**Expanded View Figures**

1. **Human primary tumors CNA**
   - CNA characterization by unsupervised PCA
   - Human Signature A: shared PC1-based CNA signature across 4 tumor types (Fig. 1; EV2-3; S1-3)

2. **In vivo & in vitro mouse models**
   - Characterization of cancer-like CNA (Fig. 2)

3. **In vitro experimental & genetic validation**
   - Normal diploid genome
   - Selective pressure
   - Genomic heterogeneity
   - Genomic instability
   - Cancer-like CNA
   - Characterization by unsupervised PCA (Fig. 4; S5-9)
   - Mouse Signature A: PC1-based and evolving CNA signatures
   - Mouse Signature B

Results:
- prediction of clinical FDG uptake by primary breast cancers (in vivo)
- prediction of glycolytic activity in breast cancer cell lines (in vitro) (Fig. 3; EV4; S4,10)

Analysis:
- cross-species consistent regions observed in additional tumor types
- cross-species consistent regions observed in mouse in vivo models and in vitro models
- enrichment of glycolysis genes in consistently altered regions (Fig. 2)

Conservation and enrichment of amplifications in glycolysis associated genes

Figure EV1. Integration of patient and model system CNA in a pan-cancer and cross-species bioinformatic and experimental analysis. Overall workflow for the analyses described in the Results section.
Figure EV2. Principal component analysis of TCGA copy number data reveals a shared CNA signature enriched for p53 mutation (related to Fig 1).

A–C Copy number profiles of 359 lung (LU) squamous carcinomas (LUSC) and 368 lung adenocarcinomas (LUAD) (A), 583 ovarian serous cystadenocarcinoma (OV) (B), and 492 uterine corpus endometrial carcinomas (UCEC) (C) tumors sorted by their PC1 score. Signature A tumors are sorted to the top of the heatmap. PC1-ranked tumors were analyzed for enrichment of known mutations or molecular subtypes (e.g., basal/luminal in breast, see D for all subtypes tested). The red horizontal bars to the right of each tumor type indicate the tumor subtype (e.g., LUAD), or the presence of the indicated mutation (e.g., TP53). Normalized enrichment score and permutation P-value for each significantly enriched mutation or clinical subtype are shown below the tick plots.

D Enrichment analysis of molecular subtypes in PC-defined signatures A and B. The molecular subtypes tested were defined by the TCGA (The Cancer Genome Atlas Research Network, 2011, 2012, 2013, 2014). Because LUAD and LUSC were combined into a single dataset LU, all LUSC and LUAD subtypes were associated with signatures A and B, respectively. Within LU, signature A was most associated with the classical subtype of LUSC, and signature B was most associated with the bronchioid subtype of LUAD. Permutation-based P-values are indicated.

E Amplification of MDM2 and deletion of CDKN2A are mutually exclusive in signature B human BRCA. Each point represents a single patient sample. Samples are colored red or blue according to their membership in signatures A or B, respectively, with intermediate colors representing intermediate PC1 scores between signatures A and B. Samples colored white represent nearly diploid tumors, as indicated by the triangle in Fig 1B.
Tumor CNAs reflect metabolic selection

Nicholas A Graham et al.

Figure EV2.
Figure EV3. Hierarchical clustering confirms the existence of the shared pan-cancer CNA signatures across multiple tumor types, and distinct signature subtypes within tumor types (related to Fig 1).

Pan-cancer clustering of 3,037 copy number profiles using a balanced, random sampling of tumors from 15 tumor types. The dendrogram reveals the presence of a cluster containing multiple tumor types and enriched in PCA-defined signature A tumors (hypergeometric P-value = 1.1 × 10^{-18} for concordance between the PCA and hierarchical clustering results). There is also a cluster enriched in the generally less uniform PCA-defined signature B tumors (hypergeometric P-value = 4.5 × 10^{-13}), which contains other tumor types as well (e.g., SKCM). In the key, the individual tumor types and pan-cancer PC1 scores are indicated. Additionally, for the core signature A tumors (OV, BRCA, UCEC, and LU), tumor-specific subtypes (e.g., BRCA basal and luminal) and tumor type-specific PC1 scores are indicated. The gene loci CNA levels in the heatmap are ordered based on their chromosomal locations (columns).
### Tumor CNAs reflect metabolic selection

**Nicholas A Graham et al**

Figure EV3.
Figure EV4. CNA-defined signature A tumors exhibit RNA expression signatures consistent with increased glycolytic activity (related to Fig 3).

A KEGG metabolism pathway enrichment analysis based on mRNA expression in BRCA and LU signature A and signature B tumors. RNA expression data are from BRCA and LU tumors with paired TCGA DNA and RNA data. To evaluate the similarity to BRCA tumors with high FDG-PET uptake, we included a gene set consisting of genes from the glycolysis and pentose phosphate pathway that were upregulated in FDG-high BRCA tumors (Palaskas et al, 2011). To combine BRCA and LU into one enrichment analysis, genes were ordered by their average rank in BRCA and LU tumor types. Enrichment scores for all KEGG metabolism pathways and for individual tumor types can be found in Table EV4.

B-D RNA expression signatures of high glycolysis (Palaskas et al, 2011) were used to predict the glycolytic phenotype of BRCA (B), LU (C), and OV (D) tumors with paired TCGA DNA and RNA data. UCEC tumors were excluded due to a lack of sufficient samples with paired RNA and DNA data. The color scale was normalized for each tumor type by the mean FDG score for tumors with integrated CNA score > 0.2. Non-normalized values are shown in the color legend and in (E).

E RNA-based WGV predictions of glycolysis show that BRCA and LU signature A tumors (top 10% PC1) are predicted to be more glycolytic than the tissue type-matched signature B tumors (bottom 10% PC1). Mann–Whitney U-test P-values are shown. Data are presented in box (median, first and third quartiles) and whisker (extreme value) plots.

F Spearman correlation values between RNA-based WGV glycolysis predictions and CNA-based PC1 values, calculated for different range windows of integrated CNA levels to control for the general increase in glycolysis predictions at increased levels of genomic instability. P-values shown are calculated by performing $1 \times 10^6$ permutations where the WGV scores were randomly shuffled, and the P-value was calculated as the percentage of random permutations that achieved a higher correlation value for any range window than did the experimentally observed relationship.
**KEGG metabolism pathway gene set (n = 76)**

<table>
<thead>
<tr>
<th>Pathway</th>
<th>NES</th>
<th>Permutation p-value</th>
<th>FDR q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDG-PET High BRCA Palaskas et al.</td>
<td>2.52</td>
<td>0.0002</td>
<td>0.0073</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>2.20</td>
<td>0.0010</td>
<td>0.0243</td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td>2.01</td>
<td>0.0051</td>
<td>0.0532</td>
</tr>
<tr>
<td><strong>Glycolysis &amp; Pentose phosphate pathway</strong></td>
<td><strong>1.96</strong></td>
<td><strong>0.0068</strong></td>
<td><strong>0.0621</strong></td>
</tr>
<tr>
<td>Glycosphingolipid biosynthesis - lacto and neolacto series</td>
<td>1.86</td>
<td>0.0107</td>
<td>0.0868</td>
</tr>
<tr>
<td>Cysteine and methionine metabolism</td>
<td>1.83</td>
<td>0.0145</td>
<td>0.0962</td>
</tr>
<tr>
<td><strong>Core glycolysis</strong></td>
<td><strong>1.80</strong></td>
<td><strong>0.0156</strong></td>
<td><strong>0.0915</strong></td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>1.73</td>
<td>0.0219</td>
<td>0.1230</td>
</tr>
</tbody>
</table>

**Figure EV4.**

**A.**

**KEGG metabolism pathway gene set (n = 76)**

**Average gene rank**

<table>
<thead>
<tr>
<th>Pathway</th>
<th>NES</th>
<th>Permutation p-value</th>
<th>FDR q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDG-PET High BRCA Palaskas et al.</td>
<td>2.52</td>
<td>0.0002</td>
<td>0.0073</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>2.20</td>
<td>0.0010</td>
<td>0.0243</td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td>2.01</td>
<td>0.0051</td>
<td>0.0532</td>
</tr>
<tr>
<td><strong>Glycolysis &amp; Pentose phosphate pathway</strong></td>
<td><strong>1.96</strong></td>
<td><strong>0.0068</strong></td>
<td><strong>0.0621</strong></td>
</tr>
<tr>
<td>Glycosphingolipid biosynthesis - lacto and neolacto series</td>
<td>1.86</td>
<td>0.0107</td>
<td>0.0868</td>
</tr>
<tr>
<td>Cysteine and methionine metabolism</td>
<td>1.83</td>
<td>0.0145</td>
<td>0.0962</td>
</tr>
<tr>
<td><strong>Core glycolysis</strong></td>
<td><strong>1.80</strong></td>
<td><strong>0.0156</strong></td>
<td><strong>0.0915</strong></td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>1.73</td>
<td>0.0219</td>
<td>0.1230</td>
</tr>
</tbody>
</table>

**B.**

**BRCA**

**C.**

**LU**

**D.**

**OV**

**E.**

**RNA-based FDG score**

**F.**

**Spearman rho values**

**Range of Integrated CNA score**
Figure EV5. Metabolic patterns are predictive of increased glycolysis in MEF lines (related to Fig 6).

A In cross-prediction tests, signature A-based predictions of signature A MEF line metabolic phenotypes perform better than signature B-based predictions. Glucose consumption was measured using a bioanalyzer.

B In cross-prediction tests, PC1 scores of signature A MEFs are more predictive of average growth fold change in signature A MEFs, while scores from signature B MEFs are more predictive in signature B MEFs.

C–G Mass spectrometry-based metabolite measurements. Samples are arranged left to right from the strongest signature B (blue) PC1 scores to the strongest signature A PC1 (red) scores. Sample D10 (purple) is a mixed signature A/B line. (C) Relative intracellular metabolite levels. (D) Isotopomer distributions for intracellular metabolites (M0, monoisotopic molecular weight). (E) Percent label for intracellular metabolites defined as percent of metabolite molecules with isotopomer mass greater than the monoisotopic molecular weight. Note, the high to low range for percent metabolite molecules with incorporated label varies for each metabolite and generally does not extend from 0 to 100%. (F) Relative metabolite consumption from and secretion into media normalized to initial media levels (i.e., cellular metabolite footprint). Here, positive and negative values indicate metabolite secretion into and consumption from the media, respectively. In (D–G), the correlation of each metabolic parameter with glucose consumption (as measured by a bioanalyzer, and as a positive variable, Fig 6A) is indicated (Pearson correlation, \( r \)). In the isotopomer cases (D), the summary metric of percent label is used for the correlation determination. Notably, cells with high glucose consumption rates tend to exhibit high consumption of serine, glycine, and glutamine, and high secretion of lactate (with concordant lactate secretion results obtained by bioanalyzer-based measurement of the same spent cell culture media). (G) Pathway schematics of differences in relative media levels (footprint, from F) between signature A and signature B MEFs. Each node is colored according the signal-to-noise (SNR) ratio. Red indicates higher post-culture media levels in signature A MEFs, and blue indicates higher post-culture media levels in signature B MEFs. Error bars indicate standard deviations of biological replicates. Footprinting data were normalized to the integrated cell number of the time course of labeling (24 h), and intracellular metabolite concentrations were normalized to the number of cells present at the time of extraction. Full metabolomic data can be found in Table EV5.
Correlation with average fold change cell number

Training set: Sig. A Sig. B Sig. A Sig. B
Prediction: Sig. A Sig. B

Relative intracellular metabolite levels

Strength of Sig. B Mixed

Isotopomer distributions for intracellular metabolites

Percent label amounts for intracellular metabolites

Media relative levels (footprint profiling)

Media relative levels

Figure EV5.