 2], *PPP1R12A* [*n* = 2], *PPP1CA*, *PPP1CC*, *EP300* [*n* = 2], *CREB3L3* [*n* = 2], *CREB3L4*, *CREBBP*), NFkB signaling (*NFKB1*, *AMH)*, fatty acid degradation (*ACOX1*), cell migration (*ARAP3*), and calcium signaling (*NFATC1*, *CAMK2D*), as well as PKA-activated P-type primary ion transport ATPases (*ATP1A1* [*n* = 2]*, ATP1A3* [*n* = 2]*, ATP1A2, ATP1A4, ATP1B3, ATP1B4, ATP2A2, ATP2B1, ATP2B2*), glutamate-regulated ion AMP acid receptors (*GRIA1, GRIA2, GRIA4),* and glutamate-regulated ion N-methyl-D-aspartate receptors (*GRIN3A* [*n* = 3]*, GRIN2A* [*n* = 2]*, GRIN1* [*n* = 2], *GRIN3B, GRIN2B*). In total, 17 mutations were predicted to result in activation, and 12 mutations inactivation, of this pathway, with mutations in aggregate being consistent with activation in 10 (15%), and inactivation in 6 (9%), metastases.

Mutations in the Carbohydrate Digestion and Absorption pathway

The Carbohydrate Digestion and Absorption pathway gene set was also identified as a metastasis-enriched SMP in our cohort (P_{perm} = 0.055) and when the Lefebvre dataset was compared with TCGA-BRCA (P_{perm} = 0.10). This pathway shares with the cAMP Signaling pathway mutations in the core PI3K/AKT pathway and in PKA-activated P-type primary ion transport ATPases. Unique to this pathway, however, are mutations in hexokinases (*HKDC1* [*n* = 3], *HK1* [*n* = 2]) and genes that regulate the initial digestion of sugars (*MGAM*, *SI*), as well as in a G protein alpha subunit (*GNAT3*) and a G-protein coupled receptor (*TAS1R3*).

Mutations in the Progesterone Mediated Oocyte Formation pathway

The Progesterone Mediated Ooctye Formation pathway gene set was also identified as a metastasis-enriched SMP in our cohort (P_{perm} = 0.022) and when the Lefebvre dataset was compared with TCGA-BRCA (P_{perm} = 0.028) and includes genes that regulate insulin and/or progesterone-mediated inhibition of G2-arrest in *Xenopus* oocytes via activation of cyclin B. Like several other metastasis-enriched pathways identified in this study, the Progesterone Mediated Oocyte Formation pathway includes the core PI3K/AKT and MAPK pathways and affects genes that regulate the cell cycle. Furthermore, like the cAMP signaling pathway, upstream activators within the Progesterone Mediated Oocyte Formation pathway, such as insulin, are expected to alter intracellular cAMP levels by regulating adenylyl cyclases. Though *PAQR8* is not explicitly included in this KEGG gene set, this pathway is described as being driven by G-protein coupled non-genomic membrane-bound progestin receptors (mPRs) due to their ability to inhibit adenylyl cyclase activity. Mutations unique to this metastasis-enriched pathway include regulators of APC/C (*ANAPC7* [n = 2], *ANAPC11*), microtubule formation (*KIF22* [n = 2], *AURKA*), CDK1/2 (*PKMYT1* [n = 2], *SPDYC*), and the nuclear progesterone receptor, *PGR* (n = 2).

Mutations in the Focal Adhesion pathway

The Focal Adhesion pathway gene set was identified as a metastasis-enriched SMP both in our tumor cohort (P_{perm} = 0.0013) and within the Lefebvre et al. dataset compared with TCGA-BRCA (P_{perm} = 0.078). Beyond shared modules with other metastasis-enriched pathways (*e.g.*, RTKs, PI3K/AKT, WNT, and MAPK signaling), mutations in metastases were identified in several components unique to the Focal Adhesion pathway (Supplemental Figure 15). For

example, within the "ECM-Receptor Interaction" module, frequent mutations were identified in metastases in collagens (26% of metastases harbored mutations in 14 members: *COL1A1* [*n* = 2], *COL4A4* [*n* = 2], *COL4A5* [*n* = 2], *COL4A6* [*n* = 2], *COL6A3* [*n* = 2], *COL6A5* [*n* = 2], *COL2A1*, *COL4A1*, *COL4A2*, *COL4A3*, *COL6A2*, *COL9A1*, *COL9A2*, *COL9A3*), integrin-α and β subunits (15% of metastases harbored mutations in 9 members: *ITGA9* [*n* = 2], *ITGB4* [*n* = 2], *ITGA1*, *ITGA11*, *ITGA4*, *ITGA7*, *ITGA8*, *ITGB3*, *ITGB6*), and laminins (11% of metastases harbored mutations in 7 members: *LAMA3* [*n* = 4], *LAMA1* [*n* = 2], *LAMB1* [*n* = 2], *LAMA4*, *LAMA5*, *LAMC1*, *LAMC3*). Other gene families mutated in this module included: platelet-derived growth factors and receptors (*PDGFC*, *PDGFD*, *PDGFRB*), integrin-binding tenascins (*TNXB* [*n* = 2], *TNC*, *TNN*), and thrombospondins (*THBS2* [*n* = 2], *THBS3*), which regulate collagen fibrillogenesis (15). Frequent mutations also occurred in *RELN* (*n* = 6), which promotes cell adhesion to the ECM via activation of integrin-β1 (16), and *VWF* (*n* = 4), which binds integrin-$\alpha_v\beta_3$ (17) and promotes platelet adhesion (18).

Downstream of integrin signaling, mutations in metastases were identified in several members of the FAK, SRC, RAS and RAC signaling cascades, including *FAK* and *SRC* genes themselves, Rho GTPase activating proteins (*ARHGAP5*, *ARHGAP35*), and guanine nucleotide exchange factor (GEF) activators of both Rac (*VAV1*, *VAV2*, *DOCK1*) and Ras (*RASGRF1* [*n* = 3]). Within the "Regulation of Actin Cytoskeleton" module of the Focal Adhesion pathway gene set, mutations in metastatic tumors were detected in the myosin light chain kinases, *MYLK* (*n* = 4), which was also identified as a metastasis-specific SMG, and *MYLK3*, as well as their substrate, *MYL10.* Metastatic tumor mutations were also identified in actin-binding actinins (*ACTN1* [*n* = 2], *ACTN4*), regulatory subunits of protein phosphatase I (*PPP1CB* [*n* = 2], *PPP1R12A* [*n* = 2], *PPP1CA*, *PPP1CC*), and regulators of actin assembly (*TLN1*, *TLN2*), cross-

linking (*FLNC* [*n* = 3]), polymerization (*PARVA* [*n* = 2], *RHOA* [*n* = 2], *ROCK1*, *DIAPH1*, *PXN*), and stabilization (*PAK4*). Finally, though not included in the Focal Adhesion gene set, we identified 2 additional metastasis-specific SMGs whose functions impact on the actin cytoskeleton: *XIRP2* and *PEAK1*.

Of the 41 mutations with putative effect within collagens, laminins, thrombospondins, integrin subunits, tenascins, and growth factors, as well FAK (*PTK2*) and *SRC*, 23 were GOF and 18 were LOF, with mutations being consistent with activation in 7 metastases and inactivation in 10 metastases (7 metastases had both GOF and LOF mutations).

Mutations in the Longevity Regulating pathway

The Longevity Regulating pathway gene set was also identified as a metastasis-enriched SMP in our cohort (P_{perm} = 0.076) and when the Lefebvre dataset was compared with TCGA-BRCA (P_{perm} = 0.0095) and includes genes that are involved in the ability of caloric restriction to increase lifespan. Like other metastasis-enriched pathways, the Longevity Regulating Pathway includes core PI3K, MAPK, mTOR, and cell-cycle pathways, as well as adenylyl cyclases and subunits of CREB. Mutations within genes specific to this metastasis-enriched pathway include: HKMTs (*EHMT1*, *EHMT2*), an adiponectin receptor, *ADIPOR1*; an AMPK kinase, *CAMKK2* (*n* = 2); a subunit of PPAR-gamma, *PPARGC1A*; a pRB-regulator, *RB1CC1*, and regulators of insulin signaling (*IRS4* [*n* = 3], *IRS2*, *KL*, *SESN3*).

Mutations in the HIF-1 Signaling pathway

The HIF-1 Signaling pathway gene set was also identified as a metastasis-enriched SMP in our cohort (P_{perm} = 0.082) and when the Lefebvre dataset was compared with TCGA-BRCA

(P_{perm} = 0.040). Like other metastasis-enriched pathways, the HIF-1 Signaling pathway includes core PI3K, MAPK, and mTOR pathways, as well as JAK-STAT. Unique to this pathway, the majority of mutations were identified in genes that are transcriptionally regulated by HIF-1. These include: glucose metabolism regulators (hexokinases: *HKDC1* [*n* = 3], *HK1* [*n* = 2]; phosphofructokinases: *PRKFB3* [*n* = 2], *PFKL*), iron binding transporters (*TF*, *TFRC*), an endothelin (*EDN1* [*n* = 2]), a nitrous oxide synthase (*NOS2*), a kinase that inhibits pyruvate dehydrogenase (*PDK1*), and the glucose transporter *SLC2A1* (also known as GLUT1).

Mutations in the Regulation of Lipolysis in Adipocytes pathway

The Regulation of Lipolysis in Adipocytes pathway gene set was also identified as a metastasis-enriched SMP in our cohort (P_{perm} = 0.0064) and when the Lefebvre dataset was compared with TCGA-BRCA (P_{perm} = 0.035). Like other metastasis-enriched pathways, the Regulation of Lipolysis in Adipocytes pathway includes core PI3K, MAPK, mTOR, and cell-cycle pathways, as well as genes that regulate cAMP. Mutations in genes unique to this pathway include insulin receptors (*IRS4* [*n* = 3], *IRS2*), isoforms of PKG (*PRKG1* [*n* = 2], *PRKG1*), which activate perilipin, *PLIN*, and hormone-sensitive lipase, *LIPE*, in the presence of cGMP, as well as downstream genes that regulate lipolysis (*PNPLA2* and *PNPLA2*).

Mutations in the VEGF Signaling pathway

The VEGF Signaling pathway gene set was also identified as a metastasis-enriched SMP in our cohort (P_{perm} = 0.018) and when the Lefebvre dataset was compared with TCGA-BRCA (P_{perm} = 0.10). Like other metastasis-enriched pathways, the VEGF Signaling pathway includes core PI3K, MAPK, and mTOR pathways. Mutations in genes unique to this pathway include

genes that hydrolyze phospholipids (*PLA2G4E* [$n = 2$], *JMJD7-PLA2G4B* [$n = 2$]), the sphingosine kinase, *SPHK2*, and calcium-dependent, calmodulin-stimulated protein phosphatase, *PPP3CB*.

Mutations in the Prolactin Signaling pathway

The Prolactin Signaling pathway gene set was also identified as a metastasis-enriched SMP in our cohort ($P_{perm} = 0.013$) and when the Lefebvre dataset was compared with TCGA-BRCA ($P_{perm} = 0.0024$). Like other metastasis-enriched pathways, the Prolactin Signaling pathway includes core PI3K, MAPK, and mTOR pathways, as well as JAK/STAT components. Mutations in genes unique to this pathway include nuclear estrogen receptors (*ESR1* [$n = 7$], *ESR2*); the G-protein coupled receptor, *LHCGR* ($n = 2$), which promotes ovarian steroidogenesis; the epithelial-specific Ets transcription factor, *ELF5*; a tumor necrosis factor, *TNFSF11*; and *GCK* (Hexokinase 4), which plays an important glucose regulatory role in liver and pancreatic islet beta cells.

Mutations in the Lysine Degradation pathway

The Lysine Degradation pathway gene set was also identified as a metastasis-enriched SMP in both our cohort ($P_{perm} = 0.020$) and when comparing the Lefebvre and TCGA-BRCA datasets ($P_{perm} = 0.041$). Although several mutations were identified in metastases in genes regulating acetyl-CoA production (*ALDH7A1*, *ALDH9A1*, *COLGALT1*, *OGDH*, *PLOD2* and *TMLHE*), the majority of mutations in this KEGG-defined pathway occurred in SET-domain-containing histone lysine methyltransferases (HKMTs) from several different gene families. The KMT2 family of lysine methyltransferases was the most frequently mutated and included

KMT2C (*MLL3*, *n* = 9), *KMT2D* (*MLL4*, *n* = 3), *KMT2E* (*MLL5*, *n* = 3), *SETD1B* (*n* = 3), *SETD1A*, *ASH1L*, and *KMT2A* (*MLL1*), each of which methylate H3K4, a mark of transcriptionally active chromatin. Mutations were also identified in KMT3 family members, *NSD1* (*n* = 3) and *SETD2* (*n* = 2), which methylate H3K36 and H3K4, markers of actively transcribed genes, and in the KMT1 (*EHMT1*, *EHMT2*, *SETDB1*, *SUV39H1*), KMT5 (*KMT5B*, *KMT5C*), and KMT8 (*PRDM2*) families, which methylate H3K9, a marker of transcriptionally repressed chromatin. One mutation was found in *SETMAR*, which is a fusion gene that contains a N-methyltransferase domain and a C-terminal transposase domain. The majority of mutations in *KMT2A/C/D* within our cohort were deleterious. Of the 20 mutations within HKMTs with putative effect, 13 (65%) are predicted to result in LOF and 7 (35%) in GOF, with mutations being consistent with inactivation in 12 metastases (18%) and activation in 6 metastases (9%).

Supplemental Methods

Breast cancer cohort

Patients with either newly suspected, untreated metastatic breast cancer or progressing disease were eligible for the METAMORPH study if they had: 1) a history of histologically-confirmed primary breast cancer; 2) clinical or imaging evidence suggestive of recurrent breast cancer in a local, regional, or distant location; 3) willingness to undergo and/or provide tissue from a recent biopsy of recurrent tumor for both clinical and research testing; and 4) willingness to undergo blood specimen collection. Patients were excluded if they were on anticoagulation that could not be interrupted for the purpose of study procedures. Additional unpaired specimens were provided for IHC analyses by the TRACR project.

Approximately 50% of tumors were provided as 10μm sections and required macro-dissection to remove non-tumor tissue, guided by top and bottom Hematoxylin and Eosin-stained (H & E) sections. Tumor tissue from intact FFPE tumor blocks was isolated using a heating block (58°C) and xylene was used for deparaffinization prior to DNA extraction. Intact OCT core biopsies were thawed in a PBS bath at 4°C and the tissue washed 3 times in PBS at 4°C. DNA was extracted from FFPE and OCT-preserved tumors using the Qiagen AllPrep DNA/RNA FFPE Kit and Qiagen AllPrep DNA/RNA Micro Kit, respectively. Germline DNA was extracted from buffy coat prepared from blood for each enrolled patient using the Qiagen Gentra Puregene Blood Kit.

Whole-exome sequencing

DNA was submitted for WES to the High-Throughput Sequencing Center at the Children's Hospital of Philadelphia and the Beijing Genomics Institute (BGI). Sequencing was performed on the Illumina HiSeq4000 using 100 bp or 150 bp paired-end reads using the SureSelect v5 library preparation kit (Agilent). OCT and germline samples were sequenced at 100x and 50x target coverage, respectively. For FFPE samples, libraries were manually generated and quality confirmed prior to sequencing to reach a target data size of 10GB (~50M reads) after removal of low-quality reads.

WES reads were first trimmed for low-quality base-pairs (Trimmomatic v0.36) (20). Paired reads and unpaired reads whose partners were removed during trimming were separately aligned to hg38 (BWA v0.7.12) (21) and then merged (samtools v1.3.1) (22). Duplicated reads were removed (PicardTools v2.6: http://broadinstitute.github.io/picard/) and the remaining reads were realigned around indels (GATK v3.6) (23). After processing, median average coverage

across target regions was 108x [IQR = 85-123x] for tumors and 70x [IQR = 66-76x] for germline samples.

Three variant callers were used to identify somatic mutations in tumors in reference to matching germline samples: MuTect1 v1.1.4 (23), MuTect2 - GATK v3.6 (23), and VarScan v2.3.4 (24). Default parameters were used for all steps of read processing and variant calling. Variants called by at least 2 of the 3 callers were considered for further analysis. Variants that were reported in at least 1 normal sample by VarScan2 or were included in the MuTect2 generated "panel-of-normals" (PON), which includes variants and recurrent artifacts occurring in at least 2 normal samples, were removed from further analysis. Variants were annotated using snpEff v4.2 (25). VarScan was used to identify germline variants in normal samples and VarfromPDB (26) was used to identify germline variants of clinical significance.

Three tiers of confidence were assigned to resulting variants – tier I: coverage $\geq$ 30 reads, alternative allele read coverage (AAC) $\geq$ 6, VAF $\geq$ 0.10; tier II: coverage $\geq$ 10 reads, VAF $\geq$ 0.10; tier III: reported by at least 2 variant callers but did not meet tier I or II criteria. Since mean concordance between sequencing replicates and differentially preserved tumor samples was highest when limited to tier I mutations, analyses used tier I mutations unless otherwise stated. Coding mutations are defined as those that were annotated as 1 of the following: missense, splice-region, nonsense, start-loss, stop-loss, exon-loss, and exonic indel.

MutSigCV2 (27) was used to identify genes that were significantly mutated (SMGs) based on tumor-specific mutation rates, mutation-specific nucleic-acid context, and gene-specific length and nucleic-acid composition, using covariates (gene expression, replication timing, and chromatin-state) provided by the Broad Institute, as well as a GC covariate provided by Ensembl (28). In lieu of the proxy coverage data file provided for MutSigCV2 by the Broad Institute, per-

base-pair/mutation-type coverage data was generated from aligned sequenced data specifically from samples in this cohort. The significance of genes was only considered for those having coding mutations in at least 4 samples in the metastatic tumor cohort.

Shallow whole-genome sequencing

For sWGS, DNA was first sheered to 350bp target size using a Covaris S220 sonicator with the following parameters: 18W peak power, 20% duty factor, 50 cycles per burst, 65 second duration, 20°C temperature, and water level at 2mm below cap of microTUBE-15. The Illumina NeoPrep system was used to prepare libraries of single-end DNA reads, which were sequenced on a NextSeq500 to a minimum of 6M 75 bp single-end reads per sample.

Reads from sWGS were trimmed of low-quality base-pairs, aligned, then removed of duplicates using the same WES processing tools described above. A median of 9.8M reads were mapped, resulting in median coverage of 0.22x [IQR = 0.18-0.30x]. QDNAseq (29) was then used to correct read counts across 15 kbp-wide genomic bins based on sequence mappability and GC content, and to remove problematic genomic regions with blacklist status in the 1000 Genomes Project (30). Hg19 was used as the reference genome for sWGS in order to take advantage of existing reference annotations (29). CN values for resulting segments were calculated by adjusting QDNAseq "signal" output by tumor-specific cellularity and ploidy identified by Sequenza (31). DNA was unavailable for sWGS for 1 primary and 3 metastatic tumors that were assayed by WES; CN calls from Sequenza were used as proxy for these samples. CNAs within chromosome X were not included in analyses due to a lack of annotated "blacklist" regions, which are used by QDNAseq for autosomal chromosomes.

Formulas for adjusted, relative, and normalized CN values are given below.

$$\boldsymbol{adjusted\ CN} = \frac{\mathbf{2}^{\boldsymbol{signal}} \times (\boldsymbol{ploidy} \times \boldsymbol{cellularity} + \mathbf{2} \times [\mathbf{1} - \boldsymbol{cellularity}]) - \mathbf{2} \times (\mathbf{1} - \boldsymbol{cellularity})}{\boldsymbol{cellularity}}$$

$$\boldsymbol{relative\ CN} = \frac{\boldsymbol{adjusted\ CN}}{\boldsymbol{ploidy}}, \boldsymbol{normalized\ CN} = \boldsymbol{relative\ CN} \times \mathbf{2}$$

The median distance between K-means clusters of sample-specific CN segments was used to identify cut-offs for low-level CNAs (CN loss [CN ≤ 1.74], CN gain [CN ≥ 2.27]) and high-level CNAs (deletion [CN ≤ 1.33], amplification [CN ≥ 2.96]). Low-level cutoffs enabled the identification of subclonal CNAs, whereas high-level cutoffs were used to limit analysis to clonal and/or CNAs with multiple amplification or deletion events.

GISTIC2 (32) was used to identify focal regions and chromosome arms with significantly increased frequencies of alterations ("SARs", FDR ≤ 0.10) after providing all CN segments as input. Focal regions are defined by GISTIC2 as those containing a high prevalence of CNAs smaller than 98% of the length of the chromosome on which they reside. SARs containing sub-regions that were significantly altered in both primary and metastatic tumor cohorts were combined into 1 distinct SAR spanning both identified SARs.

Multi-region sequencing and representative pairs

MRS was performed on tumors to evaluate variability resulting from intratumor heterogeneity and technical sources. To accomplish this, WES and sWGS were performed on DNA samples extracted from multiple tissue blocks from the same tumor. In the WES cohort, 3 regions were sequenced in 2 metastatic tumors, and 2 regions were sequenced in 1 primary and 5 metastatic tumors. In the sWGS cohort, 3 regions were sequenced in 1 metastatic tumor, and 2 regions were sequenced in 1 primary and 6 metastatic tumors. For 4 patients, 2 metastases from different time-points were assayed. The majority of statistical analyses assessed 1 DNA sample

per tumor. Representative primary and metastatic tumor blocks were chosen to optimize the similarity of preservation type and estimated tumor cellularity between matched primary and metastatic tumors. Of these, 100% of primary tumors and 71% of metastatic tumors were preserved in FFPE.

Genomic concordance

Genomic concordance based on somatic coding mutations was determined by percent-overlap (Jaccard index) of tier I somatic coding-mutations in a pairwise fashion. Only loci that were covered by at least 30 reads by WES in both tumors were considered. Genome-wide CN concordance was determined by measuring the correlation (R^2) of normalized CN across 15 kbp-wide bins across the genome (n = 176,901 bins). CN values < 0.50 or > 8.00 were capped at those values for this analysis.

External datasets

To compare the frequency of genomic events between tumors in our cohort and tumors from independent external cohorts, we processed raw sequencing data from 1,043 TCGA-BRCA tumors (2) using the same alignment and variant calling pipeline used for our tumor cohort. Input fastq files were extracted using biobambam (https://github.com/gt1/biobambam) from pre-aligned BAM files downloaded from the NIH GDC data portal (https://gdc.cancer.gov/, date:10/20/2017). Metadata for the full list of TCGA-BRCA samples included in this analysis are included in Supplemental Table 1. Since the number of called variants decreases with larger sized PONs, a representative PON was generated that was analogous to the one used for variant calling in tumors in our study. First, 20 different PONs were generated using normal and artifact

calls from 68 randomly selected normal samples from TCGA-BRCA. The representative PON was created using variant and artifacts calls present in at least 9 of 20 sub-selected PONs. Variants in this representative PON were removed from TCGA-BRCA.

In addition, variant calls reported by TCGA from 775 primary tumors were downloaded from the NIH GDC Legacy Archive (Level_2.5.3.0) and were re-annotated by snpEff with the same version and annotation database configuration used for our data. When using established variant thresholds employed by TCGA (2), the number of mutations per tumor sample was highly correlated between the variant calling pipelines used in our analysis and by TCGA (R^2 = 0.98, $P \leq$ 2.2E-16). In an analogous manner, mutation calls from 211 metastatic tumors (10) were also re-annotated using snpEff and are referred to as the Lefebvre et al. dataset throughout the manuscript.

SMG frequency comparison

In order to determine whether metastases are enriched for mutations in SMGs compared with primary tumors, mutational frequencies were compared with those from the TCGA-BRCA primary tumor dataset (n = 1,042) using mutations called from our variant pipeline using stringent filtering criteria (>30x coverage, VAF ≥ 0.10, AAC ≥ 6) and criteria established by TCGA (2) (>8x coverage, VAF ≥ 0.10, AAC ≥ 2) using a 2-sided Fisher's exact test. To evaluate whether differences in SMG frequencies were dependent on the per-gene read coverage in TCGA-BRCA samples, separate additional analyses were iteratively limited to the TCGA-BRCA samples with highest coverage for a gene of interest (Supplemental Figure 3).

Pathway analysis

Gene sets were downloaded from the KEGG pathway database (n = 320 gene sets) (33). Seventy-three gene sets (the last set of gene sets listed in the database) were removed from analysis because they did not refer to specific cellular pathways, but to genes involved with specific diseases (*e.g.*, Type II Diabetes Mellitus, Prion Diseases). Twelve additional gene sets were removed that contained fewer than 10 genes.

SMPs were identified within primary and metastatic tumors using an implementation of the approaches described in POGSEA (34) which compares the observed number of tumors with at least 1 mutation in a gene set of interest to the number of mutant tumors identified from 10,000 randomly chosen sets of genes with the same size as the gene set of interest. Genes were sampled from those that are included in at least 1 gene set in KEGG (n = 6,509 genes). The Benjamini-Hochberg method was then used to correct for multiple testing and identify the final set of SMPs (FDR $\leq$ 0.10).

To identify pathways that were preferentially mutated in metastases compared with primary tumors, mutations shared between paired tumors (tier I in 1 tumor and tier I-II in the paired tumor) were removed and frequencies of private mutations within each pathway were compared between tumors from the same patient using a McNemar test. Limiting to private mutations allows the analysis to focus on metastatic tumor mutations that were either acquired and selected following metastatic dissemination or were selected during dissemination from small, undetectable populations of cells within the primary tumor. To address the concern that an increased mutation frequency in a given pathway gene set may simply be a consequence of a generally increased number of mutations in metastases, the test statistic generated from the applied statistical method was compared with a distribution of test statistics generated from

10,000 equally-sized, randomly generated gene sets in order to compute a permutation *P*-value (P_{perm}). Significant metastasis-enriched SMPs were defined as those with $P \leq 0.05$, FDR ≤ 0.10, and $P_{perm} \leq 0.10$.

A similar approach was used to compare mutation calls generated by our pipeline for 1,042 TCGA-BRCA breast primary tumors and calls reported by Lefebvre et al. (10). In this approach, TCGA-BRCA mutations were restricted to those satisfying filtering criteria used in Lefebvre et al. (coverage ≥ 10, VAF ≥ 0.10, AAC ≥ 5) (10). Since shared mutations cannot be removed from external, unpaired datasets, *PIK3CA* and *TP53* mutations were excluded since they were typically shared between primary and metastatic tumors in our cohort and because they occur at high frequency (>30% in both primary tumors and metastases). Frequencies were then compared using a 1-sided Fisher's exact test and permutation, with significance being defined as $P \leq 0.05$, FDR ≤ 0.10, and $P_{perm} \leq 0.10$.

Kaplan-Meier curve survival analysis

RFS curves were generated using Kaplan-Meier estimates and compared using a log-rank test to determine if RFS in TCGA-BRCA patients was associated with primary tumor mutations for each SMG. Time was defined by period between diagnosis and recurrence, measured in months ("Disease_Free_months"), events were defined by whether the patient recurred ("Disease_Free_Status"), and strata were defined by mutant status. Re-called TCGA-BRCA variants that passed filtering criteria established in (2) were included in the analysis.

For metastasis-enriched SARs, DSS curves were generated using Kaplan-Meier estimates and compared using a log-rank test to determine whether each SAR was associated with shortened DSS in METABRIC patients.

Immunohistochemistry

H & E stained slides were reviewed and representative sections chosen from each case for immunohistochemical staining (IHC) for 4 markers (Supplemental Methods Table 1). A positive control and negative rabbit or mouse isotype control (Leica Biosystems Inc, Buffalo Grove, IL, Cat. PA0777 and PA0996) were included in each staining run. Prior to staining, heat-induced epitope retrieval was performed for 20 min with Leica Bond™ Epitope Retrieval Solution 1 (Leica, Cat. AR9961). IHC was then performed on 5μm FFPE or OCT sections on a Leica Bond III auto-stainer for 15 min according to the manufacturer's instructions. Slides were arranged to eliminate covariates between staining (row, column, and run batch effects) and tumor type. Detection was carried out with the Leica Bond™ Polymer Refine Detection System (DS9800) with post-primary x 8 min and HRP polymer x 8 min, and Diaminobenzidine (DAB) x 10 min. Prior to staining OCT sections, slides were removed from -80°C storage and allowed to air dry for 30 min to remove moisture, then rinsed in 1X Leica wash buffer (Cat. AR9590) twice for 2 min each, fixed in neutral buffered formalin for 10 min and rinsed in 1X Leica Wash Buffer. Without allowing slides to dry, slides were loaded onto the Leica Bond III for retrieval and staining. Isotype controls, which were performed for each stained section, did not show antibody staining in any case.

Supplemental Methods Table 1

Antibody	Short Name	Source	Antigen Dilution
Phospho-S6 Ser235/236 (clone D57.2.2E) rabbit mAb (cat. #4858)	p-S6	Cell Signaling, Danvers, MA	1:40
Phospho-Rb Ser807/811 (clone D20B12) rabbit mAb (cat. #8516)	phospho-RB	Cell Signaling, Danvers, MA	1:50
β-catenin (clone 14) mouse mAb (cat. #610154)	β-catenin	BD Biosciences, San Jose, CA	1:250
Anti-PKA alpha/beta/gamma (catalytic subunit) phospho T197 (clone EP2606Y) rabbit mAb (cat. #ab75991)	p-PKA	Abcam, Cambridge, MA	1:4000

p-S6 was stained in samples from 2 cohorts: paired primary and metastatic tumors that arose within the same patient (15 patients), and matched primary and metastatic tumors that arose in separate patients, wherein primary tumors were matched by tumor block age and preservation type of the metastasis and for the receptor subtype of the primary tumor from which the metastasis arose (16 matched sets, Figure 5), with the exception of 2 primary tumors for which receptor status was unavailable. In total, p-S6 staining was performed on 31 metastatic tumors from liver (n = 19), distant lymph nodes (n = 6), brain (n = 3), and soft tissue (n = 2), with receptor subtypes: $HR^+/HER2^-$ (n = 18), $HR^+/HER2^+$ (n = 4), $HR^-/HER2^-$ (n = 5), $HR^-/HER2^+$ (n = 2), and unknown (n = 2). Phospho-RB was stained in a subset of samples from the paired (n = 8) and matched (n = 15) tumor cohorts. In total, phospho-RB staining was performed on 23 metastatic tumors from liver (n = 16), distant lymph nodes (n = 5), and brain (n = 2), with receptor subtypes: $HR^+/HER2^-$ (n = 14), $HR^+/HER2^+$ (n = 2), $HR^-/HER2^-$ (n = 3), $HR^-/HER2^+$ (n = 2), and unknown (n = 2). β-catenin and p-PKA were stained in a subset of samples from the paired (n = 7) and matched (n = 8) tumor cohorts. In total, β-catenin and p-PKA staining were each performed on 15 metastatic tumors from liver (n = 11), distant lymph nodes (n = 2), and brain (n = 2), with receptor subtypes: $HR^+/HER2^-$ (n = 10), $HR^+/HER2^+$ (n = 1), $HR^-/HER2^-$ (n = 2), $HR^-/HER2^+$ (n = 1) and unknown (n = 1).

Slides were separately scored for staining intensity in invasive tumor and adjacent non-tumor compartments, each of which was further subdivided into epithelial and stromal sub-compartments. Within each sub-compartment, the proportion of cells and/or nuclei that were $marker^+$ was estimated using the following scoring schema: 0 = no signal, 1 = low signal, 2 = moderate signal, 3 = high signal. Mean staining intensities in each sub-compartment were calculated using the percent of positive cells as weights. FFPE material was used in the majority

of cases, except for 7 metastases in the paired tumor cohort. No significant difference was detected in mean p-S6 intensity between metastases preserved in OCT (mean = 0.85) and those preserved in FFPE (mean = 1.3, $P = 0.55$). Scoring of phospho-RB within non-tumor compartments could not be performed for the majority of non-liver metastatic tumor samples and were excluded from analysis. Scoring for β-catenin and p-PKA were limited to the tumor epithelium.

Associations between genomic and clinical attributes

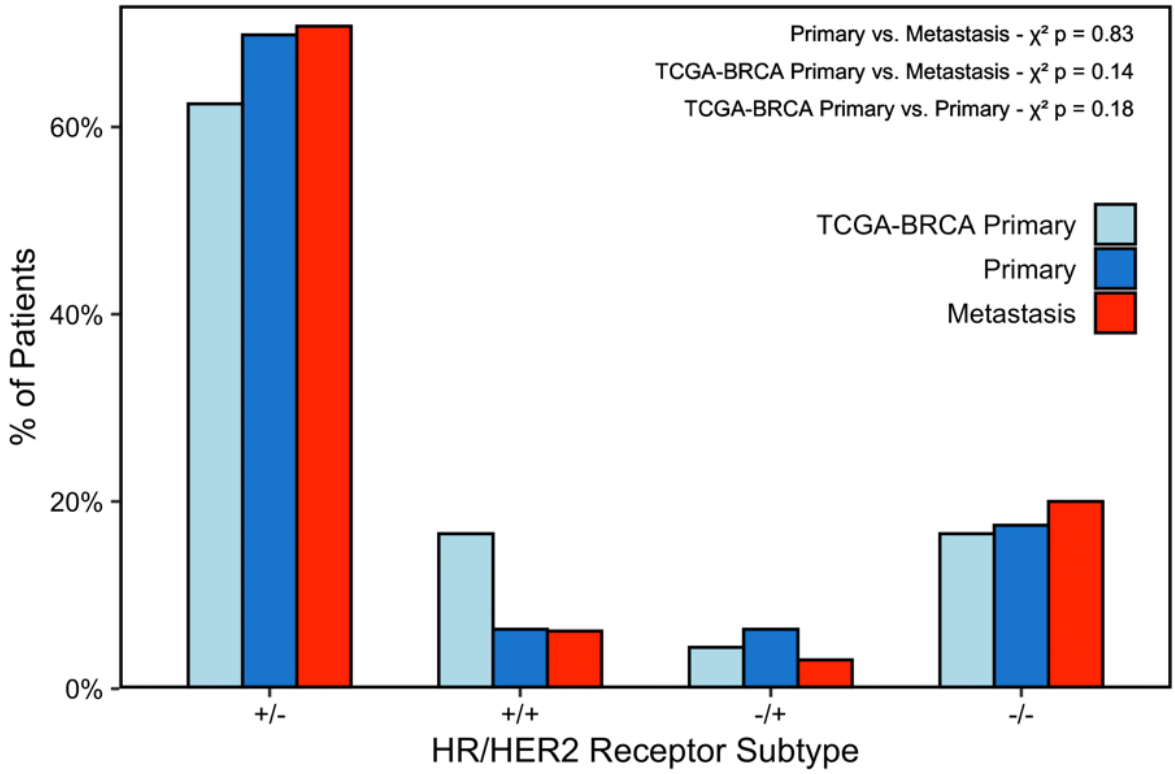
Six distinct analyses were performed to determine the level of co-occurrence between metastasis-specific SMGs, CNAs, and pathways across receptor subtypes, and to evaluate the level of association of genomic features with IHC staining and metastatic tumor site. The first analysis compared the co-occurrence of mutations between all SMGs and *ERBB2* with receptor subtype in the entire metastatic tumor cohort. The second analysis tested the association within the entire metastatic tumor cohort between metastasis-enriched CNAs (*STK11*, *CDKN2A*, *PTK6*, *PAQR8*) both via low-level CN dichotomized events and continuous CN values, mutations in metastasis-enriched SMPs, and receptor subtypes. The third analysis was the same as the second but was limited to paired primary and metastatic tumors. The fourth analysis investigated enrichment of metastatic tumor site with receptor subtype, as well as mutations and alterations within SMGs, SMPs, and SARs. The fifth analysis evaluated the association of p-S6 staining intensity and positivity with genomic features identified in this study. Only features that were present in at least 3 affected metastases were included: mutations in SMGs (*PIK3CA*, *TP53*, *KMT2C*, *RUNX1*, *MYLK*, *PEAK1*, *EVC2*) and therapeutic targets (*ESR1*, *ERBB2*), CNAs (loss of *TP53*, *PTEN*, *RB1*, *CDKN2A*, and *STK11*; gain of *CCNE1*, *CDK2*, *CDK4*, *CDK6*, and *PTK6*), as

well as whether metastases harbored ≥ 1 or ≥ 2 non-*PIK3CA* mutations in the mTOR Signaling pathway. The sixth analysis, which sought to identify covariates of phospho-RB signaling, used the same features as the fifth analysis, but also included the number of pRB-inactivating CNAs and mutations as features. In the first 4 analyses, Fisher's exact tests and Wilcoxon rank-sum tests were 2-sided. The fifth and sixth analyses used Wilcoxon rank-sum 1-sided tests. For each of these analyses, multiple test correction was implemented separately, with FDR values provided in the text.

Two additional analyses tested specific hypotheses and were not amenable to multiple test correction. The first tested an association of *PAQR8* gain with *ESR1/PGR* mutations in endocrine treated tumors. The second tested an association of high-level CNAs in the mTOR pathway (deletion of *STK11* and amplification of *PTK6*) with high-level CNAs in the CDK/RB pathway that are (1) predicted to inactivate pRB and (2) exhibited some level of enrichment in the total metastatic tumor cohort compared with the primary tumor cohort *(P* < 0.20, deletion of *CDKN2A* and amplifications of *CCNE1*, *CDK2*, *CDK4*, *CDK6*). Both tests used 2-sided Fisher's exact tests.

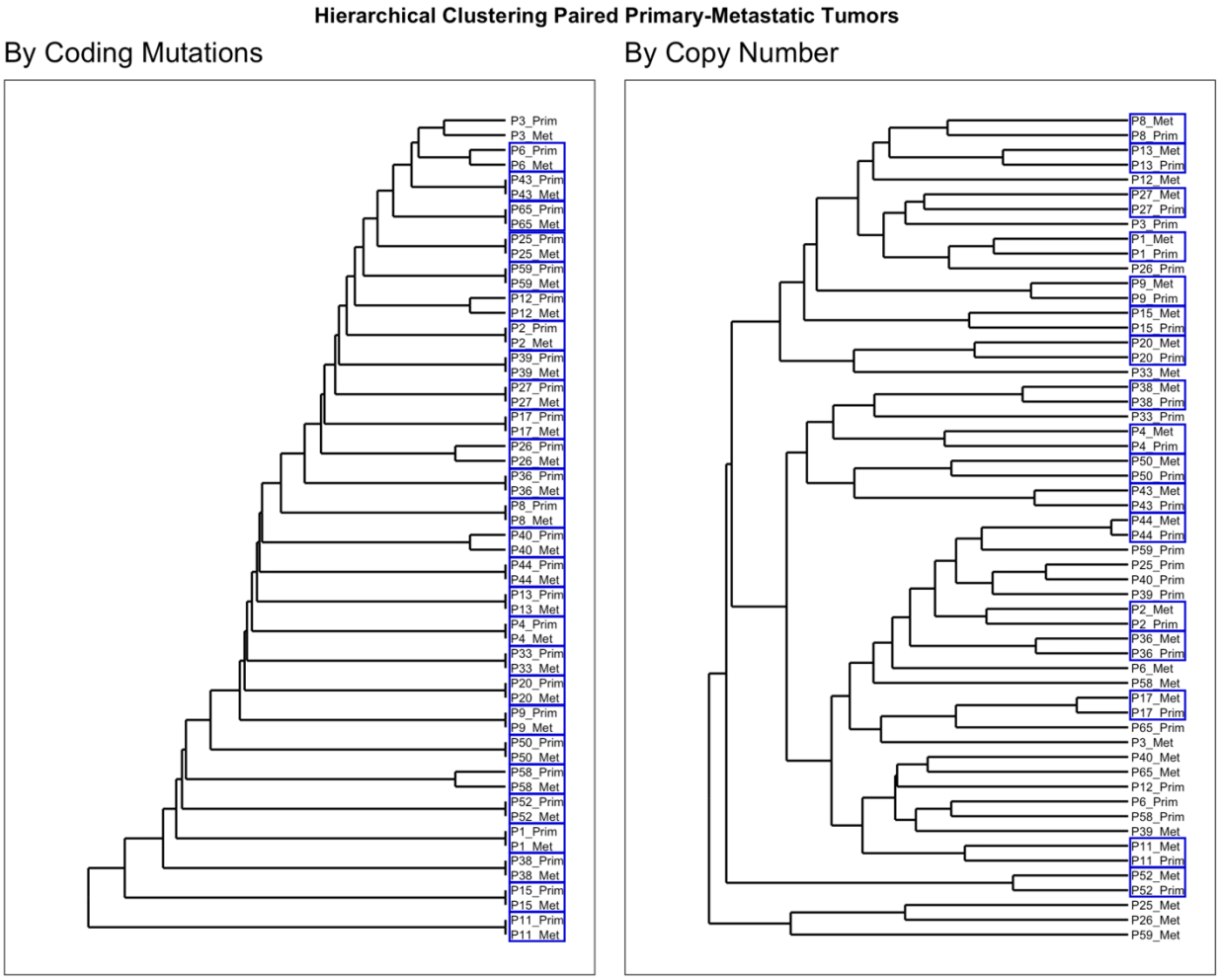
Supplemental Figures

Supplemental Figure 1: Global distribution of HR/HER2 receptor subtype is not significantly different between cohorts in the present study and TCGA-BRCA.



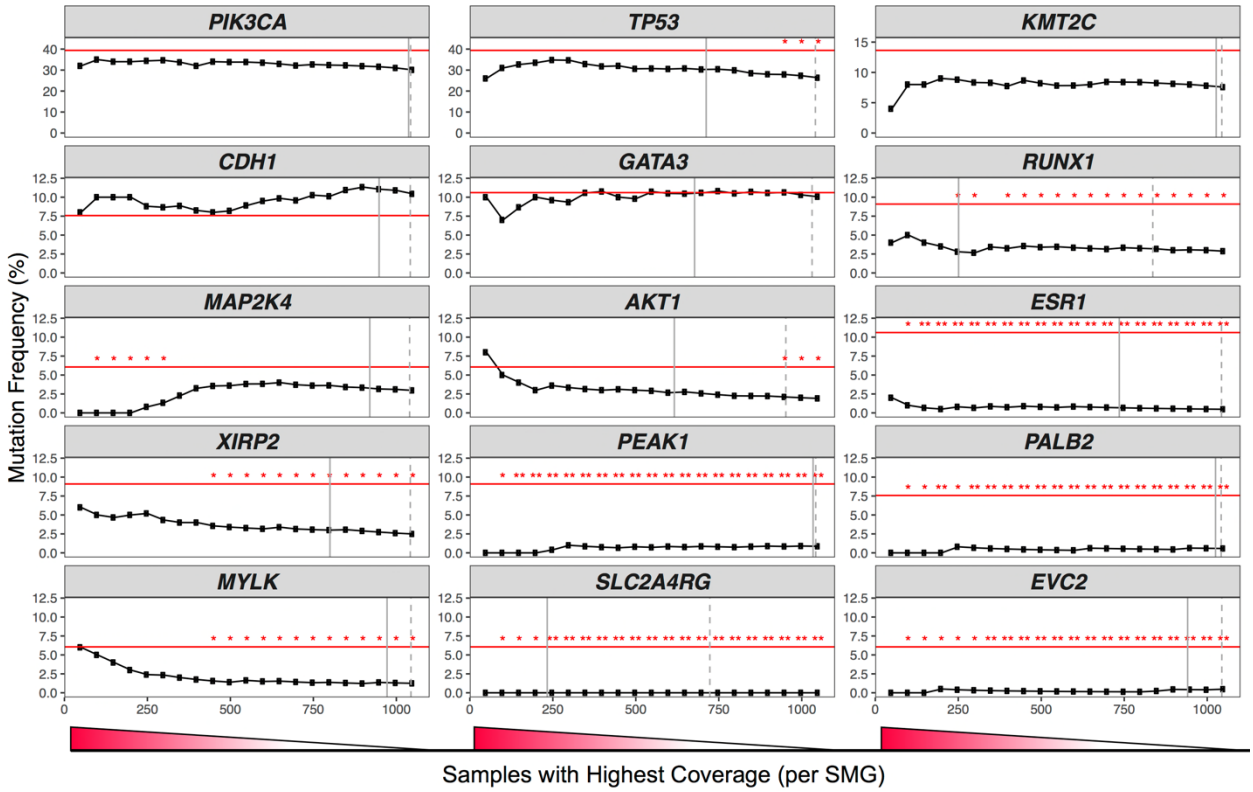
Bar-plots show the % of HR/HER2 subtypes in paired primary tumors (dark blue, $n = 63$), metastatic tumors (red, $n = 66$), and TCGA-BRCA primary tumors (light blue, $n = 906$) with complete receptor subtype information. χ^2 *P*-values indicate no significance difference in receptor subtype distributions between the 3 cohorts.

Supplemental Figure 2: Primary-metastasis tumor pairs cluster by recurrent mutations and genome-wide CN.



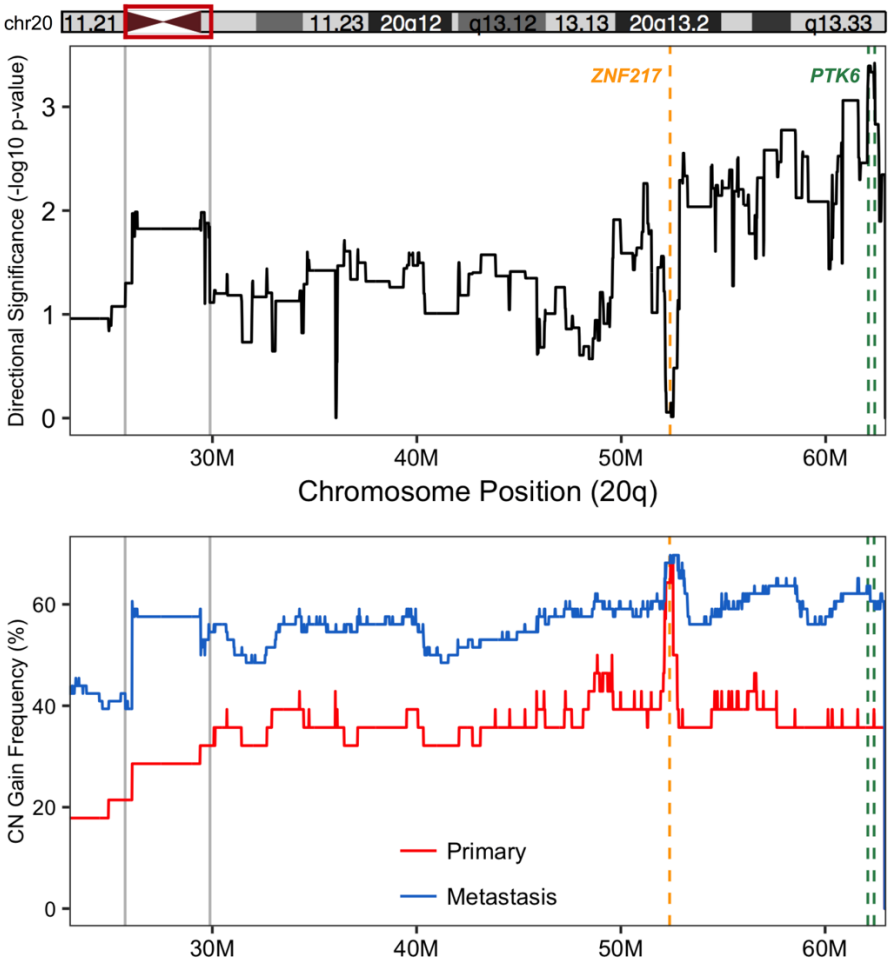
Dendrograms showing hierarchical clustering of all tumors in the paired tumor cohort using recurrent mutations (left) or genome-wide CN similarity across the top 50% most variable genomic bins (right). Tumors that co-cluster with their paired tumor are indicated by blue boxes. 27 out of 28 primary-metastatic tumor pairs cluster by shared mutations. Despite tumors from patient P3 appearing to co-cluster, they do not share mutations with each other. A majority of tumor pairs co-clustered by genome-wide CN similarity.

Supplemental Figure 3: Identification of metastasis-specific SMGs is largely independent of sequencing coverage in TCGA-BRCA.



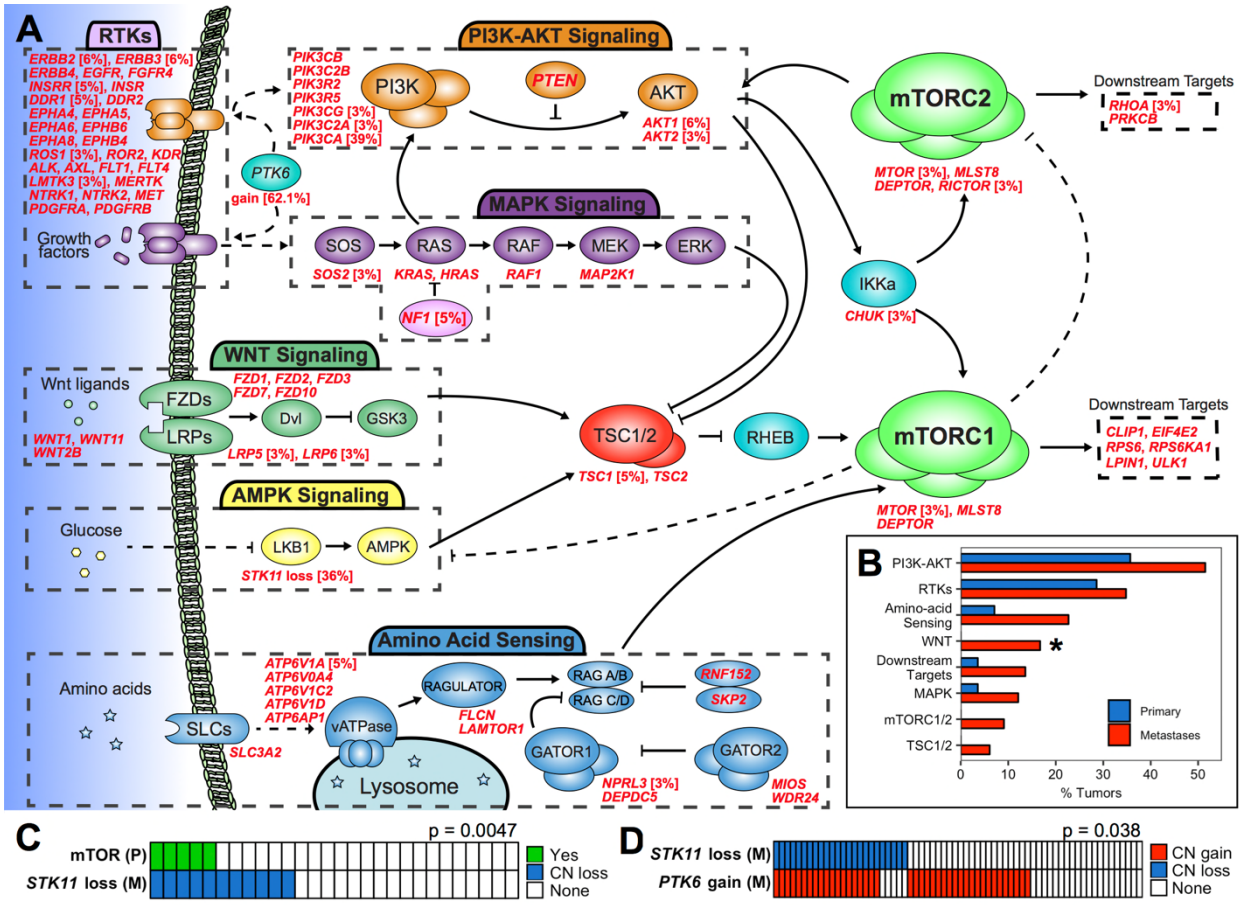
The degree to which SMG mutation frequencies are significantly different between metastases in the present cohort and TCGA-BRCA primary tumors, as shown across iteratively smaller subsets of TCGA-BRCA primary tumors with the highest coverage for each gene in question. Red horizontal lines indicate mutation frequencies for metastases in the present cohort. Black dotted lines indicate mutation frequencies across increasingly larger subsets of TCGA-BRCA primary tumors using high-confidence criteria. TCGA samples are ordered by decreasing sequencing coverage, with sample sets left of the grey solid and dashed lines having at least half of the gene coding region covered by at least 30 and 10 reads, respectively. Asterisks denote level of significance (2-sided Fisher's exact test, ** FDR<0.01; * FDR<0.10).

Supplemental Figure 4: Inference of *PTK6* as a driver of 20q amplification.



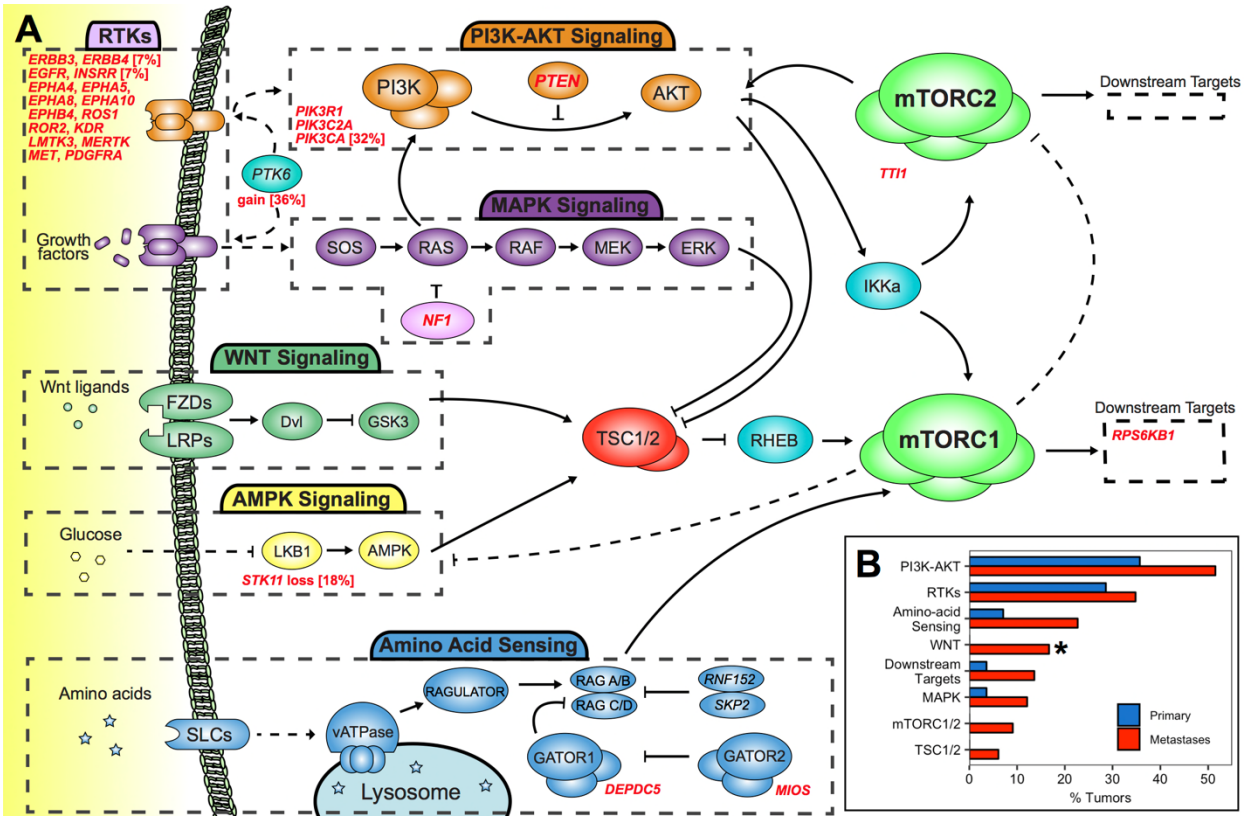
(Top) Degree to which the difference in normalized CN between paired primary and metastatic tumors is significantly different across chr20q (2-sided Wilcoxon signed-rank test). (Bottom) Frequencies of CN gain in all assayed primary and metastatic tumors. The metastasis-enriched SAR, 20p11.1-q11.21, is indicated by grey lines and tightly corresponds to the centromere of chr20 (red box above). Dashed green lines indicate the region with peak significance difference between paired primary tumors and metastases, which contains *PTK6* and *ZBTB46*. Dashed orange line indicates a region containing *ZNF217* where the frequencies of CN gain in both primary and metastatic tumors were the highest and most similar to each other.

Supplemental Figure 5: Mutations in the mTOR pathway in metastases.



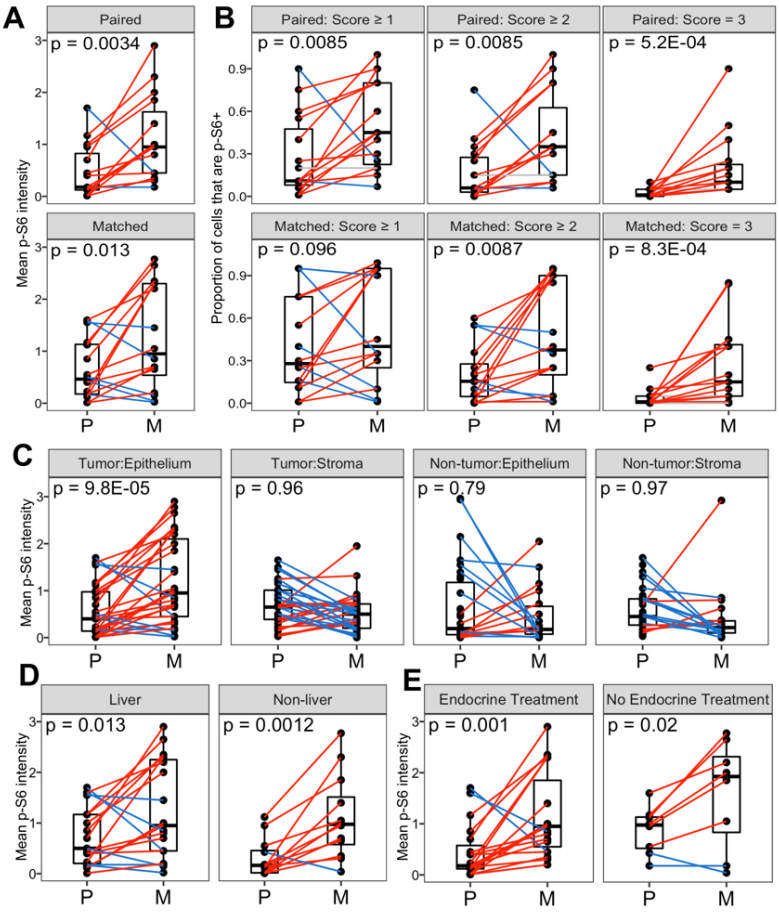
(A) Mutations in the mTOR pathway in metastatic tumors occur within multiple signaling modules. Gene names in red indicate genes mutated in ≥1 metastases (1.5%), with frequencies provided for genes mutated >1. Frequencies of *PTK6* CN gain and *STK11* CN loss are indicated. Mutations in the mTOR pathway that occurred in primary tumors are provided in Supplemental Figure 6. (B) Mutation frequencies of each module in primary (blue) and metastatic (red) tumors. Metastases exhibited a significantly higher mutation frequency of the WNT signaling module compared with primary tumors ($P = 0.010$). (C) Co-occurrence analysis showing that primary tumors (P) with non-*PIK3CA* mutations in the mTOR Signaling pathway gene set were more likely to give rise to metastases (M) with loss of *STK11*. (D) Co-occurrence analysis showing that *PTK6* gain and *STK11* loss significantly co-occur in the total metastatic tumor cohort.

Supplemental Figure 6: Fewer non-*PIK3CA* mTOR pathway mutations in primary tumors.



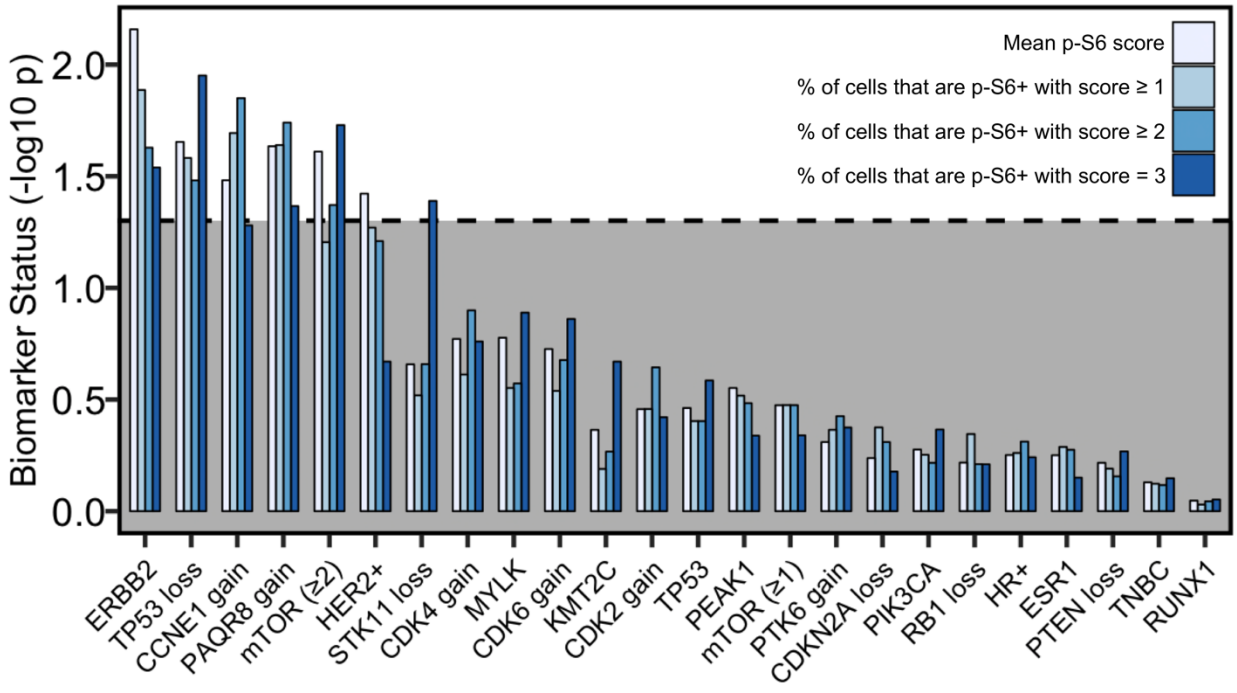
(A) Non-*PIK3CA* mutations in the mTOR pathway occur infrequently in primary tumors compared with metastases. Gene names in red indicate genes mutated at least once in primary tumors (4%), with frequencies provided for genes mutated more than once. Frequencies of CN gain of *PTK6* and CN loss of *STK11* are also indicated. (B) Frequencies of tumors with a mutation in each of the mTOR signaling modules, as also given in Figure 5. The majority of mTOR pathway mutations in primary tumors occur in RTKs and PI3K/AKT signaling components.

Supplemental Figure 7: mTOR pathway hyperactivation in sub-cohorts of metastases.



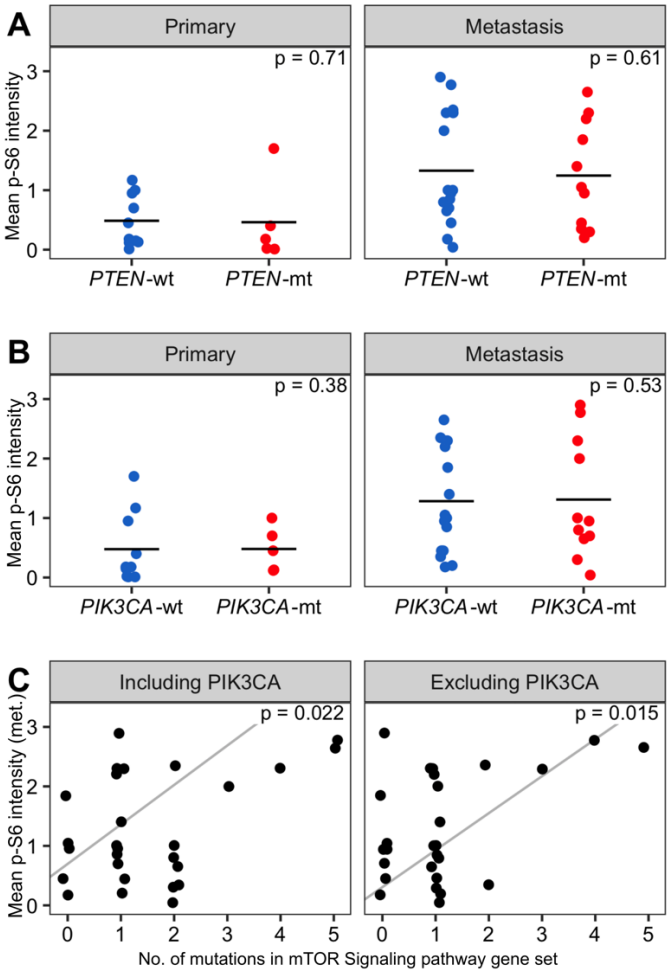
Differences in p-S6 staining intensity in primary and metastatic tumors as a function of tumor compartment, scoring metrics, and tumor or patient characteristics. Line colors indicate the change in p-S6 from primary tumors to their paired or matched metastases (red = increase, blue = decrease). *P*-values indicate the degree to which changes in p-S6 measures are concordant (1-sided Wilcoxon signed-rank test). (A) Mean p-S6 IHC staining intensity and (B) proportion of p-S6^{+} cells in primary-metastasis tumor pairs from individual patients (top), and primary and metastatic tumors matched by receptor status, tumor block age and fixation method (bottom). (C-E) Differences in mean p-S6 intensity within (C) epithelial and stromal compartments in tumor and non-tumor tissue, (D) liver and non-liver metastases, and (E) metastases in patients with, or without, exposure to endocrine therapy.

Supplemental Figure 8: Identification of p-S6 covariates.



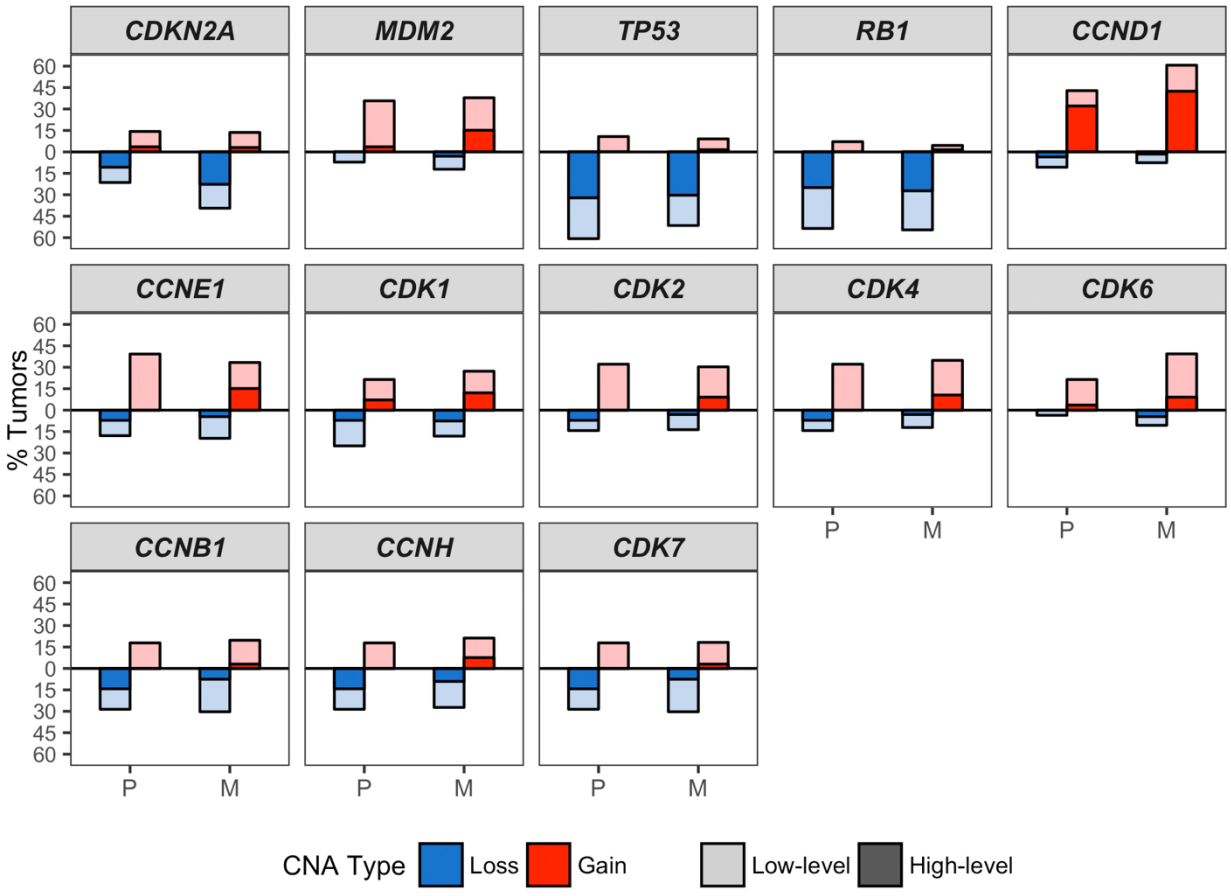
Evaluation of genomic features as biomarkers of mTOR activity (p-S6 staining intensity) in metastases (1-sided Wilcoxon rank-sum tests). Gene names refer to mutations, unless in reference to low-level CN gain or loss. "mTOR (≥X)" indicate the presence of at least X number of non-*PIK3CA* mutations in the mTOR Signaling pathway gene set. HR^+, $HER2^+$, and TNBC refer to receptor subtypes of assayed metastases. Manhattan plot indicating levels of significance for associated features with either mean p-S6 intensity or percent cells with p-S6 scores ≥1, ≥ 2, or = 3. The dash line indicates a *P*-value cutoff of 0.05.

Supplemental Figure 9: Loss of *PTEN* and mutations in *PIK3CA* are not associated with mTOR activity.



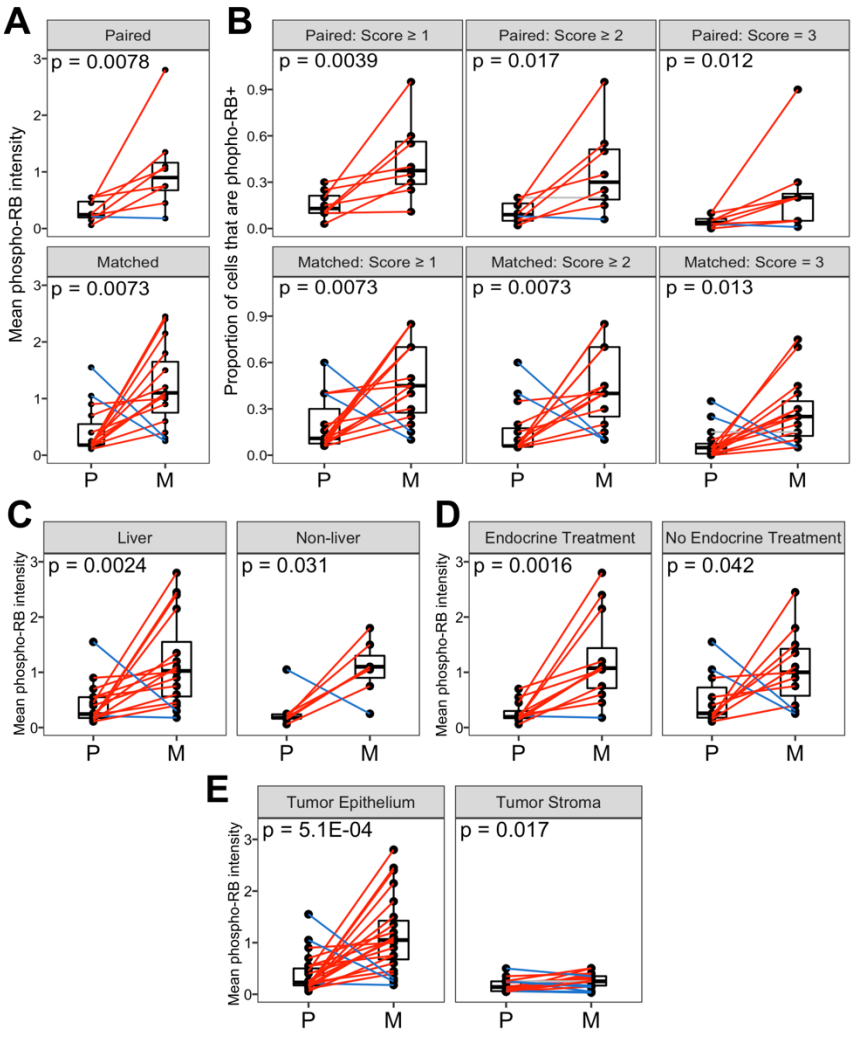
(A, B) Mean p-S6 staining intensity in primary (left) and metastatic (right) tumors for those mutant ("mt", red) or WT ("wt", blue) for (A) loss of *PTEN* and (B) mutation in *PIK3CA*. *P*-values indicate that mean p-S6 activity was not significantly different in tumors mutant for these genes (2-sided Wilcoxon rank-sum test). (C) Mean p-S6 staining in metastases as a function of the number of mTOR pathway mutations in each tumor. Mean p-S6 staining intensity in metastases was significantly correlated with the total number of mutations in the mTOR Signaling pathway gene set when including *PIK3CA* (left) or excluding *PIK3CA* (right) mutations from the analysis (univariate linear regression).

Supplemental Figure 10: Frequencies of low- and high-level CNAs in the CDK/RB pathway.



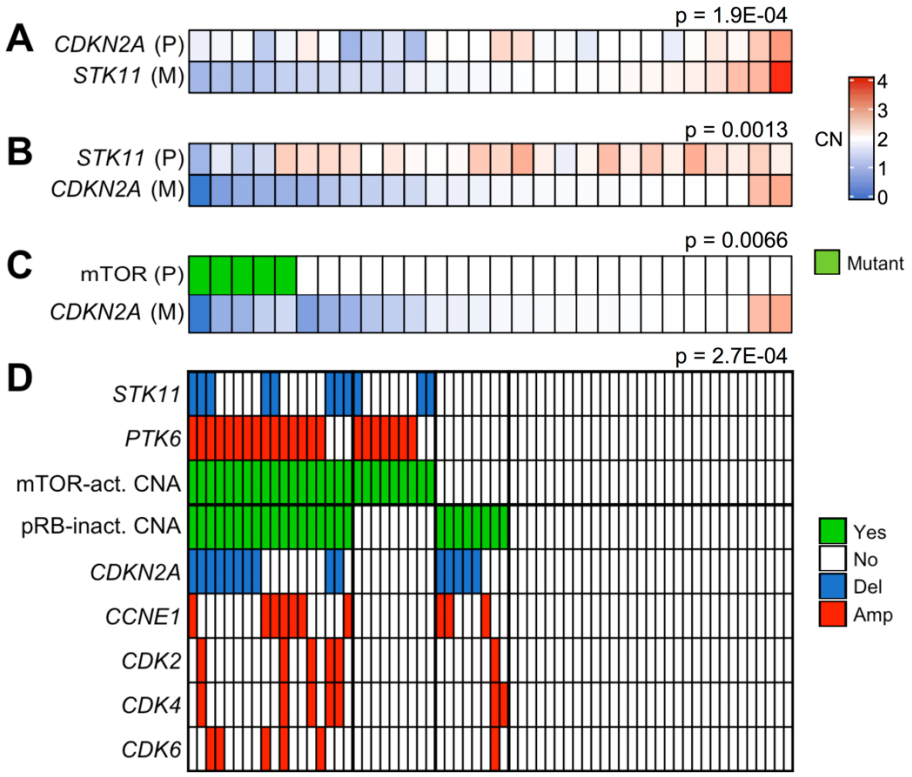
Frequencies of CN gain (red) and CN loss (blue) within primary and metastatic tumors are shown for genes downstream of $p16^{INK4A}$/$p14^{ARF}$ and/or upstream of pRB. Low-level CNAs (CN gain and loss) and high-level CNAs (amplification and deletion) are shown as lighter and darker-colored bars, respectively.

Supplemental Figure 11: Preferential inactivation of pRB in sub-cohorts of metastases.



(A) Mean phospho-RB IHC staining intensity and (B) proportion of phospho-RB^{+} cells in primary-metastatic tumor pairs arising in the same patient (top), or in different patients that were matched by receptor status, tumor block age and preservation type (bottom). (C-E) Differences in mean phospho-RB intensity in paired and matched (C) liver and non-liver metastases, (D) metastases in patients with, or without, exposure to endocrine therapy, and (E) epithelial and stromal compartments within tumors. Colored lines indicate the change in phospho-RB measures for primary tumors compared with their paired or matched metastases (red = increase, blue = decrease, 1-sided Wilcoxon signed-rank test).

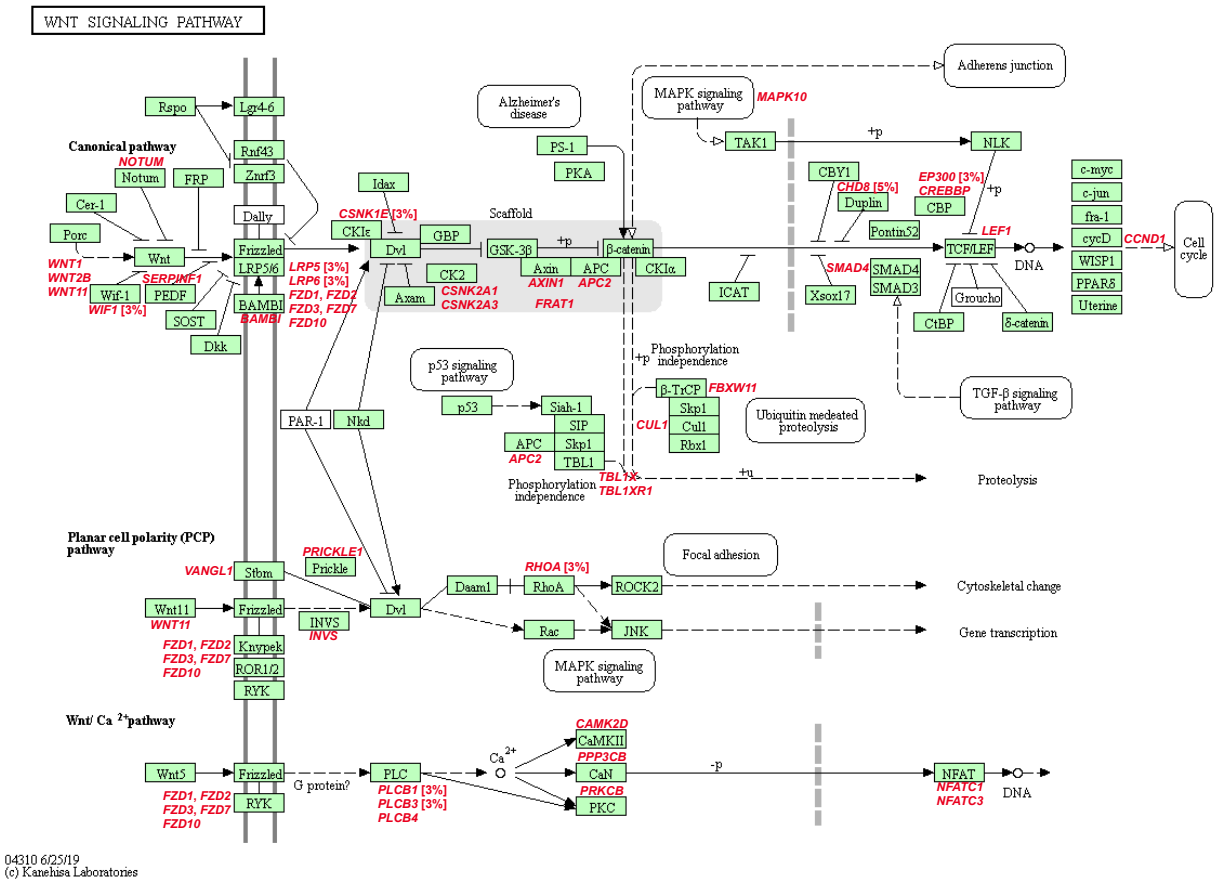
Supplemental Figure 12: Co-occurrence of alterations in the mTOR and CDK/RB pathways.



(A-C) Patients whose primary tumors exhibited genomic alterations in either the mTOR or CDK/RB pathways were more likely to give rise to metastases bearing genomic alterations in the other pathway. Each column represents a primary-metastasis tumor pair from 1 patient. (A, B) CN values (shaded bars) of *STK11* and *CDKN2A* in primary (P) and metastatic (M) tumors from the same patient ordered by CN values in the metastatic tumor. *P*-values indicate the degree of correlation (univariate linear regression). (C) Primary tumors that harbored a non-*PIK3CA* mutation in the mTOR Signaling pathway gene set (green, P) were more likely to give rise to metastatic tumors with lower values of *CDKN2A* (blue, M) (2-sided Wilcoxon's signed-rank test). (D) High-level CNAs that are predicted to activate mTOR or inactivate pRB are correlated within the total metastatic tumor cohort. Each column represents a metastasis from 1 patient.

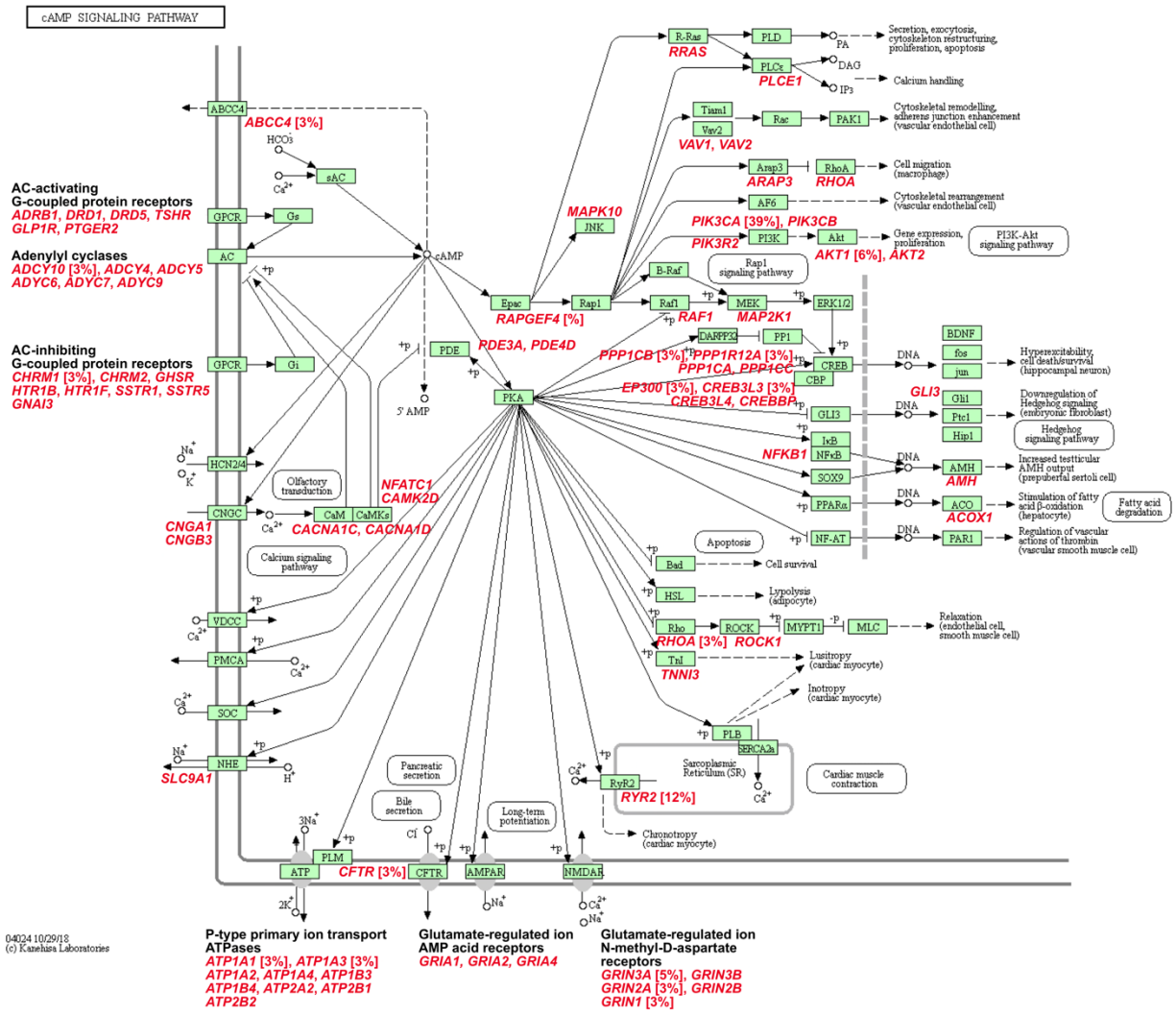
Blue bars indicate tumors with *STK11* deletion or *CDKN2A* deletion. Red bars indicate tumors with amplification of *PTK6*, *CCNE1*, *CDK2*, *CDK4* or *CDK6*. *P*-value indicates level of co-occurrence of at least 1 mTOR-activating or pRB-inactivating CNA in tumors (green) (2-sided Fisher's exact test).

Supplemental Figure 13: Mutations in the WNT Signaling pathway gene set in metastases.



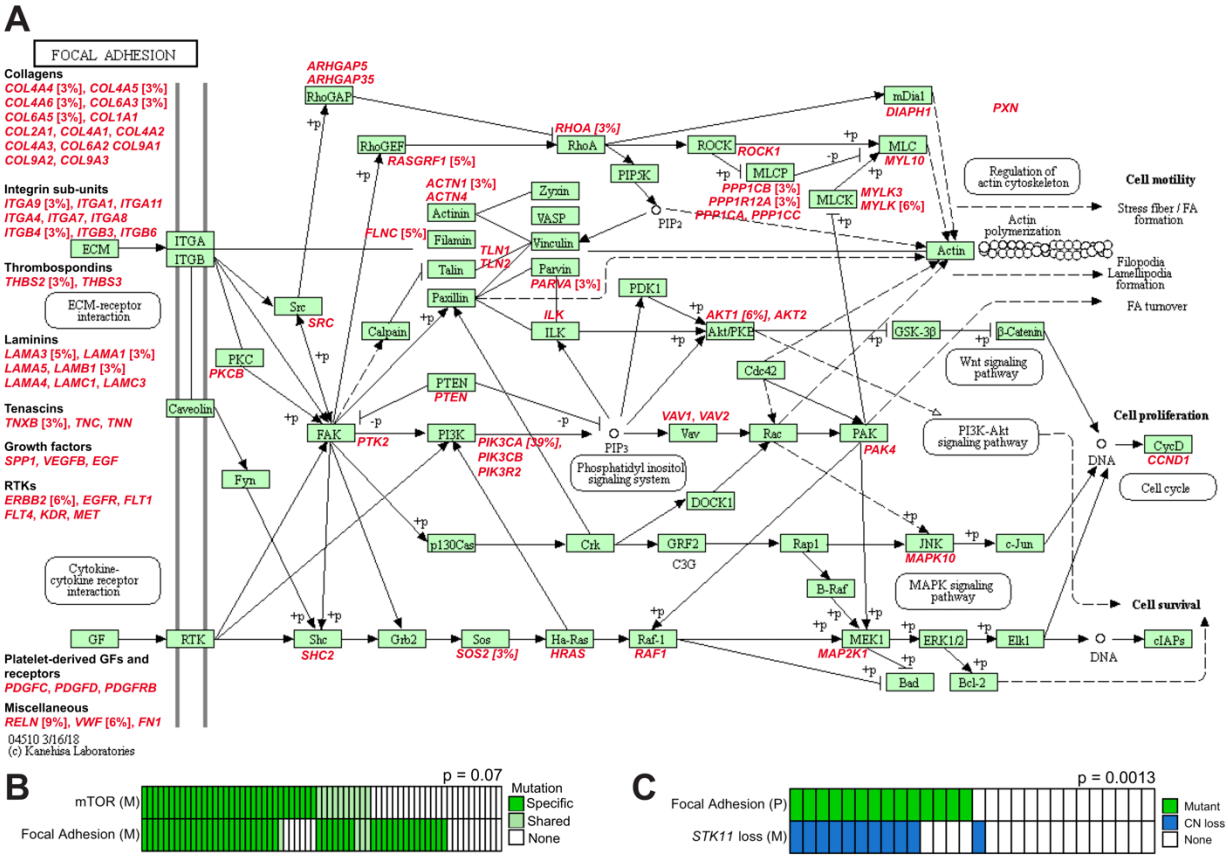
Annotated KEGG diagram indicating each gene mutated in the WNT Signaling pathway gene set within metastases (red). All mutation frequencies are 1.5%, unless otherwise indicated.

Supplemental Figure 14: Mutations in the cAMP Signaling pathway gene set in metastases.



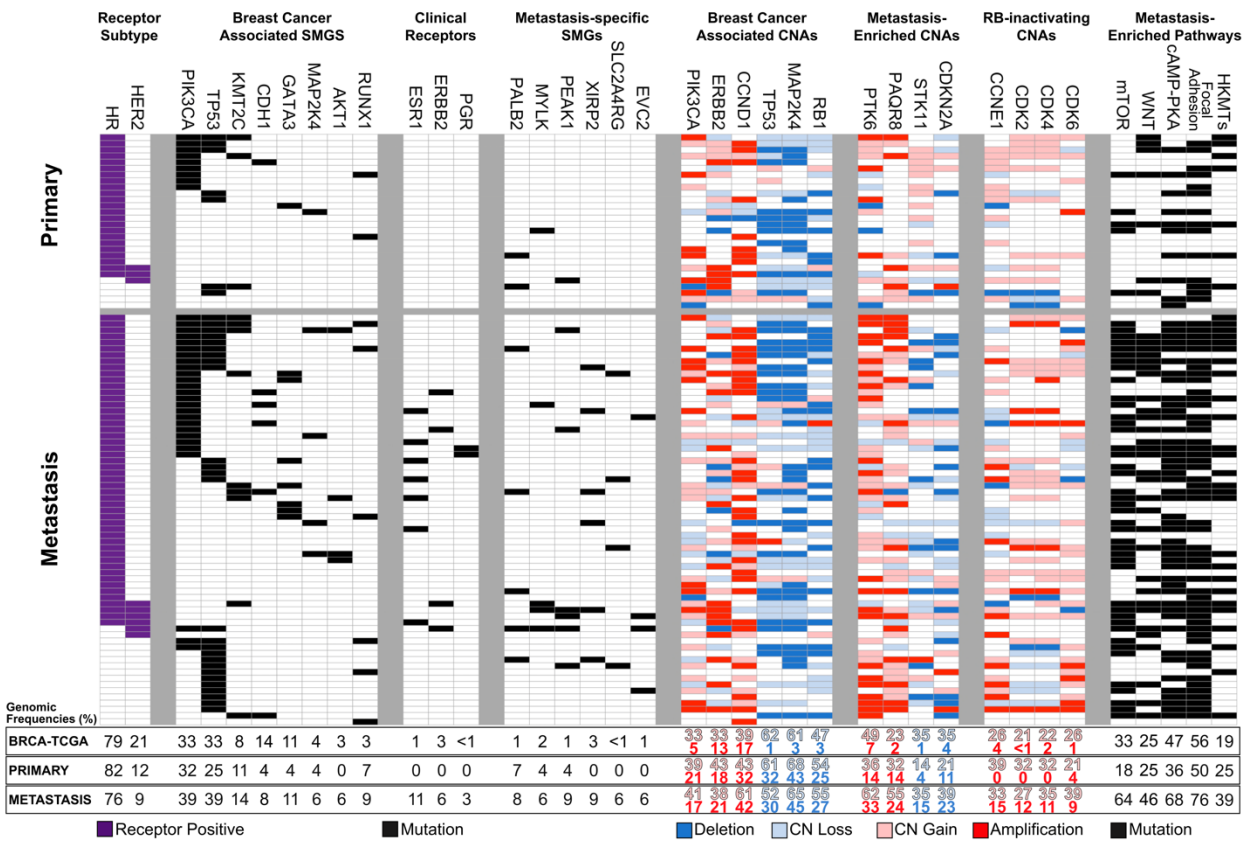
Annotated KEGG diagram indicating each gene mutated in the cAMP Signaling pathway gene set in metastases (red). All mutation frequencies are 1.5%, unless otherwise indicated.

Supplemental Figure 15: Mutations in the Focal Adhesion pathway gene set in metastases.



(A) Annotated KEGG diagram indicating each gene mutated in the Focal Adhesion pathway gene set in metastases (red). All mutation frequencies are 1.5%, unless otherwise indicated. (B, C) Co-occurrence of mutations and alterations in the Focal Adhesion and mTOR Signaling pathway gene sets. Bars indicate the presence of mutations or CNAs in paired primary tumors (P) and paired metastases (M). (B) A large proportion of metastases exhibit mutations specific to the mTOR Signaling pathway gene set and to the Focal Adhesion pathway gene set with trending co-occurrence (2-sided Fisher's exact test). Tumors that only harbor mutations in genes that are included in both pathways are in light green. (C) Primary tumors with a mutation in the Focal Adhesion pathway gene set were more likely to give rise to metastases with *STK11* loss (2-sided Fisher's exact test).

Supplemental Figure 16: Summary of genomic features implicated in metastases



Clinical and genomic features of primary tumors (top) and metastases (bottom). From left to right: receptor subtype; presence of mutations in primary tumor-associated SMGs, genes encoding ER, PR, and HER2, and metastasis-specific SMGs; presence of low and high-level CNAs in known 'driver' genes for primary breast cancer, metastasis-enriched CNAs, and additional genes whose encoded proteins both inactivate pRB and exhibit trending enrichment in metastases. Far-right: presence of non-*PIK3CA/TP53* mutations in SMPs and gene families (HKMTs) that are preferentially mutated in metastases. Frequencies of each feature are given at the bottom for TCGA-BRCA primary tumors as well as for primary and metastatic tumors in the present study. Mutation frequencies of SMGs, clinical receptors, and pathways within TCGA-BRCA were calculated based on mutations re-called using the same variant calling pipeline employed in our study and variant thresholds established in (2). Frequencies of low- and high-

level CNAs in TCGA-BRCA result from a GISTIC2 analysis reported by TCGA that used array competitive genomic hybridization assays and per-sample CNA level cut-offs.

Supplemental Tables

Supplemental Table 1: Sheet 1: Clinical annotation for each genomically assayed patient, including information on metastatic tumor site, receptor subtype, summary treatment information, and whether the paired primary tumor from the same patient was genomically assayed. Sheet 2: Clinical annotation provided by TCGA is given for the 1,044 primary tumors used as a reference set throughout the study.

Supplemental Table 2: Sheet 1: MutSigCV2 output used to identify the 15 SMGs identified in the total metastatic tumor cohort. Sheet 2: Mutations in clinically relevant receptors (*ESR1*, *ERBB2*, *PGR*). Sheet 3: Mutations in SMGs determined by MutSigCV2. Sheet 4: Comparison of SMG frequencies between metastatic tumors in our cohort and TCGA-BRCA primary tumors using low and high-confidence variant calling thresholds. Variants within the TCGA-BRCA data set were re-called using the same variant calling pipeline employed in this study.

Supplemental Table 3: Sheets 1-4: GISTIC2 results for genomically assayed primary and metastatic tumors. Sheet 5: Analysis of differences in normalized copy number and CNA frequencies in SARs identified by GISTIC2 between paired primary and metastatic tumors and between total tumor cohorts.

Supplemental Table 4: Sheets 1-3: SMPs identified in paired primary tumors, paired metastases, and all metastases, respectively. Sheet 4: Mutation frequencies and enrichment *P*-values for each pathway. Sheet 5: Information on mutations detected in metastasis-enriched SMPs.

Supplemental Table 5: Sheets 1 and 2: SMPs identified within the TCGA-BRCA primary tumor cohort and the Lefebvre et al. metastatic tumor cohort, respectively. Sheet 3: Mutation frequencies and enrichment *P*-values are given for each pathway.

Supplemental Table 6: The number of WT and mutant samples for each p-S6 biomarker within the paired primary-metastatic tumor cohort assayed by p-S6 IHC, as well as resulting *P*-values when comparing p-S6 intensities.

References

1. Popova T, Manie E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res.* 2012;72(21):5454-62.
2. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell.* 2015;163(2):506-19.
3. Rudolph M, Anzeneder T, Schulz A, Beckmann G, Byrne AT, Jeffers M, et al. AKT1 (E17K) mutation profiling in breast cancer: prevalence, concurrent oncogenic alterations, and blood-based detection. *BMC Cancer.* 2016;16:622.
4. Yi KH, and Lauring J. Recurrent AKT mutations in human cancers: functional consequences and effects on drug sensitivity. *Oncotarget.* 2016;7(4):4241-51.
5. Zheng Y, Peng M, Wang Z, Asara JM, and Tyner AL. Protein tyrosine kinase 6 directly phosphorylates AKT and promotes AKT activation in response to epidermal growth factor. *Mol Cell Biol.* 2010;30(17):4280-92.
6. Li X. Brk/PTK6 Sustains Activated EGFR Signaling through Inhibiting EGFR Degradation and Transactivating EGFR. *Oncogene.* 2012;31(40):4372-83.
7. Aubele M, Spears M, Ludyga N, Braselmann H, Feuchtinger A, Taylor KJ, et al. In situ quantification of HER2–protein tyrosine kinase 6 (PTK6) protein–protein complexes in paraffin sections from breast cancer tissues. *Br J Cancer.* 2010;103:663.
8. Kamalati T, Jolin HE, Fry MJ, and Crompton MR. Expression of the BRK tyrosine kinase in mammary epithelial cells enhances the coupling of EGF signalling to PI 3-kinase and Akt, via erbB3 phosphorylation. *Oncogene.* 2000;19(48):5471-6.
9. Bhat BM, Allen KM, Liu W, Graham J, Morales A, Anisowicz A, et al. Structure-based mutation analysis shows the importance of LRP5 beta-propeller 1 in modulating Dkk1-mediated inhibition of Wnt signaling. *Gene.* 2007;391(1-2):103-12.
10. Lefebvre C, Bachelot T, Filleron T, Pedrero M, Campone M, Soria JC, et al. Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS Med.* 2016;13(12):e1002201.
11. Shaw RJ. mTOR signaling: RAG GTPases transmit the amino acid signal. *Trends Biochem Sci.* 2008;33(12):565-8.
12. Zoncu R, Bar-Peled L, Efeyan A, Wang S, Sancak Y, and Sabatini DM. mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H(+)-ATPase. *Science.* 2011;334(6056):678-83.
13. Masui K, Tanaka K, Ikegami S, Villa GR, Yang H, Yong WH, et al. Glucose-dependent acetylation of Rictor promotes targeted cancer therapy resistance. *Proc Natl Acad Sci.* 2015;112(30):9406-11.
14. Cortes JR, Ambesi-Impiombato A, Couronne L, Quinn SA, Kim CS, da Silva Almeida AC, et al. RHOA G17V Induces T Follicular Helper Cell Specification and Promotes Lymphomagenesis. *Cancer Cell.* 2018;33(2):259-73.e7.
15. Zhao JH, Reiske H, and Guan JL. Regulation of the Cell Cycle by Focal Adhesion Kinase. *J Cell Biol.* 1998;143(7):1997-2008.
16. Damsky W, Micevic G, Meeth K, Muthusamy V, Curley DP, Santhanakrishnan M, et al. mTORC1 activation blocks BrafV600E-induced growth arrest but is insufficient for melanoma formation. *Cancer Cell.* 2015;27(1):41-56.
17. Huang J, Roth R, Heuser JE, and Sadler JE. Integrin α(v)β(3) on human endothelial cells binds von Willebrand factor strings under fluid shear stress. *Blood.* 2009;113(7):1589-97.
18. Kruse-Jarres R, and Johnsen JM. How I treat type 2B von Willebrand disease. *Blood.* 2018;131(12):1292-300.

19. Chin KV, Yang WL, Ravatn R, Kita T, Reitman E, Vettori D, et al. Reinventing the wheel of cyclic AMP: novel mechanisms of cAMP signaling. *Ann N Y Acad Sci.* 2002;968:49-64.
20. Bolger AM, Lohse M, and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-20.
21. Li H, and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-60.
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
23. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-8.
24. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012;22(3):568-76.
25. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* 2012;6(2):80-92.
26. Cao Z, Wang L, Chen Y, Cai R, Lu J, Yu Y, et al. VarfromPDB: An Automated and Integrated Tool to Mine Disease-Gene-Variant Relations from the Public Databases and Literature. *J Proteom Bioinform.* 2017;10.
27. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499(7457):214-8.
28. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic Acids Res.* 2018;46(D1):D754-D61.
29. Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* 2014;24(12):2022-32.
30. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
31. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol.* 2015;26(1):64-70.
32. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, and Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12(4):R41.
33. Kanehisa M, Furumichi M, Tanabe M, Sato Y, and Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353-d61.
34. Boca SM, Kinzler KW, Velculescu VE, Vogelstein B, and Parmigiani G. Patient-oriented gene set analysis for cancer mutation data. *Genome Biol.* 2010;11(11):R112.