

**Table S1. Number of patients for breast cancer subtype in the TCGA dataset according to IHC status.** Sample sizes of the initial dataset, and for samples that were able to be included at each stage (mutation burden calling, neoepitope load estimation, and then gene expression analysis), are reported for each of the three breast cancer subtype categories (ER/PR(+)HER-2(-), HER-2(+), TNBC).

<b>Breast cancer subtype</b>	<b>Sample size</b>		
	<b>Mutation burden</b>	<b>Neoepitope load</b>	<b>Expression data</b>
<b>ER/PR(+)HER-2(-)</b> ER(+) only, ER(+)/PR(+), PR(+) only, HER-2(-) for all	630	586	583
<b>HER-2(+)</b> HER-2(+) only, ER(+)/PR(+)/HER-2(+), ER(+)/HER-2(+), PR(+)/HER- 2(+)	141	138	138
<b>TNBC</b> ER(-)/PR(-)/HER-2(-)	99	93	92
<b>Total</b>	870	817	721