Supplementary Note for Ref 1

1. Generating candidate clone genotypes from predicted variant clusters by FastClone (CloneFinder+)

FastClone predicts variant clusters and their evolutionary relationships for a given tumor sample, i.e., ancestor-descendant, sibling, or monoclonal. Based on the predicted variant clusters and their evolutionary relationships, CloneFinder+ constructs candidate clone sequences by accumulating all predicted mutations (variants) from the root cluster to a target cluster.

Each cluster is composed of variants with >99% estimated probabilities for the assignment because most of the correct predictions received a very high probability in our simulation study. An ambiguous base is assigned for a variant with a probability between 1% and 99%, where CloneFinder+ predicts that a variant with the probability <1% is not present within a given cluster.

For variant clusters in an ancestor-descendant relationship, only a clone genotype that has mutations up to the ancestral cluster is generated because sibling clusters were often incorrectly predicted to be ancestor-descendant clusters in our simulation study. This error is reasonable as observed VAFs are not precise. For predicted sibling clusters, only tip (descendant) clones are generated because the presence of the ancestral clone is not assessed in the FastClone analysis, i.e., a by-product. Clone genotypes that have a larger number of ambiguous bases than the number of mutations, i.e., limited resolution, are not included in the collection of candidate clones. When all clones are excluded for a tumor sample or when FastClone has predicted only a single cluster, the genotype of the tumor sample is used as the candidate clone for a sample, i.e., the predicted monoclonal sample. To generate a tumor genotype, all variants with VAFs greater than zero are assumed to be present.

2. Parameter settings

CloneFinder+: In the CloneFinder+ analysis, we clustered variants without giving the tumor purity and used variants with at least 50 reference read counts and two mutant read counts to assess candidate clones. During the analysis, candidate clone genotypes with <1% clone frequencies for all tumor samples were discarded.

PathFinder: PathFinder was performed by giving the information of the correct primary tumor samples, and CloneFinder+ clones that were predicted with >5% clone frequencies were considered to be present for a given tumor sample.

LICHeE: In the LICHeE analysis, variants were considered robustly present in a sample at VAF > 0.005 (robust variants). We set the VAF error margin to be 0.1, and variants with VAF > 0.6 were excluded. LICHeE groups variants based on the pattern of presence/absence of mutations across the samples, and each variant group was required to contain at least two robust variants. LICHeE also clusters variants by VAF similarities. We required that a variant cluster contained at least two variants unless a variant was sample-specific. All the variant groups/clusters were initially kept in the network. Two groups/clusters could collapse when the mean VAF difference was <0.01.

CloneFinder: Similarly, CloneFinder was performed using variants with at least 50 reference read counts and two mutant read counts, and we discarded clones when estimated clone frequencies were <1%.