Frequency-dependent selection of neoantigens fosters tumor immune escape and predicts immunotherapy response

Shaoqing Chen[1,2#], Duo Xie[2,3#], Jiguang Wang[4,5,6,7], Zheng Hu[2*], Da Zhou[1,8*]

[1]School of Mathematical Sciences, Xiamen University, Xiamen, China
[2]CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[3]Faculty of Health Sciences, University of Macau, Taipa, Macau, China
[4]Division of Life Science and State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China
[5]Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China
[6]Hong Kong Center for Neurodegenerative Diseases, InnoHK, Hong Kong SAR, China
[7]HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, China
[8]National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China
[#]These authors contributed equally
*Corresponding authors: zheng.hu@siat.ac.cn; zhouda@xmu.edu.cn;

Abstract

Cancer is an evolutionary process undergoing stringent immune selection. However, recent studies have revealed that certain tumors undergo neutral evolution following the malignant transformation. Here, we propose negative frequency-dependent selection (or NFDS), where the immune response against cancer cells depends on the clonality of neoantigens, can lead to an immunogenic landscape that is similar to neutral evolution. With a stochastic model of NFDS, we demonstrated that NFDS promotes early immune evasion in hypermutable tumors, leading to neutral-like evolutionary dynamics with high antigenic heterogeneity and worse response to immune checkpoint blockade (ICB) therapy. Our model also revealed that NFDS is characterized by a negative association between average clonality and total burden of neoantigens. Indeed, NFDS was validated with whole-exome sequencing datasets (357 tumor samples from 275 patients) from four melanoma cohorts with ICB therapy. Altogether, our study provides the first quantitative evidence supporting the theory of NFDS in immune-driven cancer evolution. These findings highlight the critical role of NFDS theory in devising more efficient and predictive immunotherapies.

Keywords: Cancer evolution, Tumor Heterogeneity, Negative frequency-dependent selection, Neoantigens, Immunotherapy

Introduction

Cancer is an evolutionary process driven by genetic or epigenetic alterations in the genome and the interactions between cancer cells and the microenvironments[1-4]. Understanding the evolutionary dynamics that are operative at different stages of progression in individual tumors might inform the earlier detection, diagnosis, and treatment of cancer[5]. Cancer evolution is particularly shaped by a complex interplay between mutated cancer cells and immune cells capable of recognizing and combating them[6]. Herein, somatic mutations can be translated into novel peptides and presented on the cancer cell surface, thus generating neoantigens[7]. Tumor neoantigens are capable of eliciting a T-cell response and are recognized by the patient's human leukocyte antigen (HLA) molecules, prompting the T cells to attack and eliminate the mutated cancer cells[8].

This antitumor immune response confers a negative selection pressure[9], and there are several mechanisms that cancers can co-opt to evade immune recognition and/or elimination[10, 11]. For instance, the programmed death ligand 1 (PD-L1) can inhibit the immune system by interacting with PD-1 on activated T cells, B cells, and myeloid cells[12]. This understanding has fueled the development of immune checkpoint blockade (ICB) therapies, which have shown significant promise in enhancing the survival rates of patients with a variety of tumor types[13]. However, certain tumor patients remain unresponsive to ICB therapies despite high tumor mutational burden and/or extensive immune cell infiltration, while the underlying mechanisms remain elusive[14, 15]. Several explanations have been proposed, such as insufficient immune cell infiltration[16], PD-L1 overexpression offsetting therapy[17], loss of an HLA haplotype[18, 19], somatic mutations in B2M gene or HLA alleles leading to dysfunctional neoantigen presentation[20, 21], high intra-tumor heterogeneity[22], *etc*.

Several studies have indicated that certain types of tumors can evolve neutrally, wherein subclonal lineages exhibit no selective difference in cell fitness[23-26]. Interestingly, these neutral tumors are often linked to immune escape and poor prognosis[27]. In fact, neutral tumor evolution leads to high intra-tumor heterogeneity (ITH)[23] and thus may result in an ineffective immune response of the developing tumor[28]. However, clonal selection has also been found in patients with poor prognosis[29], challenging the view of neutral evolution[23-26]. The limitations of these evolutionary features necessitate in-depth investigation of immune selection in tumors.

Noting the parallels between cancer-immune dynamics and prey-predator interactions in ecological systems[30, 31], here we propose that immune-driven cancer evolution adheres to organism evolution under negative frequency-dependent selection (NFDS), leading to effectively neutral evolution. Essentially, predators are inclined to prey on the most abundant species, inadvertently providing survival advantages to less abundant species[30]. Analogously, immune cells could target cancer cells expressing a certain abundance of neoantigens allowing those with fewer neoantigens to dodge immune responses. In fact, mouse experiments have shown that immunogenic tumor antigens do not lead to immune-mediated cell rejection when the fraction of cells bearing each antigen ('clonal fraction') is low[32], thus supporting NFDS in immune-driven cancer evolution. However, this theory remains to be tested in human patient samples.

In this study, we employed stochastic modeling to investigate the impact of NFDS on the evolutionary dynamics and neoantigen subclonal structure of hypermutated tumors. We found that tumors subject to NFDS exhibit high tumor heterogeneity and low clonality of neoantigens. We then simulated tumor immunotherapy and discovered that NFDS tumors were less responsive to ICB therapies than tumors under conventional purifying selection (PS). Furthermore, pre-therapy PS and NFDS *virtual* tumors exhibited distinct features regarding the clonality of neoantigens, which associated with different ICB therapy responses. We then analyzed WES datasets from four melanoma patient cohorts including 357 tumor samples from 275 patients[16, 22, 33, 34] and found that NFDS was supported by the unique evolutionary patterns in these datasets.

Results

Modeling neoantigen evolution under different immune selection scenarios

We utilized a stachastic branching model to investigate tumor evolution under four immune selection scenarios: (1) purifying selection (PS); (2) NFDS; (3) PS with subclonal immune escape (IE); and (4) NFDS with subclonal immune escape (IE). A cancer cell harboring a sufficient amount of neoantigens is defined as immunogenic. While both PS and NFDS undergo negative Darwinian selection (selective coefficient $s<0$) against immunogenic cancer cells, PS allows the immune system to respond even only one immunogenic cell exists, independent of the frequency of immunogenic cells in a population[9, 35, 36]. In contrast, NFDS requires a sufficient number of immunogenic cells to trigger an unremitting immune response (**Fig. 1a**). Here highly immunogenic neoantigens that present in high clonality (measured by cancer cell fraction or CCF) will be more susceptible to negative selection[37] (see **Methods**). For the remaining two scenarios of immune selection, we considered subclonal immune escape upon the PS or NFDS by allowing cancer cells to stochastically acquire immune escaped alterations, leading to neutral evolution of the immune escaped lineages (selective coefficient $s = 0$), regardless of the cell's antigenicity (**Fig. 1b** and **Methods**).

The stochastic branching model of the tumor growth initiated with an antigenically neutral cancer cell (**Fig. 1b**). Cancer cells in a lineage can divide (with a division probability b), die (with a death rate d_c), or enter a quiescent state (quiescent probability is $1-b-d_c$; cells remain viable but stop dividing). Each cell division resulted in m unique mutations that follow a Poisson distribution, $m \sim \text{Poisson}(p_m)$. The newly occurred mutations were neoantigens at rate p, or as passengers (evolutionarily neutral) at rate $1-p$. Each neoantigen exhibited an antigenicity score (a_k) drawn from an exponential distribution. The tumor cell death rate d_c was determined by the cumulative antigenicity, i.e. $A_c=\sum a_k$, and the negative selection intensity s as Eq. (1). Here A_c was defined as the sum of the antigenicity of neoantigens harbored in the lineage. NFDS was modeled by setting a threshold of tumor immunogenicity, $\frac{T_{A_c}}{T} > c_2$, where T_{A_c} and T were the number of immunogenic tumor cells and total number of tumor cells, respectively (**Fig. 1b, Methods**).

With this model, we first examined the growth dynamics of PS and NFDS tumors, respectively (**Fig. 1c-d**). We found that NFDS tumors grew faster than PS tumors, indicating that PS imposed a stronger suppression on tumor growth than NFDS. As selection intensified, PS tumors grew more slowly, while NFDS tumors still maintained a steady growth rate (**Fig. 1d**), indicating that NFDS tended to protect the cancer cells from elimination even under stringent negative selection (s=-0.8).

We then examined the antigenic mutation landscape under each of four scenarios of immune selection. As expected, the simulation results showed that the number of antigenic mutations in a PS tumor decreased as selection intensified (**Fig. 1e**). Interestingly, although NFDS tumors had a lower antigenic mutation burden as selection intensity increased from low to moderate (**Fig. 1e**), further increases in selection intensity (*s*=-1.2) did not significantly reduce neoantigen burden, but rather increased it. These patterns were observed across different cancer cell fraction (CCF) thresholds as well as different cell and tumor immunogenicity thresholds (**Supplementary Figs. 1-3**). Tumors with subclonal immune escape showed similar rollback patterns as NFDS (**Fig. 1f**) because mutations accumulate rapidly in the immune-escaped lineages following neutral evolution. However, the cause of the same phenomenon was different for NFDS where low-cellularity neoantigens of immunogenic cells were maintained by unsustainable negative selection. Our simulations indicated that tumors experienced interleaved immune activation and suppression under intense NFDS (**Supplementary Fig. 5b**). This mechanism is similar to immune exhaustion, where the immune system faces constant pathogen infections and inflammations[38]. The amplitude and frequency of the oscillation determine the profound influence of NFDS on antigenic mutation accumulation, with a more significant effect observed under a relatively weaker selection intensity. We also observed this effect from the difference in growth rates between PS tumors and NFDS tumors (**Supplementary Fig. 5c-d**).

This finding is intriguing because the pattern of neoantigens accumulation under NFDS resembles subclonal immune escape. It is widely accepted that tumor immune escape is achieved through specific genomic mutations in the genome[9, 39-41]. However, our results suggest that the specific interactions between tumor cells and the immune microenvironment confer immune-escape-like evolutionary dynamics.

NFDS promotes evolutionary rescue in hypermutated tumors

To systematically compare the evolutionary dynamics of four immune selection scenarios, we simulated 20 hypermutated tumors (mutation rate μ=5.5 per cell division per genome). Consistent with the above simulations, all PS cancer cells without immune escape were eventually eliminated, and all NFDS tumors were able to grow to the maximum population size (100,000 cells) (**Fig. 2a**). We also observed that *virtual* tumors with subclonal immune escape experienced evolutionary rescue and all grew to the maximum population size (**Fig. 2b**). In contrast to the rescue patterns of NFDS, PS tumor growth was slower soon after initiation, but grew faster after randomly acquiring immune escape alterations (**Fig. 2b**). These results suggest that NFDS may act as an early immune escape mechanism. However, when the mutation rate was a little higher (μ=6 per cell division), all *virtual* tumors without subclonal immune escape were eventually eliminated by negative selection, although most NFDS tumors reached a larger population size (15,000 cells) than PS tumors (**Supplementary Fig. 6**). In contrast, tumors with subclonal immune escape still grew to the maximum population size (**Supplementary Fig. 6**). The extinction of tumors recapitulated Muller's ratchet-like effect of mutation-driven meltdown of a population[42] (**Supplementary Fig. 6**). Together, these results on one hand suggest the immune escape conferred by NFDS is dependent on the mutation rate. On the other hand, NFDS may coexist with conventional immune escape mechanisms (e.g. by escape mutations) in tumors with high mutation rate.

Given that NFDS can diversify the ecological system in species evolution[43-45], we next ask whether this type of balancing selection will increase the clonal diversity of antigenic mutations in cancer cells. To address this question, we characterized the antigenic mutation load, clonality and Shannon diversity of virtual tumors in different scenarios. As expected, NFDS tumors harbored higher antigenic mutation burden and clonal diversity as compared to the conventional purifying selection (**Fig. 2c** right panel; **Supplementary Fig. 7a**), demonstrating that NFDS increases the diversity of subclonal neoantigens. We also noted that the average CCFs of antigenic mutations in NFDS tumors were significantly lower than in PS tumors (**Fig. 2c**). This is because NFDS tended to remove high-frequency neoantigens[37]. Subclonal immune escape increased the average neoantigen CCF, antigenic mutation load and Shannon diversity of PS tumors but not for NFDS tumors (**Fig. 2c-d, Supplementary Fig 7a-b**), suggesting that further acquirement of immune escape alterations has little impact on the neoantigen landscape in NFDS tumors. We also found that tumors under PS eventually had larger fractions of immune escaped cells than ones under NFDS (**Supplementary Fig. 7c**). These results suggest that subclonal immune escape has significant influence on the antigenic mutation landscape under PS but little influence under NFDS (**Fig. 2c-d**; **Supplementary Fig. 7-8**).

To test whether PS and NFDS tumors have different evolutionary patterns of neoantigens, we then simulated 100 *virtual* tumors for each of four scenarios and calculated the antigenic mutation load and average CCF for each *virtual* tumor. We found that in PS tumors, the average neoantigen CCF was positively associated with antigenic mutation load (**Fig. 2e** left panel). In contrast, this association is negative in NFDS tumors, regardless of cell-immunogenic or tumor-immunogenic thresholds (**Fig. 2e** right panel**; Supplementary Fig. 9**). However, in all tumors with subclonal immune escape, we also observed a positive correlation. To investigate the underlying mechanisms of these correlations, we then performed a mathematical analysis using the ordinary differential equation (ODE) model. According to Eqs. (4) and (7) in **Supplementary Note**, the correlations between the average CCF and the number of antigenic mutations can be inferred, which confirmed the negative correlations between the average neoantigen CCF and the antigenic mutation load under NFDS (**Fig. 2e and 2f**).

In PS tumors with immune escape, we also found that the average CCF was strongly correlated with the proportion of immune escaped cells (**Supplementary Fig. 10a**). However, this correlation was not observed in NFDS tumors with immune escape (**Supplementary Fig. 10b**), suggesting that in NFDS tumors, higher intra-tumoral antigenic heterogeneity provides the potential for greater antigenic mutation burden at lower frequencies. In contrast, in tumors subject to intense purifying selection, a higher antigenic mutation load typically arises from earlier immune escape, resulting in a higher average CCF of antigenic mutations. These observations suggest that tumors may adopt different strategies to maintain a substantial burden of antigenic mutations. Importantly, the distinct association between average neoantigen CCF and antigenic mutation load offers a way to test NFDS in real patient data.

Lack of response to ICB therapy in NFDS *virtual* tumors

To further explore the impact of immune checkpoint blockade (ICB) therapy on PS and NFDS tumors, respectively, we modeled ICB therapy by restoring negative selection on immune escaped cells. Simulations showed that PS tumors can be eliminated by the immune system after ICB treatment, which can be classified as "responders" (**Fig. 3a**). In contrast, although NFDS tumors shrank earlier in size after ICB therapy, they eventually manifested as progressive disease and therefore can be classified as "non-responders" (**Fig. 3a**). We then generated 50 treated and 50 untreated virtual patients with PS and NFDS tumors, respectively. We defined a *virtual* patient as deceased when the tumor reached a maximum cell population size (100,000 cells). Survival analysis showed that ICB-treated *virtual* patients with NFDS tumors had significantly lower survival rates than patients with PS tumors (**Fig. 3b**). In untreated virtual patients, NFDS tumors still exhibited worse survival (**Fig. 3c**), suggesting that NFDS tumors may represent a more aggressive subtype. These results also were also consistent with the clinical observations that heterogeneous tumors have a worse response to immunotherapy[15, 28, 34]. Notably, the negative correlation between average CCF and subclonal antigenic mutation load was maintained even after ICB therapy (**Fig. 3d**).

We observed characteristic CCF distributions with distinct subclones in untreated PS tumors (**Fig. 3e**). However, these distinct subclones were not found in untreated NFDS tumors, in line with effectively neutral evolution[46] (**Fig. 3e**). By fitting the CCF distributions of simulated pre-therapy tumors to the power-law model of neutral evolution[25], we found a significantly higher R^2 fitting score in NFDS tumors as compared to PS tumors (**Fig. 3f**). The results were similar when taking account of varying sample sequencing depths and tumor purities (**Supplementary Fig. 11**). Of note, for PS tumors, the average CCF was significantly decreased after immunotherapy (**Fig. 3g**), while the average CCF of NFDS tumors was only slightly affected by immunotherapy (**Fig. 3h**). Taken together, these results suggest that the pre-therapy CCF distributions might inform the prognosis of immunotherapy. Namely, tumors with a CCF distribution containing distinct high-frequency subclones may respond better to immunotherapy.

Predicting response to ICB therapy based on pre-therapy CCF distributions

Next, we sought to test where NFDS represents an immune escape mechanism in real-world patients. In particular, our simulations have generated a testable prediction for NFDS, namely the negative correlation between average CCF of neoantigens and subclonal antigenic mutation load. We analyzed whole-exome sequencing (WES) data from tumor biopsies of four publicly available melanoma patients with ICB therapy[16, 22, 33, 34]. Since melanoma generally exhibit a high mutation burden[47], these sequenced biopsies provide an excellent dataset to test our model. An in-house bioinformatics pipeline[48, 49] was used to identify somatic single nucleotide variants (sSNVs), insertions/deletions (indels) and somatic copy number alterations (sCNAs) while neoantigens was predicted for sSNVs and indels using NeoPredPipe[50] (**Methods**). We also estimated the cancer cell fraction (CCF) of sSNVs and indels to discriminate clonal (the upper bound of the 95% confidence interval (CI) of CCF ≥ 1) versus subclonal (the upper bound of the 95% CI of CCF < 1) sSNVs (**Methods**).

Indeed, we found a significant negative correlation between the average CCF of neoantigens and the number of subclonal antigenic mutations in pre-therapy samples of melanoma patients from Reuben et al.[34] (**Fig. 4a**), indicating this dataset fits well with the expectation of NFDS model[33]. We also observed a similar negative correlation in the pre-therapy patient cohort of Amato et al. (**Fig. 4b**). To investigate the effect of immune escape, we analyzed post-therapy non-responder samples (labeled as stable disease/progression disease (SD/PD) according to RECIST 1.1 criteria[51]) from this cohort and also found a trend of negative correlation between the two variables (Spearman's ρ= -0.90, **Fig. 4c**). In another cohort Riaz et al.[16], we found a significant positive correlation in both pre- and post-therapy samples (**Fig. 4d-e**), which rejected NFDS model. However, when focusing on the post-therapy non-responder samples (SD/PD) with high mutation rate (subclonal neoantigen burden N>50), a negative trend (Spearman's ρ= -0.62) between average CCF and subclonal neoantigen burden was still observed, although not statistically significant due to limited sample size (**Fig. 4f**). These results support NFDS of neoantigens in hypermutated melanoma tumors even after ICB immunotherapy.

One of the well-established predictors of response to ICB therapy is T-cell infiltration, as also shown in the Riaz cohort[16] (**Supplementary Fig. 12a**). However, our results suggest the existence of another effective biomarker when T-cell infiltration is comparable between responders (partial response /complete response, PR/CR) and non-responders (SD/PD), that is the evolutionary mode of pre-therapy tumors (**Fig. 3b-c, e-f**). In particular, we observed no significant difference in cytolytic scores (representing T-cell infiltration) between responders and non-responders in Liu et al's cohort[22] (**Supplemnentary Fig. 12b**). However, through neutrality testing on this dataset, a significant difference in R^2 values between responders and non-responders was observed (**Fig. 4g**). This result was consistent with our model prediction that neutral-like evolution is associated with worse response to ICB therapy (**Fig. 3b-c, e-f**). This pattern is mirrored when performing the same test in the Amato et al's cohort[33] (**Fig. 4h**). To evaluate the clinical significance of neutrality in CCF distributions, we set a threshold for high R^2 values and performed survival analysis on patients from Liu et al's cohort where the patient survival information was available[22]. Indeed, patients with high R^2 values (more neutral-like) showed a significantly shorter overall survival as compared to patients with low R^2 values (more selection-like) (**Fig. 4i**). In conclusion, our results indicate that the distribution of CCF in pre-therapy tumor samples can serve as a predictive marker for the response to ICB therapy in melanoma patients. This discovery provides a new evolutionary biomarker for prioritizing patients for immunotherapy.

Discussion

As a model of balancing selection, NFDS has been proved to be a robust mechanism for maintaining genetic diversity within a population[43-45]. Its role has been posited to drive coevolution between biological antagonists[52, 53]. The diversity of major histocompatibility complexes (MHC) is famous example. Tasked with recognizing foreign antigens, MHC molecules posseses an extremely high allelic diversity within the human population[54]. A popular theory suggests that NFDS is a major contributor to this extensive diversity[55], as pathogens evolve to avoid the most prevalent MHC in the immune system[56].

Although neoantigens and immune systems are antagonists in tumor evolution[30], the potential role of NFDS in this context remains underexplored. Current theory posits that neoantigens undergo conventional purifying selection[9]. Furthermore, the extensive mutational diversity in tumors is often attributed to neutral evolution[23-26]. Although several studies have argued for strigent selection in cancer, most of them assume that each genotype has a steady fitness effect, regardless of its relative frequency in the population[57, 58]. Our current study proposes and tests the hypothesis of NFDS-driven tumor evolution, where cancer cells with higher neoantigen abundance are toward elimination and cancer cells with lower neoantigen abundance exhibit higher fitness. Through stochastical modeling, we demonstrate that tumors subject to NFDS exhibit high intra-tumor heterogeneity and unique antigenic mutation landscape, similar to effective neutral evolution. In particular, a negative association between the average clonality of neoantigens and the number of subclonal antigenic mutations represents a distinct hallmark of NFDS. We then provide the quantitative evidencethat support NFDS occurs in melanoma patients.

Our findings provide important insights into our understanding of immunotherapy and could potentially guide the development of more effecitve treatment strategies. We propose a potential explanation for why tumors with a high neoantigen load often prove resistant to ICB therapies. Our evidence suggests that NFDS contributes to the evolution of heterogeneous tumors that subsequently exhibit reduced responsiveness to ICB therapies. Additionally, the neutral tail of the neoantigen CCF distribution could serve as an indicator of NFDS and potentially predict which tumors are undergoing NFDS. Another potential application is to inform tumor treatment strategies, such as the adaptive therapy approach, in which patients are repeatedly given the minimum dose necessary to prevent the growth of treatment-resistant tumor cells and to control tumor size[59]. Further study of the NFDS will hopefully make it possible to tailor treatment to individual patients using their specific sequenced biopsy and adaptive therapy strategy combined with neoantigen-specific therapies[7].

Our study, while valuable as a proof of concept, acknowledges the limitations by the inclusion of only four melanoma patient cohorts[16, 22, 33, 34]. This small sample size may not fully capture the heterogeneity seen across diverse tumor types. Expanding the study to include a larger set of pan-cancer patient cohorts for hypermutated tumor would allow for a more general examination of NFDS in immune selection and a more comprehensive analysis of the evolutionary dynamics of tumors. It is also important to note the potential challenges in distinguishing between NFDS and other immune escape mechanisms, particularly when a patient has not received ICB therapy. However, with the growing prevalence of ICB therapy, NFDS may emerge as another hurdle in tumor treatment, warranting further research and solutions.

In summary, our theoretical framework of tumor evolution offers critical insights into the evolutionary dynamics that occur between cancer cells and the immune system. We substantiate these findings by validating them with data derived from actual patients. This research not only enhances our understanding of tumor growth and immune interaction but also provides a valuable perspective that could guide future studies and the development of effective treatment strategies.

Methods

Stochastic branching process of neoantigen evolution

A stochastic model was developed to capture the evolutionary dynamics of neoantigens under NFDS. The model is based on a branching process that simulates the exponential growth of cancer cells[9, 60, 61]. In each step of the model, tumor cells belonging to a lineage can proliferate at a rate of b and survive at a basal rate of b_0 in the absence of negative selection pressure. The two daughter cells resulting from each division accumulate mutations at an overall rate of p_m, where m represents the number of generated mutations (m ~ Poisson(p_m)). A fraction of these mutations are assigned antigenicities at rate p, drawn from an exponential distribution, which determines their potential to be recognized as neoantigens.

To initialize the model, the founder cell was set to be antigenically neutral. This assumption was made to avoid the rapid extinction of tumor lineages in cases where the founder cell carries a highly immunogenic neoantigen that elicits a strong immune response (**Supplementary Fig. 4**). In the model, a tumor cell can die with a death rate of d_C determined by the accumulated antigenicity A_C and the intensity of negative selection, denoted by s (where s is a non-positive number). The cumulative antigenicity A_C is defined as the sum of antigenicities $\sum a_k$ carried by the neoantigens in the cell, where a_k represents the k-th antigen. If a cell is considered immunogenic, its basal survival rate is reduced based on the cumulative antigenicity and the intensity of negative selection within the tumor microenvironment. As the number of neoantigens in a cell increases, the likelihood of becoming an immunogenic cell in the lineage also increases. This effect was modelled by setting a threshold c_1 for the cumulative antigenicity of tumor cells. Once a tumor cell accumulates enough neoantigens to exceed the threshold, it is considered immunogenic, i.e.

$$\sum a_k > c_1 \,. \qquad (1)$$

The Gillespie algorithm[60] was used to simulate populations of cells. Each time a cell divided, it acquired m new unique mutations following the Poisson distribution. The newly derived mutations were assigned as neoantigens at rate p, or as passengers (evolutionary neutral) at rate $1-p$. Each antigenic mutation was assigned an antigenicity value (denoted a_k for the k-th antigen in a given cell) sampled from an exponential distribution with mean equals to 0.2[9]. The death rate of each cell was determined by the complement of the survival rate reduced by the immunogenicity of the tumor cell. Then we have

$$d_C = 1 - (1 + sA_C)b_0 \,. \qquad (2).$$

Tumor growth and mutation accumulation were simulated using different parameter sets to capture the effects of tumor mutational characteristics and tumor microenvironment. Tumor growth was simulated up to a maximum population size of 100,000 cells. To account for the effect of NFDS, a weighting mechanism was introduced into the cell death rate. Highly immunogenic antigens present in a larger fraction of cells were assigned a higher weight, resulting in a higher likelihood of cell death. In this context, NFDS is defined as a mechanism whereby the presence of highly immunogenic antigens in a larger proportion of cells leads to an increased impact on the cell death rates, i.e.

$$s = \begin{cases} [-2,0), & \frac{T_{A_C}}{T} > c_2 \\ 0, & \frac{T_{A_C}}{T} \le c_2 \end{cases} . \qquad (3)$$

Here T_{A_C} is the number of immunogenic tumor cells, T is the total number of tumor cells, and c_2 is the threshold for the effect of NFDS. $-2 \le s \le 0$ was used in all simulations[9].

Immune escape was modelled by stochastically setting s to 0 for immune escaped cells. Since our focus was on ICB therapies, cancer cells were allowed to acquire immune escape with a probability of p_e. Once a cell acquired immune escape, its survival rate was fixed to the basal survival rate b_0 regardless of its antigenicity. Furthermore, its daughter cells can inherit immune escape, giving them a survival advantage. The probability of immune escape was constant and independent of the accumulation of antigenic mutation.

Parameters were chosen to represent different tumor-immune environments and literatures was reviewed to estimate parameter values in our model. The following parameters were used in all simulations: $b = 0.5$[62, 63], $b_0 = 0.4$, $\mu = 5$, $p = 0.1$, $p_e = 10^{-4}$ (probability of immune escape, if applicable)[9, 41]. For the analyses where cells and tumors were classified as immunogenic or non- immunogenic, the cell and tumor immunogenicity thresholds $c_1 = 0.5$ and $c_2 = 0.5$ were used[9], unless otherwise stated.

To calculate the overall growth rate of the simulated tumors, we performed a linear fit on the simulated tumor growth data (**Supplementary Fig. 5a**). The result showed a good fit between the fitted curve and the simulation curve. Consequently, the exponential growth of the simulated tumors can be characterized as $T \approx T_0 e^{rt}$, where T_0 and T represent the initial and final cell population sizes respectively, and r represents the tumor growth rate. Therefore, we can estimate the tumor growth rate based on the fitted exponential growth model:

$$r \approx \frac{ln\frac{T}{T_0}}{t} \qquad (4)$$

Simulation of VAF/CCF distributions

Tumor evolution was simulated, and each mutation of every simulated tumor was assigned a unique index. To replicate actual sequencing data, mutations with a cancer cell fraction (CCF) greater than 0.04 were retained, and sequencing errors caused by sequencing depth and purity were simulated. The VAF (CCF/2) was considered as true values, and a statistical model that recreates the sequence noise seen in actual data generated the observed VAF. A binomial distribution was utilized in this model, where

$$D_{obs} \sim Bin(n = N,\ p = VAF) \tag{5}$$

Here N is the read depth of a given site an D_{obs} is the number of observed coverage of a mutation. Mutations with $VAF = 0$ were filtered.

Whole-exome sequencing (WES) datasets from four melanoma cohorts

In this study, the publicly available WES data of normal-tumor pairs from the same patient were collected by accessing the Sequenced Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra). These cohorts are as follows:

Liu et al. anti-PD1 melanoma cohort[22]: This cohort consisted of 144 melanoma patients who received anti-PD1 therapy while a subset of 64 patients had progressed from prior ipilimumab treatment. Pre-treatment samples from all 144 patients were included in our analysis. We also used RNA-seq data and cytolitic scores from this cohort for our analysis.

Riaz et al. anti-PD1 melanoma cohort[16]: This cohort consisted of 68 melanoma patients who received anti-PD1 therapy, with 35 patients having progressed from prior ipilimumab treatment. We analyzed pre-therapy and on-therapy samples from this cohort, totaling 68 pre-therapy tumor samples and 41 on-therapy tumor samples. The cytolitic scores were also obtained from the original study[16].

Reuben et al. melanoma cohort with multiple therapeutics[64]: This cohort consisted of 14 patients with synchronous melanoma metastases who received either targeted therapy, immune checkpoint blockade, or were treatment-naive, with a total of 40 tumor samples, after excluding 2 tumor samples due to lack of paired normal tissues.

Amato et al. anti-PD1 melanoma cohort[33]: This cohort comprised 49 patients with melanoma who had undergone immunotherapy. A total of 64 tumor samples were analyzed, including 53 pre-therapy samples and 10 on-therapy samples, after excluding 2 tumor samples for lack of paired normal tissue.

These cohorts have the best overall response (BOR) to anti-PD1 and anti-CTLA4 combination ICB immunotherapy from the original studies. The BOR was determined based on Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 criteria, with complete responders (CR) and partial responders (PR) classified as responders, and stable disease (SD) and progressive disease (PD) as non-responders. Patients with a mixed response or non-evaluable response were excluded from the study.

sSNV and indels calling and neoantigen prediction

For three cohorts[16, 33, 34], whose raw sequencing reads were available in SRA, the sSNV and indel mutations were called using a uniform pipeline. For Liu et al. 's cohort[22], the raw sequencing reads were not available and the processed mutations from its supplementary was used. The sequencing reads were first aligned to the human reference genome (hg38) using Burrows–Wheeler Aligner (bwa; 0.7.17-r1188)[65]. These aligned reads were then processed using GATK (v4.2.6.1; MarkDuplicates, BaseRecalibrator, ApplyBQSR, Mutect2 and FilterMutectCalls)[66]. MarkDuplicates marked duplicate reads, and BaseRecalibrator and ApplyBQSR recalibrated base quality scores together. Mutect2 was used to call somatic sSNVsand indels for each tumor/normal pair. The germline resource (af-only-gnomad.hg38.vcf.gz) and a panel-of-normals generated by CreateSomaticPanelOfNormals, which takes multiple normal sample call sets generated by Mutect2's tumor-only mode, were used. These sSNVs and indels were finally filtered using FilterMutectCalls and annotated using ANNOVAR (v.20200608)[67].

HLA alleles were called using POLYSOLVER 1.0.0 (https://anaconda.org/compbiocore/hla-polysolver). These HLA alleles, along with the called mutations, were used to predict neoantigen using NeoPredPipe[50], a pipeline designed for neoantigen prediction and evaluation. Only single nucleotide variants resulting in a single amino acid change were considered; novel peptides of eight, nine and ten amino acids were considered. Mutations were annotated using ANNOVAR with default parameters. We considered a peptide to be a neoantigen if its predicted affinity ranked 2% compared to a set of random natural peptides to the patient's HLA types.

CCF estimation and analysis

The CNA and CCFs were estimated for each sSNV/indel, and then CCF analysis for these samples were performed. Allele-specific absolute copy number, ploidy and tumor purity were estimated using TitanCNA[68], a hidden Markov model-based method. For each patient, the germline heterozygous SNP at dbSNP 146 loci in the normal biopsy were identified using SAMtools[69] and SnpEff v5.1[64]. Read counts for 1,000 base pair (bp) bins across the genome for all tumor samples were generated using HMMcopy1.36.0[70]. Then TitanCNA was used to determine the allelic ratios at the germline heterozygous SNP loci in the tumor biopsy and the depth ratios between the tumor and normal biopsies in the bins containing those SNP loci. When estimating allele-specific absolute copy number profiles, we only included SNP loci within WES-covered regions. Then 12 runs of TitanCNA with 12 combinations of subclone number, purity and ploidy were performed. In 9 out of the 12 combinations, different numbers of subclones (n=1, 2 and 3), purity (0.2, 0.4, and 0.6), and ploidy (2) were set. In the remaining 3 combinations, we set a different number of subclones (n=1, 2, 3) combined with the purity and ploidy estimated by Sequenza[71]. Then one run was selected for each tumor biopsy based on manual inspection of the fitted results, prioritizing the results with a single subclone unless results with multiple subclones better visibly fit to the data. Three samples were excluded because the results were difficult to interpret.

Then the CCFs and their variation (95% confidence interval) for each sSNV/indel were estimated using CHAT v1.1[72]. CHAT estimates the CCF for each sSNV by adjusting its VAF based on local allele-specific copy numbers at the sSNV locus. Then the CCFs for all sSNVs in the autosomes were calculated using the sSNV frequencies and copy number profiles estimated from the previous steps. Finally, we adjusted the CCFs using tumor purity estimated by TitanCNA. We distinguish subclonal mutations by selecting the upper bound of the 95% CI to be less than 1.

Identification of effectively neutral evolution

The R^2 by fitting the VAF distributions of subclonal mutations to the expectated power-law distribution in neutral model was calculated[25]. For each sample or simulated tumor, the inverse cancer cell frequency (1/f) and the cumulative frequency were calculated. We then fitted the cumulative frequency to a linear regression and obtained R^2, the ratio of the sum of squares of the regression to the sum of squares of the deviations. The closer R^2 is to 1, the more likely the tumor follows neutral evolution. We then compared the R^2 between patients with poor prognosis and patients with good prognosis in the patient cohorts.

Statistical and survival analysis

All statistical analyses were conducted using R version 4.1.3 or Python version 3.9.6. The Wilcoxon rank-sum test was used to assess differences between two groups, while the Kruskal-Wallis test was employed for comparisons among more than two groups. Spearman's ρ coefficient was used to quantify correlations between variables. Survival analyses were performed using the Kaplan-Meier estimator to generate survival curves, and the nonparametric log-rank test was used to compare the survival curves.

Data availability

All patient data analyzed in this article are publicly available through online sources with accession numbers PRJNA639866, PRJNA316754 and PRJNA359359. Other data described in this article were obtained from the original studies.

Code availability

The codes of model simulation, data analysis and visualization of this study are available at https://github.com/Mandy1726367303/immuneNFDS.

Acknowledgments

We thank members of Zhou and Hu laboratories for constructive discussions. This work was supported by National Natural Science Foundation of China (11971405 to D.Z., 32270693 & 82241236 to Z.H.), Guangdong Basic and Applied Basic Research Foundation (2021B1515020042 to Z.H.) and Fundamental Research Funds for the Central Universities (20720230023 to D.Z.).

Author contributions

Z.H. and D.Z. designed the study. S.C. constructed the models and performed the simulation studies. D.X. analyzed the patient sequencing data. S.C., D.X., J.W, Z.H., and D.Z. interpreted the data. Z.H. and D.Z. supervised the study. S.C., D.X., Z.H. and D.Z. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have no competing interests.

References

1. Nowell, P.C. The Clonal Evolution of Tumor Cell Populations. *Science* **194**, 23-28 (1976).
2. Greaves, M. & Maley, C.C. Clonal evolution in cancer. *Nature* **481**, 306-313 (2012).
3. Wu, C.-I., Wang, H.-Y., Ling, S. & Lu, X. The Ecology and Evolution of Cancer: The Ultra-Microevolutionary Process. *Annu. Rev. Genet.* **50**, 347-369 (2016).
4. Vendramin, R., Litchfield, K. & Swanton, C. Cancer evolution: Darwin and beyond. *EMBO J* **40**, e108389 (2021).
5. Fittall, M.W. & Van Loo, P. Translating insights into tumor evolution to clinical practice: promises and challenges. *Genome Medicine* **11**, 20 (2019).
6. Thol, K., Pawlik, P. & McGranahan, N. Therapy sculpts the complex interplay between cancer and the immune system during tumour evolution. *Genome Med.* **14**, 137 (2022).
7. Yarchoan, M., Johnson, B.A., Lutz, E.R., Laheru, D.A. & Jaffee, E.M. Targeting neoantigens to augment antitumour immunity. *Nature Reviews Cancer* **17**, 209-222 (2017).
8. Liu, X.S. & Mardis, E.R. Applications of Immunogenomics to Cancer. *Cell* **168**, 600-612 (2017).

9. Lakatos, E. et al. Evolutionary dynamics of neoantigens in growing tumors. *Nat. Genet.* **52**, 1057-1066 (2020).
10. Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J. & Schreiber, R.D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat. Immunol.* **3**, 991-998 (2002).
11. Liu, Y. & Cao, X. Immunosuppressive cells in tumor immune escape and metastasis. *J. Mol. Med.* **94**, 509-522 (2016).
12. Hudson, K., Cross, N., Jordan-Mahy, N. & Leyland, R. The Extrinsic and Intrinsic Roles of PD-L1 and Its Receptor PD-1: Implications for Immunotherapy Treatment. *Front. Immunol.* **11**, 568931 (2020).
13. Auslander, N. et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat. Med.* **24**, 1545-1549 (2018).
14. Hu, Z.I. et al. Evaluating Mismatch Repair Deficiency in Pancreatic Adenocarcinoma: Challenges and Recommendations. *Clin. Cancer Res.* **24**, 1326-1336 (2018).
15. Mandal, R. et al. Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. *Science* **364**, 485-491 (2019).
16. Riaz, N. et al. Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **171**, 934-949.e916 (2017).
17. Alsaab, H.O. et al. PD-1 and PD-L1 Checkpoint Signaling Inhibition for Cancer Immunotherapy: Mechanism, Combinations, and Clinical Outcome. *Front. Pharmacol.* **8**, 561 (2017).
18. Martínez-Jiménez, F. et al. Genetic immune escape landscape in primary and metastatic cancer. *Nature Genetics* **55**, 820-831 (2023).
19. McGranahan, N. et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **171**, 1259-1271.e1211 (2017).
20. Grasso, C.S. et al. Genetic Mechanisms of Immune Evasion in Colorectal Cancer. *Cancer Discov.* **8**, 730-749 (2018).
21. Sade-Feldman, M. et al. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat. Commun.* **8**, 1136 (2017).
22. Liu, D. et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat. Med.* **25**, 1916-1927 (2019).
23. Ling, S. et al. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6496-6505 (2015).
24. Sottoriva, A. et al. A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209-216 (2015).
25. Williams, M.J., Werner, B., Barnes, C.P., Graham, T.A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nat. Genet.* **48**, 238-244 (2016).
26. Sun, R. et al. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat. Genet.* **49**, 1015-1024 (2017).
27. Johnson, D.C. et al. Neutral tumor evolution in myeloma is associated with poor prognosis. *Blood* **130**, 1639-1643 (2017).
28. Wolf, Y. et al. UVB-Induced Tumor Heterogeneity Diminishes Immune Response in Melanoma. *Cell* **179**, 219-235.e221 (2019).
29. Angelova, M. et al. Evolution of Metastases in Space and Time under Immune Selection. *Cell* **175**, 751-765.e716 (2018).
30. Korolev, K.S., Xavier, J.B. & Gore, J. Turning ecology and evolution against cancer. *Nat. Rev. Cancer* **14**, 371-380 (2014).

31. Hamilton, P.T., Anholt, B.R. & Nelson, B.H. Tumour immunotherapy: lessons from predator-prey theory. *Nat. Rev. Immunol.* **22**, 765-775 (2022).
32. Gejman, R.S. et al. Rejection of immunogenic tumor clones is limited by clonal fraction. *eLife* **7**, e41090 (2018).
33. Amato, C.M. et al. Pre-Treatment Mutational and Transcriptomic Landscape of Responding Metastatic Melanoma Patients to Anti-PD1 Immunotherapy. *Cancers* **12** (2020).
34. Reuben, A. et al. Genomic and immune heterogeneity are associated with differential responses to therapy in melanoma. *NPJ Genom Med* **2** (2017).
35. Butner, J.D. et al. Early prediction of clinical response to checkpoint inhibitor therapy in human solid tumors through mathematical modeling. *Elife* **10** (2021).
36. Kim, P.S. & Lee, P.P. Modeling protective anti-tumor immunity via preventative cancer vaccines using a hybrid agent-based and delay differential equation approach. *PLoS Comput. Biol.* **8**, e1002742 (2012).
37. Zapata, L. et al. Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome. *Genome Biol.* **19**, 67 (2018).
38. Philip, M. & Schietinger, A. Heterogeneity and fate choice: T cell exhaustion in cancer and chronic infections. *Current Opinion in Immunology* **58**, 98-103 (2019).
39. Beatty, G.L. & Gladney, W.L. Immune Escape Mechanisms as a Guide for Cancer Immunotherapy. *Clinical Cancer Research* **21**, 687-692 (2015).
40. Castro, A., Zanetti, M. & Carter, H. Neoantigen Controversies. *Annu Rev Biomed Data Sci* **4**, 227-253 (2021).
41. Hamzeh, K. et al. Mutation Rate Evolution Drives Immune Escape In Mismatch Repair-Deficient Cancer. *bioRxiv*, 2022.2003.2006.482973 (2022).
42. Haigh, J. The accumulation of deleterious genes in a population--Muller's Ratchet. *Theor. Popul. Biol.* **14**, 251-267 (1978).
43. Turelli, M. & Barton, N.H. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G x E interactions. *Genetics* **166**, 1053-1079 (2004).
44. Fitzpatrick, M.J., Feder, E., Rowe, L. & Sokolowski, M.B. Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. *Nature* **447**, 210-212 (2007).
45. Kazancıoğlu, E. & Arnqvist, G. The maintenance of mitochondrial genetic variation by negative frequency-dependent selection. *Ecol. Lett.* **17**, 22-27 (2014).
46. Caravagna, G. et al. Subclonal reconstruction of tumors by using machine learning and population genetics. *Nature Genetics* **52**, 898-907 (2020).
47. Lawrence, M.S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218 (2013).
48. Hu, Z. et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* **51**, 1113-1122 (2019).
49. Hu, Z., Li, Z., Ma, Z. & Curtis, C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nature Genetics* **52**, 701-708 (2020).
50. Schenck, R.O., Lakatos, E., Gatenbee, C., Graham, T.A. & Anderson, A.R.A. NeoPredPipe: high-throughput neoantigen prediction and recognition potential pipeline. *BMC Bioinformatics* **20**, 264 (2019).
51. Eisenhauer, E.A. et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer* **45**, 228-247 (2009).
52. Burdon, J.J., Thrall, P.H. & Ericson, L. Genes, communities & invasive species: understanding the ecological and evolutionary dynamics of host-pathogen interactions. *Curr. Opin. Plant Biol.* **16**, 400-405 (2013).

53. Hori, M. Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. *Science* **260**, 216-219 (1993).
54. Robinson, J. et al. IPD-IMGT/HLA Database. *Nucleic Acids Res.* **48**, D948-D955 (2020).
55. Clarke, B. & Kirby, D.R. Maintenance of histocompatibility polymorphisms. *Nature* **211**, 999-1000 (1966).
56. Borghans, J.A.M., Beltman, J.B. & De Boer, R.J. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* **55**, 732-739 (2004).
57. Attolini, C.S.-O. et al. A mathematical framework to determine the temporal sequence of somatic genetic events in cancer. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17604-17609 (2010).
58. Mumenthaler, S.M. et al. Evolutionary modeling of combination treatment strategies to overcome resistance to tyrosine kinase inhibitors in non-small cell lung cancer. *Mol. Pharm.* **8**, 2069-2079 (2011).
59. Gatenby, R.A. & Brown, J.S. Integrating evolutionary dynamics into cancer therapy. *Nat. Rev. Clin. Oncol.* **17**, 675-686 (2020).
60. Gillespie, D.T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22**, 403-434 (1976).
61. Williams, M.J. et al. Quantification of subclonal selection in cancer from bulk sequencing data. *Nature Genetics* **50**, 895-903 (2018).
62. Dudley, M.E. et al. Cancer Regression and Autoimmunity in Patients After Clonal Repopulation with Antitumor Lymphocytes. *Science* **298**, 850-854 (2002).
63. Eladdadi, A., Kim, P. & Mallet, D. Mathematical models of tumor-immune system dynamics, Vol. 107. (Springer, 2014).
64. Ruden, D. et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Frontiers in Genetics* **3** (2012).
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
66. McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
67. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164-e164 (2010).
68. Ha, G. et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Research* **24**, 1881-1893 (2014).
69. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
70. Ha, G. et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Research* **22**, 1995-2007 (2012).
71. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of Oncology* **26**, 64-70 (2015).
72. Li, B. & Li, J.Z. A general framework for analyzing tumor subclonality using SNP array and DNA sequencing data. *Genome Biology* **15**, 473 (2014).

Supplementary Information

Supplementary Figures 1-12

Supplementary Note

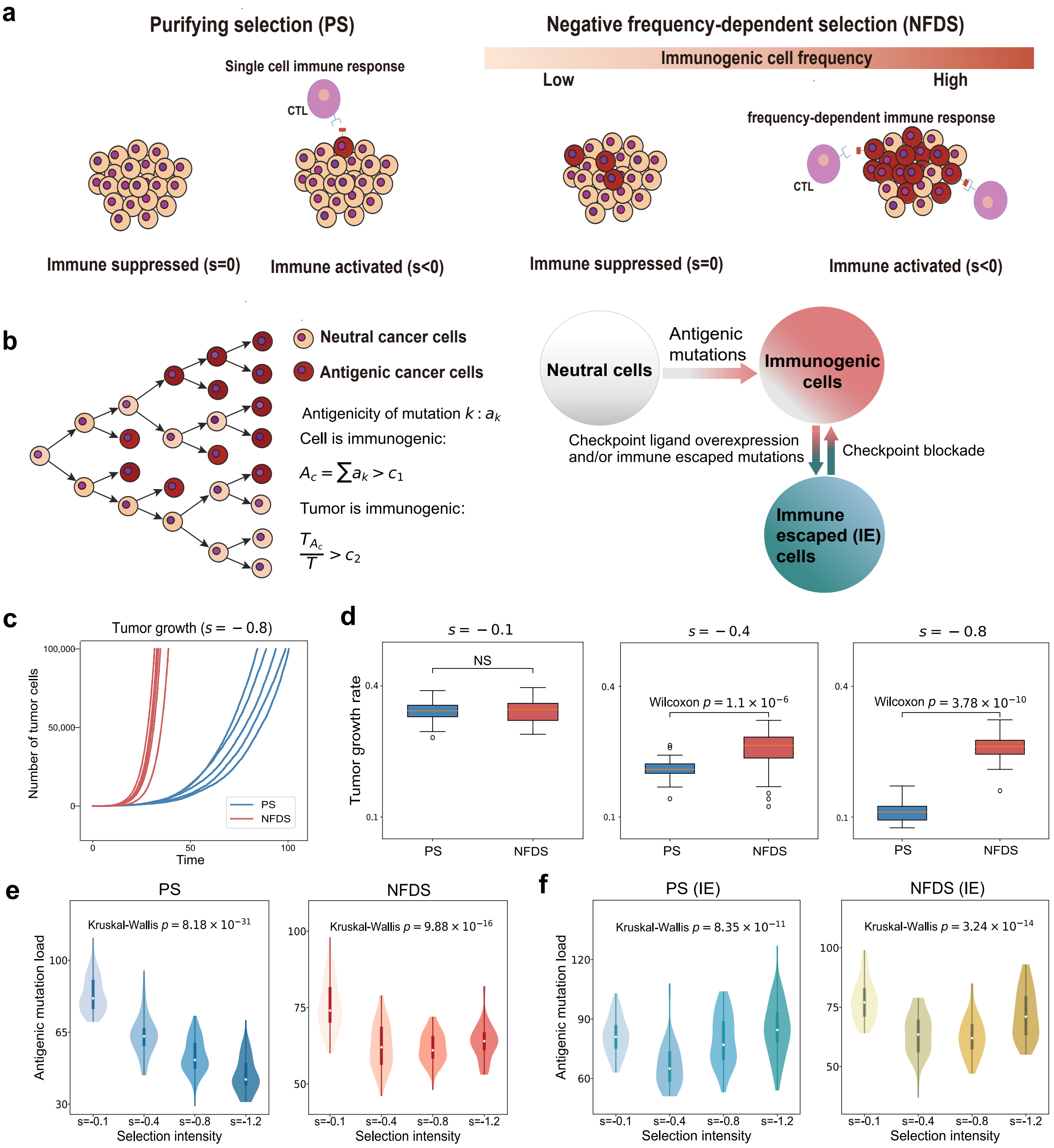

Fig. 1. Models of neoantigen-driven tumor evolution a, Schematic of purifying selection (PS) and negative frequency-dependent selection (NFDS). The red filled circles represent antigenic cancer cells and purple filled circles represent cytotoxic T lymphocytes (CTLs). **b,** Schematic of stochastic branching process with accumulation of neutral and antigenic mutations. c_1 is the threshold of cell immunogenicity while c_2 is the threshold of tumor immunogenicity. T_{A_c} and T are the number of immunogenic and total number of tumor cells, respectively. **c,** Growth curves of five simulated tumors under PS and NFDS, respectively with selection coefficient $s = -0.8$. **d,** Tumor growth rates of simulated tumors $(n = 50)$ under PS and NFDS, respectively at different varying selective intensities. Each simulation was terminated when the tumors reached 100,000 cells. Box plots show median, quartiles (boxes) and range (whiskers). P values, Wilcoxon rank-sum one-sided test. **e,** Number of neoantigens accumulated in simulated tumors at varying selective intensities (each with 50 simulations). **f,** Number of neoantigens accumulated in simulated tumors with subclonal immune escape (IE, immune escape probability $p_e = 10^{-4}$) at varying selective intensities (each with 50 simulations). All tumors were simulated with a tumor mutation rate of $\mu = 5$ per cell division per genome. Violin plots show median, quartiles and range (whiskers). P values, the Kruskal-Wallis test.

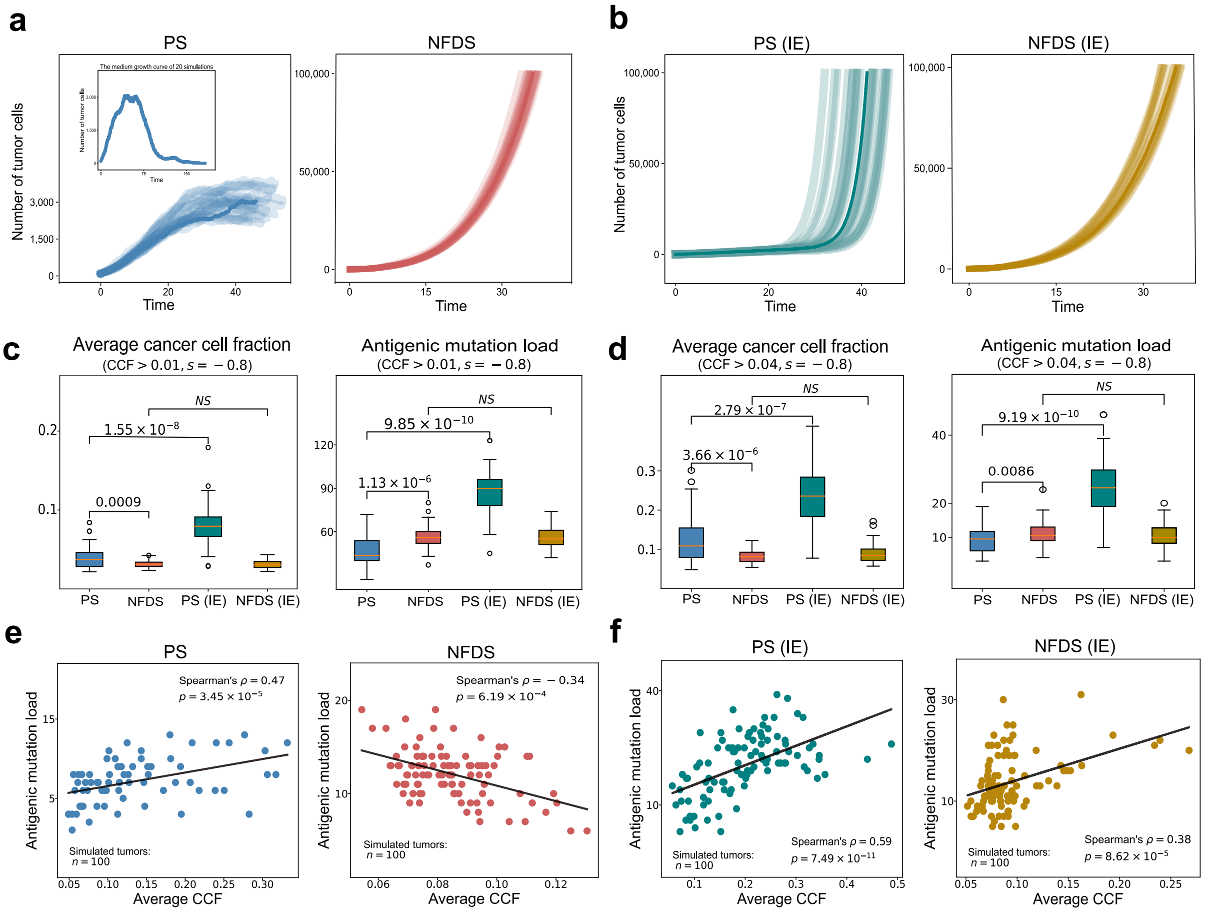

Fig. 2. Evolutionary rescue of tumor growth dynamics and and characteristic neoantigen landscape of NFDS. a, Growth curves of 20 simulated tumors under PS and NFDS, respectively. **b,** Growth curves of 20 simulated tumors under PS (subclonal immune escape or IE) and NFDS (subclonal IE), respectively. Here mutation rate $\mu = 5.5$ and selection coefficient $s = -0.8$ are used. **c-d,** Box plots showing the antigenic mutations accumulated in 50 simulated tumors with cancer cell fraction (CCF)$>$0.01 (**c**) and CCF$>$0.04 (**d**), respectively. The average CCF and mutation load of antigenic mutations from 50 simulated tumors under four evolutionary scenarios are shown. Box plots show median, quartiles (boxes) and range (whiskers). P values, Wilcoxon rank-sum one-sided test. **e,** Correlation analysis between average CCF and subclonal antigenic mutation load of 100 simulated tumors undergoing PS and NFDS, respectively. **f,** Correlation analysis between average CCF and subclonal antigenic mutation load of 100 simulated tumors undergoing PS and NFDS with subclonal immune escape (immune escape probability $p_e = 10^{-4}$, CCF$>$0.04), respectively. The line indicates the linear regression.

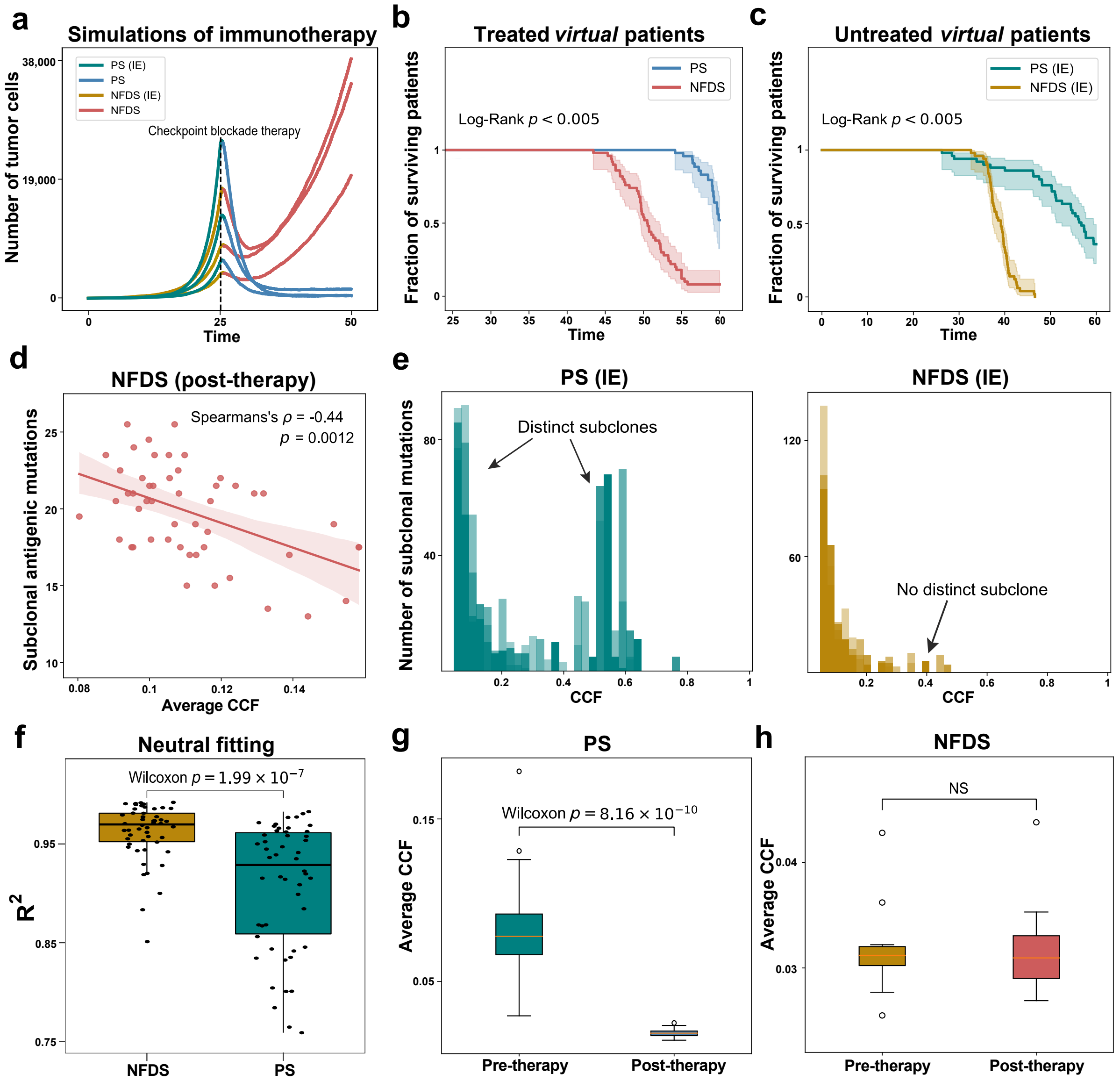

Fig. 3. Computational simulations of immune checkpoint blockade (ICB) therapy. a, Outcome of immunotherapy for six virtual patients under four evolutionary scenarios, respectively. Here mutation rate $\mu = 5.5$ at selection coefficient $s = -0.8$ are used. Virtual patients receive ICB therapy at simulation time of 25. **b,** Kaplan-Meier curves of 50 *virtual* patients that receive immunotherapy at simulation time 25. **c,** Kaplan-Meier curves of 50 untreated *virtual* patients. **d,** Correlation analysis between average CCF and subclonal antigenic mutation load of ICB-treated virtual patients (n=50) who underwent NFDS. Only neoantigens with CCF$>$0.04 are used. The line indicates the linear regression and the shading indicates the 95% CI of the regression. **e,** CCF distributions of all mutations (including neutral passengers) of five simulated tumors under immune escape mode of PS and NFDS, respectively. The model parameters used are immune escape probability $p_e = 10^{-4}$, selection coefficient $s = -0.8$ and mutation rate $\mu = 5$. **f,** Identification of neutrality for simulated pre-therapy tumors. The R^2 represents degree of fitting to power-law distribution. **g-h,** Average CCF of simulated tumors that underwent PS (**g**) and NFDS (**h**), respectively before and after ICB therapy. Box plots show median, quartiles (boxes) and range (whiskers). P values, Wilcoxon rank-sum one-sided test.

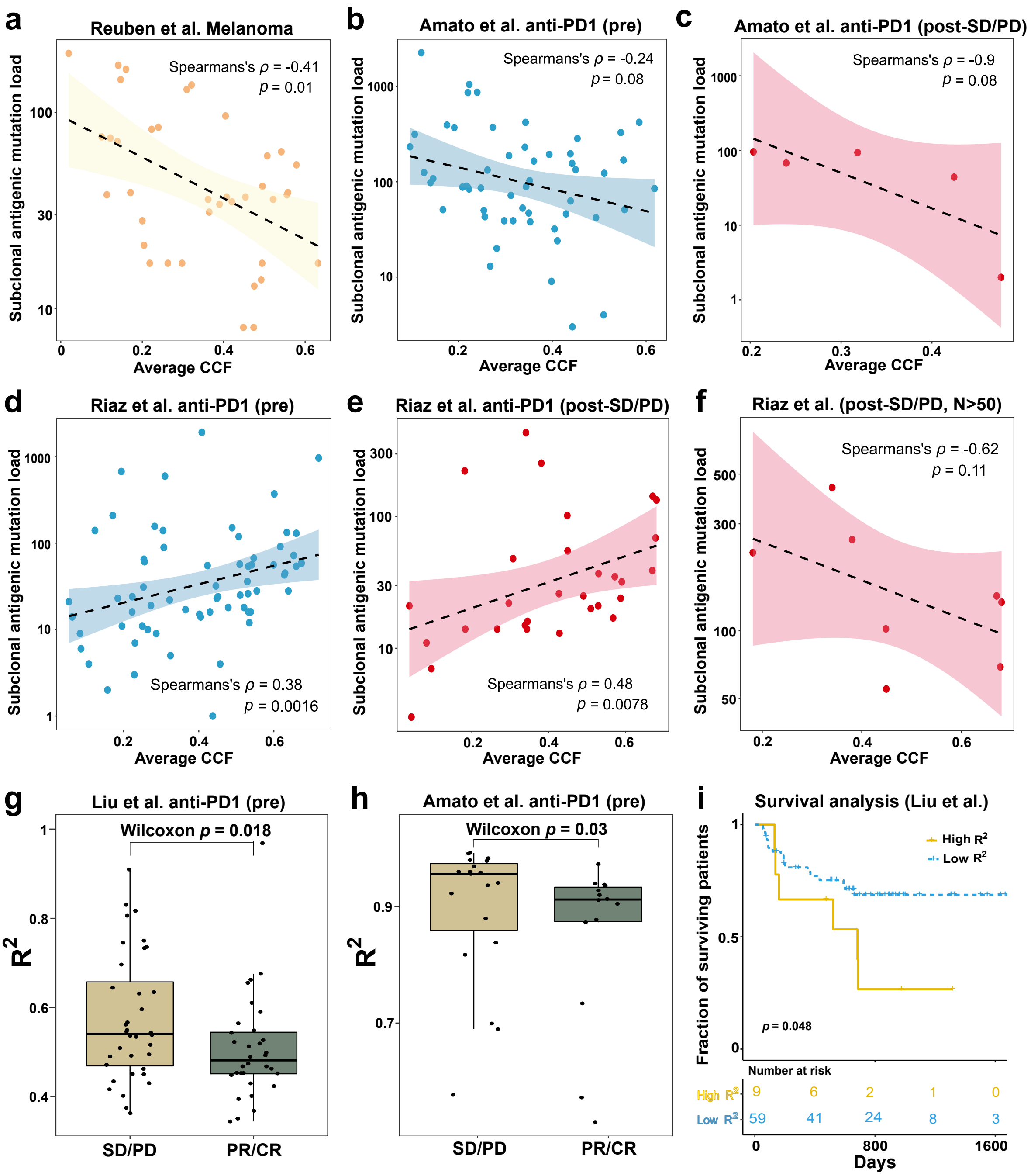

Fig. 4. Testing NFDS with four whole-exome sequencing datasets in melanoma. a-f, Correlation analysis between the average CCF and number of subclonal antigenic mutations for samples from Reuben et al. (**a**), pre-therapy samples from Amato et al. (**b**), post-therapy SD/PD samples from Amato et al. (**c**), pre-therapy samples from Riaz et al. (**d**), post-therapy SD/PD samples from Riaz et al. (**e**), and samples with high subclonal neoantigen load (more than 50) from Riaz et al. (**f**). The line indicates the linear regression and the shading indicates the 95% CI of the regression. **g-h,** Power-law model of neutrality fitting for responders (PR/CR) vs non-responders (SD/PD) from Liu et al.' s cohort (**g**) and Amato et al.' s cohort (**h**), respectively. Box plots show median, quartiles (boxes) and range (whiskers). P values, one-sided Wilcoxon rank-sum test. **i,** Survival analysis of patients from Liu et al.' s cohort stratified by neutrality fitting score R^2 (high, $R^2 \geq 0.7$; low, $R^2 < 0.7$).