SUPPLEMENTARY METHODS

Pathogenicity score

A combination of Mutation Taster [1], CHASM (breast) [2] and FATHMM [3] was used to define the potential functional effect of each missense SNV. Missense SNVs defined as non-deleterious/passenger by both MutationTaster [1] and CHASM (breast) [2] a combination of mutation function predictors shown to have a high negative predictive value [4], were considered likely passenger alterations. The remaining missense SNVs were defined as likely pathogenic if they were predicted to be “driver” and/or “cancer” by CHASM (breast classifier) and/or FATHMM [3] respectively, or affected cancer genes included in the cancer gene lists described by Kandoth *et al.* (127 significantly mutated genes) [5], the Cancer Gene Census [6] or Lawrence et al. (Cancer5000-S gene set) [7], or affected hotspot residues [8]. In-frame indels defined as “neutral” by MutationTaster [1] and PROVEAN [9] were defined as likely passengers. The remaining in-frame indels, as well as frameshift, splice-site and truncating mutations were considered likely pathogenic if they were targeted by loss of the wild-type allele (see below) or affected haploinsufficient genes or affected cancer genes [5-7] or affected hotspot residues [8]. Mutations that were neither likely pathogenic nor likely passenger were considered of indeterminate pathogenicity.

Phylogenetic tree construction

A starting tree was constructed using the Neighbor-joining method and Hamming distance and optimized using the parsimony ratchet method [10] implemented in the R package Phangorn [11]. Trees were rooted at the hypothetical normal where all somatic alterations are absent. Branch lengths were determined according to the ACCTRAN criterion as implemented in the Phangorn package and were drawn to scale.

References

1. Schwarz JM, Rödelsperger C, Schuelke M, et al. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010;7:575-576.
2. Carter H, Chen S, Isik L, et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res*. 2009;69:6660-6667.
3. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 2013;34:57-65.
4. Martelotto LG, Ng CK, De Filippo MR, et al. Benchmarking mutation effect prediction algorithms using functionally validated cancer-related missense mutations. *Genome Biol.* 2014;15:484.
5. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature* 2013;502:333-339.
6. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. *Nat Revi Cancer* 2004;4:177-183.
7. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014;505:495-501.
8. Chang MT, Asthana S, Gao SP, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol.* 2016;34:155-163.
9. Choi Y, Sims GE, Murphy S, et al. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012;7:e46688.
10. The parsimony ratchet: a new method for rapid parsimony analysis and broad sampling of tree islands in large data sets. *Proceedings of the Program of the 17th meeting of the Willi Hennig Soc*, Sao Paulo, Brazil; 1998.
11. Schliep KP. phangorn:phylogenetic analysis in R. *Bioinformatics* 2011;27:592-593.