common in both cancer cells [21, 22] and the human genome [27], and it can lead to local fitness maxima when it affects mutations that individually increase fitness—see also [28]. Thus, to examine if CPMs can be used to predict paths of tumor progression we will need to assess how the quality of the predictions is affected by multi-peaked fitness landscapes.

The second question addressed in this paper is whether we can use CPMs to estimate evolutionary unpredictability, regardless of the performance when predicting the actual paths of tumor progression. A model could be useful if it suggests few paths are possible, even if its actual predictions about the distribution of paths are not trustworthy. Conversely, predicting correctly the distribution of paths of tumor progression might be of little importance in scenarios where the true evolutionary unpredictability itself is very large (where disease progression follows a very large number of possible paths); for practical purposes, forecasting here would be useless.

Inline note <

To address the above questions (can we predict the paths of tumor progression using CPMs?; can we estimate evolutionary unpredictability using CPMs?) we use evolutionary simulations on 1260 fitness landscapes that include from none to severe deviations from the assumptions that CPMs make about the structure of fitness landscapes, and we analyze the data with four different CPMs, whose predictions about restrictions in the order of accumulation of mutations we have adapted to provide probabilities of paths of tumor progression. This paper does not attempt to understand the determinants of evolutionary (un)predictability (see, e.g., [5, 6, 25, 29, 30]) but, instead, we focus on the effects of evolutionary unpredictability for CPMs. This is why we use variation in key determinants of evolutionary unpredictability (e.g., variation in population sizes and mutation rates) but these factors are only used to generate variability in unpredictability, and not themselves the focus of the study. To better assess the quality of predictions, we use sample sizes that cover the range from what is commonly used to what are much larger sample sizes than currently available. We also include variation in the cancer detection process or detection regime (when cancer samples are taken, or when patients are sampled), since previous studies have shown that it affects the quality of inferences from CPMs [31].

Typewriter note

We have shown before [31] that the performance of two CPMs (CBN and CAPRI) for predicting accessible genotypes degrades considerably when the fitness landscapes contain reciprocal sign epistasis. That study focused on predicting accessible genotypes and its results cannot provide an answer to the questions about predicting paths of tumor progression and estimating evolutionary unpredictability. We are extending our previous study to answer whether CPMs can be used to predict paths of progression and to estimate evolutionary unpredictability. To address these questions we need to look directly at the prediction of paths (not genotypes), and compare them with the true paths of progression, as we do in the current work. Thus, the two studies differ in objectives, methods (here we use a larger number of CPMs, we follow evolution until fixation, and we develop procedures to compare predicted with true paths of tumor progression), and scenarios considered (the types fitness landscapes used and the extent of evolutionary unpredictability); see details in S1 Text.

Here we find that the agreement between the predicted and true distributions of paths is generally poor, unless sample sizes are very large and fitness landscapes conform to the assumptions of CPMs. Both detection regime and evolutionary unpredictability itself have major effects on performance. But in spite of the unreliability of the predictions of paths of tumor progression, we find that CPMs can be useful for estimating upper bounds to the true evolutionary unpredictability.

What are the implications of our results for the analysis and interpretation of the use of CPMs with cancer data sets? We analyze twenty-two real cancer data sets with H-CBN, the best performing CPM in the simulations. We cannot examine how close predictions are to the