Analysis of mutation, selection, and epistasis: an informed approach to cancer clinical trials

Jon F. Wilkins[1], Vincent L. Cannataro[2], Brian Shuch[3,4] and Jeffrey P. Townsend[2,5,6]

[1] Ronin Institute, Montclair, NJ, USA

[2] Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, CT, USA

[3] Department of Urology, Yale School of Medicine, New Haven, CT, USA

[4] Department of Radiology, Yale School of Medicine, New Haven, CT, USA

[5] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

[6] Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA

Correspondence to: Jeffrey P. Townsend, *email*: Jeffrey.Townsend@Yale.edu

ABSTRACT

Currently, drug development efforts and clinical trials to test them are often prioritized by targeting genes with high frequencies of somatic variants among tumors. However, differences in oncogenic mutation rate—not necessarily the effect the variant has on tumor growth—contribute enormously to somatic variant frequency. We argue that decoupling the contributions of mutation and cancer lineage selection to the frequency of somatic variants among tumors is critical to understanding—and predicting—the therapeutic potential of different interventions. To provide an indicator of that strength of selection and therapeutic potential, the frequency at which we observe a given variant across patients must be modulated by our expectation given the mutation rate and target size to provide an indicator of that strength of selection and therapeutic potential. Additionally, antagonistic and synergistic epistasis among mutations also impacts the potential therapeutic benefit of targeted drug development. Quantitative approaches should be fostered that use the known genetic architectures of cancer types, decouple mutation rate, and provide rigorous guidance regarding investment in targeted drug development. By integrating evolutionary principles and detailed mechanistic knowledge into those approaches, we can maximize our ability to identify those targeted therapies most likely to yield substantial clinical benefit.

The sequencing of the human genome at the beginning of this century ushered in a revolution in genomic medicine [1, 2], and many promising therapeutic approaches are in development that make primary use of a patient's genotype data, such as genetic markers in tumor cells. The synergistic combination of personalized medicine with molecular therapeutics that target specific cancer drivers has yielded some remarkable successes, but also a string of disappointments [3-5]. More than thirty targeted cancer therapeutics have been approved for clinical use, but the efficacy of these new drugs has proven variable, with limited durable responses seeming to depend on tumor type and secondary biomarkers [6-12]. Therapeutics that specifically target a single gene product—and especially those that target the mutant, oncogenic form—have the potential to produce fewer unwanted effects on non-tumor cells relative to conventional chemotherapeutics that target all dividing cells. For targeted therapeutics to achieve optimal effectiveness, though, requires accurate application of personalized medicine. If a drug targets a specific genetic variant, it will likely only benefit a patient whose tumor is carrying that variant, and in whom that variant is driving tumorigenesis.

Although this "matching" of targeted therapeutic to a precision-medicine profiled patient is straightforward—

in principle—the list of potential molecular targets (and drugs targeting them) continues to expand [13, 14], as does our understanding of the frequency at which specific genetic variants are found in various cancers [15, 16] and our understanding of the genetic heterogeneity of tumors [17-19]. This expansion of targets across cancer types naturally generates a huge number of plausible hypotheses: specific patient populations who might benefit from specific targeted therapies, or combinations of therapies [20, 21]. As it is not possible to test every mutation type in every cancer, a tissue-agnostic approach bringing an agent forward to the clinic may be necessary once there is strong preclinical rationale that inhibiting a specific pathway has efficacy in a different cancer model. Testing the hypothesis of efficacy based on a different preclinical model with strong laboratory and genomic evidence still requires subsequent clinical trials [22].

As the numbers of candidate therapies and putative targets increase, the development of theoretical frameworks providing rational, quantitative methods for identifying drug/target combination becomes increasingly important. The number of ideas that can be defended on the grounds of plausibility is large, but the fraction of these ideas that result in superior therapies is likely to be small. Not only considerations of cost- and time-effectiveness, but also moral considerations regarding patient welfare, demand that we develop approaches that have predictive value regarding the potential success or failure of targeted drug development [23], derived from what we know about the genetic architecture of cancer.

While precision medicine tumor boards and programs like NCI-MATCH [24] explicitly aim to pair patients with treatment options based on the tumor genomic profile, we feel that these and future efforts can be enhanced by informing decisions with all available sources of data and insight on cancer progression and treatment. One such additional source of information potentially derives from taking an evolutionary perspective on tumor progression [25, 26], employing tools and insights from population genetics. In this paper, we discuss the value and importance of considering the strength of selection favoring the persistence and spread of specific high-fitness genetic variants within the context of the heterogeneous tumor environment. "Selection" or "fitness" of somatic genetic variants in this context refers to the differential survival and reproduction of cell lineages within the patient, and not the fitness of the patient. Estimates of the strength of this selection suggest an upper bound for the potential efficacy of therapies targeting oncogenic mutations, before pre-clinical data can be obtained that provides pharmacokinetic guidance or biologically effective doses (e.g. [27, 28]). We also discuss how estimates of selection will be affected by variation in the oncogenic mutation rate and by epistasis among oncogenic mutations.

For newly developed targeted therapies, it is tempting to focus on oncogenic targets found in a high fraction of patients (although some drugs work well for targets found in a low fraction). This approach is sensible not only because there is a larger patient population available and because the required accrual for a clinical trial is easier, but also for effectiveness and impact. First, a variant found at high frequency in a particular type of tumor is likely to play a causal role in tumor growth. Second, should the clinical trial prove effective, a therapy targeting a more common variant will have a greater downstream impact on public health. For example, mutations in the *BRAF* gene are found in over 50% of melanomas [29], which made this cancer an attractive candidate for clinical trials on therapies that target *BRAF*-mutant tumors via the oncogenic variant gene product. One decision rule for the execution of novel clinical trials is to test therapies that have proven effective against one cancer by applying them to other tumor types [30]. As mentioned above, there may not be a sufficient preclinical model to test the efficacy in the laboratory for a particular cancer in which extensive testing suggests benefit. In this situation, there may be merit to a tissue-agnostic approach, as it focuses on drugs with *in vivo* anti-tumor activity and known side effects, in contrast with early stage drugs with unknown safety profiles. But if an anti-*BRAF* therapy proves effective against melanoma, should we expect that effectiveness to carry over to patients with *BRAF* mutations in other cancers in which *BRAF* mutations are found at much lower frequencies?

The "causal-role" logic of the argument to perform clinical trials that focus on therapeutics targeting high-frequency somatic variants argues against trials of their efficacy in cancers in which the variants are rarely found. But there is more guidance that can be derived from the somatic genetic architecture of cancer types than this simple logic would dictate. The frequency at which genetic variants are found in particular cancers is determined by two factors: 1) the rate at which oncogenic mutations arise, and 2) the effect of natural selection on those mutations within the population of cancerous (or pre-cancerous) cells. Both mutation rate, which varies with gene location [31] and gene expression levels [32], and mutation target size, vary tremendously across the genome [16]. Strength of selection is also enormously variable across the genome and ontogeny [33], highly dependent on the tissue type and cellular environment (including the presence of anti-cancer therapy), and complicated by epistatic interaction among oncogenic variants [34]. Decoupling the contributions of mutation and selection to the frequency of somatic variants is critical to understanding—and predicting—the therapeutic potential of different interventions within individual cancer types. The potential benefit of a therapeutic will relate to the strength of selection favoring the oncogenic variants, and will generally be unrelated to the rate at which the oncogenic mutations occur. However, the observable

data—the frequency of a variant in the context of a given cancer—is affected by both the strength of selection and rate of mutation. Thus, estimations of selection, and the associated predictions regarding therapeutic benefit, require an appropriate accounting for any variation in the mutation rate.

VARIATION IN THE ONCOGENIC MUTATION RATE

The per-gene mutation rate varies across the genome by more than an order of magnitude. This extensive variation is attested to and accounted for by established tools like MutSigCV [16] that aim, by correcting for the mutation rate, to help identify the genes burdened with mutations in a higher-than-expected fraction in a particular cancer. The use of a well-estimated mutation rate to identify genes contributing to cancer represents an important first-order correction necessary for disentangling the effects of mutation and selection. In addition to the variation in mutation rate across the genome, there is substantial variation in the rates of specific types of mutations by nucleotide and cell type. For instance, G→A and C→T transitions are more common than A→T and T→G transversions due to biochemical properties of the nucleotides, and lung cancers have a characteristically high G→T transversion incidence attributed to tobacco-specific mutagens [35].

Equally important for decoupling the effects of mutation and selection is understanding the variation in the oncogenic mutation target size. For genes where loss of function is oncogenic, there will be a variety of missense, nonsense, deletion, and insertion mutations that will disrupt gene function and drive tumorigenesis. However when the druggable target is the result of an activating missense mutation, activation likely depends on a small number of very specific mutations. For example, of the approximately 50% of melanoma patients whose tumors carry mutations in *BRAF*, nearly 90% have a substitution at amino acid position 600, and 90% of these substitutions are V600E, which occurs only from a nucleotide position 1799 T → A substitution [36]. Transition mutations occur at roughly twice the frequency of transversions [37]. At position 1799 T, approximately 50% of new mutations would be T → C transitions while T → G and T → A transversions would account for 25% each. Thus, the most common substitution, observed in over 40% of melanomas, results from the least probable point mutation at that single nucleotide position . This combination of high frequency and mutational rarity implies that the V600E substitution must confer a high selective advantage in melanoma tumor cell populations.

Other oncogenic mutations have a broader distribution of substitutions at a single site. For instance, approximately 20% of melanoma patients have tumors carrying mutations in *NRAS*, and of these approximately

80% have amino acid substitutions at position 61, including Q61R (38%), Q61K (24%), and Q61L (15%) [38]. Like *BRAF* V600E, the Q61K and Q61L substitutions result from transversion mutations (181 C → A and 182 A → T, respectively). The Q61R substitution results from a transition mutation (182 A → G). Thus the *BRAF* V600E substitution is observed at twice the frequency of *NRAS* Q61 substitutions in melanoma—despite the fact that oncogenic mutations at *NRAS* Q61 likely arise at a higher frequency, both because there are more oncogenic mutations available (Q61K, Q61L, and Q61R, *versus* V600E), and because one of those mutations relies on the more frequent transition process. These ratios imply that the impact of selection on the frequency of observing the *BRAF* V600E substitution is many times greater than the impact of selection on the frequency of the common oncogenic *NRAS* Q61 substitutions. The relative maximum therapeutic benefits of targeting *BRAF* and *NRAS* should scale similarly.

This example comparing *BRAF* and *NRAS* in melanoma illustrates a general point that, in addition to the variation among genes in the per-nucleotide mutation rate, there are at least two other sources of variation that affect what we might call the "oncogenic mutation rate," meaning the rate of occurrence of mutations that actually contribute to cancer growth. The first source of variation is the number of distinct substitutions that result in an oncogenic phenotype. The second is the difference in rates of different types of mutational processes (e.g., transitions *versus* transversions). The frequency at which a gene is observed as a high-frequency variant in tumor tissues of a given cancer depends both on the rate at which oncogenic mutations occur and on the selective advantage that they confer. For a variant occurring at a given frequency, the rarer the mutation, the greater the selective advantage the alteration must have conferred. The potential for variation in the oncogenic mutation rate becomes even more pronounced when we consider mutation processes other than nucleotide substitutions, such as deletions that remove entire exons, or translocations that give rise to specific fusion proteins. The rates at which these other types of mutations occur are generally less well known than nucleotide substitution rates, but understanding them is critical for our interpretation of the frequencies at which the resulting variants are found in different cancers.

SELECTION, THERAPEUTIC POTENTIAL, AND EPISTASIS

If a particular mutation seldom occurs, yet is found at high frequency among cancers, this contrast implies that the mutation confers a strong positive selective advantage to the cells possessing it. An intervention that specifically targets that mutation, and thereby removes that selective advantage, would have a substantial therapeutic impact. The mutation rate for the *BRAF* V600E mutation described

above is very low, but the mutation is nonetheless present in tumors at an incommensurate frequency in primary skin cutaneous melanoma and colon adenocarcinoma, where it is a known driver, and in low-grade glioma and lung adenocarcinoma where it has been calculated to be intensely selected [39]. It is not surprising then, that therapies that specifically target that variant are particularly effective for patients with melanoma [40], at least in the short term. At the other end of the spectrum are examples like *TTN*, which, with its ~33,000 amino acids, is the largest known human protein. *TTN* has shown up in many tumor sequencing projects as a putative cancer driver but, as a structural component of muscle sarcomeres, is widely discredited as such. Rather, its extreme size means that it presents an extraordinarily large

mutational target; tumors will often contain one or more *TTN* mutations, even if little or no selection is involved [16]. The extent to which natural selection favors a given oncogenic variant is directly connected to that variant's contribution to cancer progression. Within a population of cancer (or pre-cancerous) cells, those genotypes that lead to the cell dividing more frequently—and those that lead to the cell undergoing cell death or terminal differentiation less frequently—increase in frequency in the same kind of evolutionary process that leads to the spread of fitness-enhancing variants in cellular populations. The fitness associated with a new mutation is directly related to its enhancement of the cancer phenotype within its current environment. And the frequency at which we observe a given variant across patients—modulated by our
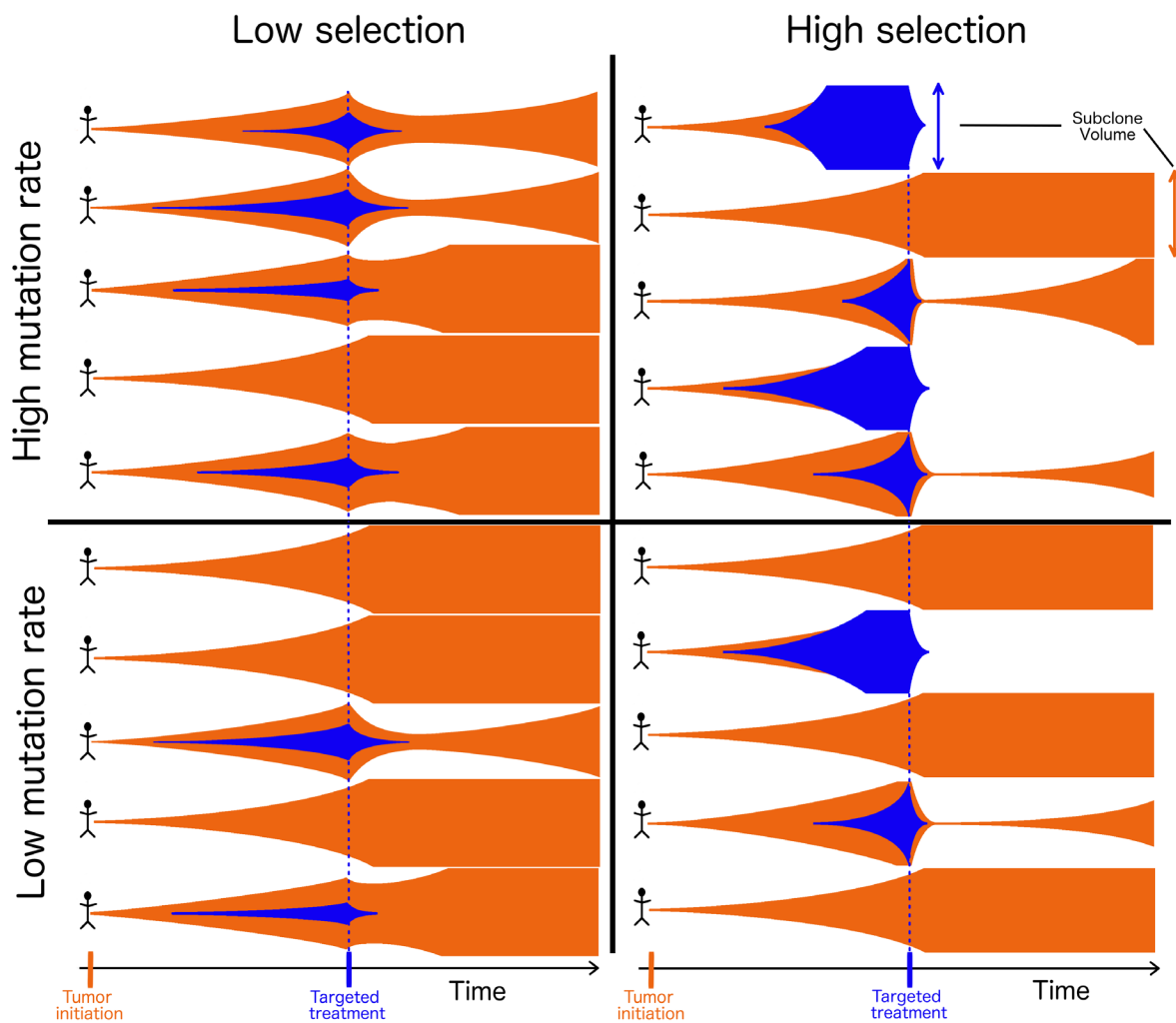
Figure 1: Tumor evolution and treatment under high and low mutation and high and low selection scenarios. Each tumor is initiated by an original cell population (orange), and populations expand in size (vertical axis) over time (horizontal axis). As the tumor grows, a new subclone (blue) can be founded via a mutation that confers a more extensive cancer phenotype, and can grow alongside or supplant the original population. Targeted treatment (blue line) specifically eliminates the new population that it targets (blue), leading to a varying consequence on the tumor that depends on the extent of intra-tumor heterogeneity, which is determined by its selective effect. Tumors with a high proportion of their cells belonging to the population affected by treatment are most likely to experience the greatest reduction of tumor burden, the longest remission before recurrence, and the greatest opportunity for cure.

expectation given the mutation rate and target size—is an indicator of that strength of selection. That is, an oncogenic driver mutation is simply a "beneficial" mutation that is under positive selection within the population of tumor cells.

Tumors exhibit some degree of intra-tumor heterogeneity. It is this heterogeneity that provides the context within which selection can act; the greater fitness conferred by particular mutation, the more common it will become within that heterogeneous population. Put another way, if we observe that a particular variant is very common among patients with a tumor type—and/or very common within heterogeneous tumor-cell populations—this high frequency suggests that the variant has a large effect on the ability to grow, divide, survive and indeed, to contribute to any of the hallmarks of cancer [41] other than increasing mutation rate. Consequently, if we have a therapeutic agent that inactivates that variant, we expect it to have a correspondingly large and negative impact on cancer cells within the tumor. If, however, that same variant is rare in a second tumor type, then that is an indication that the mutation provides less of a selective advantage in that cell type. Consequently, counteracting that mutation with a targeted drug would likely have a less severe impact on the ability of cancer cells to thrive, and a lower salutary impact on the patient (Figure 1). Clearly, for heterogeneous tumors, the efficacy of any targeted therapy will be limited to the subset of cells expressing the appropriate target. Thus, targeting oncogenic driver mutations that are more strongly favored by selection will also tend to lead to the greatest reductions in tumor volume.

The strength of selection favoring a particular genetic variant will predict not only the potential efficacy of targeted therapeutics, but also the strength of selection favoring the evolution of resistance following the application of those therapeutics [42]. The fitness of the mutation within the tumor environment is dependent on the context within which mutations and therapies interact. The introduction of any therapy alters that landscape, creating new selective pressures favoring mutations that interfere with the therapy or bypass its locus of action [43]. Moreover, other factors can modulate the selective advantage conferred by a mutation. For instance, the *BCR-ABL1* fusion is present in 90–95% of patients with chronic myeloid leukemia [44]. This fusion arises from a t(9;22) (q34.1;q11.21) translocation that results in the Philadelphia chromosome [45]. While the exact rate at which this translocation spontaneously occurs is unknown, it appears that it occurs frequently in white blood cells of healthy individuals [46]. The translocation is considered a major driver of chronic myeloid leukemia, apparently because as cellular fitness declines with age or potentially other influences, the translocation becomes increasingly strongly favored by natural selection [47, 48]. It would be no coincidence, then, that therapeutics specifically targeted

against the BCR-ABL fusion protein (e.g., imatinib) have provided an outstanding example of successful targeted chemotherapy, potentially providing patients with long-term health outcomes similar to those of the general population [49, 50]. The substantial rate of evolution of resistance and recurrence is consistent with imatinib imposing a strong selective pressure.

The strength of natural selection favoring each genetic variant in each cancer type can be inferred using theoretical approaches and tools derived from population genetics [39, 42]. Then it can be used as a predictor for the efficacy of therapeutics targeting specific oncogenic variants in different cancers or in individual molecular subsets of a cancer type defined by epigenetic states or cellular context [51]. However, estimates of selection intensity based solely on somatic variant frequency and mutation rate will have to be interpreted with caution, due to the complicating effects of epistasis—the phenomenon in which the strength of selection acting on a mutation depends on the presence or absence of other mutations elsewhere in the genome. Epistasis is known to be present among causal somatic variants in cancer, usually identified in cancer genomics by mutual exclusivity or co-occurrence of mutations [52-55]. However, statistical identification of epistatic interactions typically requires strong epistasis and large sample sizes. Fortunately, mechanistic knowledge of the biochemical interactions in the cell and the signaling pathways where dysregulation often leads to cancer provides a way to focus a search for epistasis in principled ways [55, 56]; in turn, knowledge of epistatic interactions can serve as a further guide to the potential efficacy of a targeted therapeutic agent. This knowledge can facilitate more accurate characterization of selection strengths acting on individual genes and a better understanding of how those selection strengths vary based on other genetic markers.

To understand the effect of epistatic interactions on the potential efficacy of a targeted therapeutic, it will help to categorize interactions as either "antagonistic" or "synergistic". When the selective advantage of two mutations is less than expected from the advantages conferred by the single mutations, the interaction is antagonistic. Antagonistic epistasis affecting selection on oncogenic mutations manifests in patterns of mutual exclusivity. Two different mutations may be favored by selection separately. However, with antagonistic epistasis, the presence of one mutation reduces or eliminates the selection favoring the other one. As a result, while we would find each mutation at high frequency in cancer cells, rarely would we find both mutations in the same cell. Biochemically, antagonistic epistasis can arise when one gene product is downstream from another in a signaling pathway—though other antagonistic interactions are known [57].

One example is the interaction between activating mutations of *EGFR* and downstream *RAS* genes.

Oncogenic mutations in these genes result in constitutive activation of a mitogenic signaling pathway, but only one of the two mutations is required to achieve an oncogenic effect. If a mutation has activated *EGFR*, activating *RAS* mutations have little additional growth effect, and do not spread through the population of cancer cells. Likewise, if a downstream *RAS* becomes constitutively active and no longer dependent on upstream activation, mutations in *EGFR* will exert no downstream effect, and no selective advantage. For example, in lung adenocarcinoma, ~30% of tumors carry a mutation in the *KRAS* gene and ~15% carry a mutation in *EGFR*, but these mutations are mutually exclusive [58, 59]. If antagonistic epistasis is sufficiently strong, the presence of both mutations together can actually be deleterious (disfavored by selection), even though each is beneficial (favored by selection) individually (Figure 2). In fact, there is evidence from transgenic mouse models and cancer cell lines that simultaneous activating mutations in *KRAS* and *EGFR* lead to cell death in lung adenocarcinoma [60]. Despite differences in their roles in signalling and in therapeutic implications, a similar mutually exclusive pattern of mutation extends to KRAS and downstream BRAF [61], and extends further down mitogen-activated protein kinase pathways.

Existing methods can estimate the intensity of selection acting on specific genes and variants within a particular cancer type [39, 42]. The most straightforward application of those methods—and possibly the only statistically supportable one—estimates selection under the assumption that there is no epistasis. To the extent that antagonistic epistasis is present, assuming it is not would lead to underestimation of the level of selection on the antagonistic mutations and thus underestimation of the potential clinical benefit of targeted therapy. Assuming no epistasis effectively yields averages over the different genetic contexts in which the mutation may arise: weak or even negative selection in the presence of the epistatic partner and strong positive selection in the absence of that partner. However, under antagonistic epistasis, most or all tumor tissues with one somatic variant of an epistatic pair will lack the partner variant, leading to a situation where no matter what the selection coefficient on acquisition of the partners are, neither can reach the highest frequencies across tumors that mutations without antagonistic epistasis can. The somatic variant frequency across tumors stalls at a virtual "ceiling" created by the incidence of its epistatic somatic variant; the strength of selection will be underestimated, even after fully accounting for the oncogenic mutation rate. The clinical benefit of effective targeted therapy for these patients will be enhanced relative to the benefit predicted from the average selective strength inferred under an assumption of no epistasis.

Patterns of antagonistic epistasis or mutual exclusivity can furthermore provide insight into possible modes of drug resistance due to selective pressure by a targeted agent [62]. These mutational patterns indicate alternative molecular mechanisms for achieving the same oncogenic cellular phenotype. For example, *RAS* mutations (most commonly at the G12 and G13 sites) represent a common form of evolved resistance to therapies targeting *EGFR* in metastatic colorectal cancer [63]. The strong positive selection on *EGFR*-activating mutations predicts a substantial clinical response to treatments targeting the activated protein. However, due to epistasis, this treatment has a side-effect of altering the strength of selection on other genes—e.g., making activating *RAS* mutations that would have been lethal in the presence of uninhibited EGFR become strongly beneficial for the cancer lineage(s) in the presence of targeted EGFR treatment. Thus, understanding the patterns of epistasis in these cells can not only provide a better ability to identify potential drug targets, but also to predict the likely modes of resistance to future therapies.

In contrast to antagonistic epistasis, the selective advantage of two mutations can also be greater than expected from the advantages conferred by the single mutations, in which case the interaction is synergistic. Synergistic epistasis in cancer manifests in patterns of co-occurrence. In the case of two synergistic mutations, each of them on their own is favored only modestly by selection (and they are only modestly oncogenic). But when both mutations are present together, selection on the pair is strong. That is, selection strongly favors one mutation only in the presence of the other. Thus, finding either mutation alone will be rare, but the pair of mutations will be highly enriched. The first case of synergistic epistasis described for cancer was between RAS and MYC, when researchers demonstrated that full tumorigenic conversion of fibroblasts required introduction of both oncogenes [64, 65]. Other genes have been found to also interact synergistically with RAS, as oncogenes promoting hyper-proliferation such as RAS can also promote cell-induced
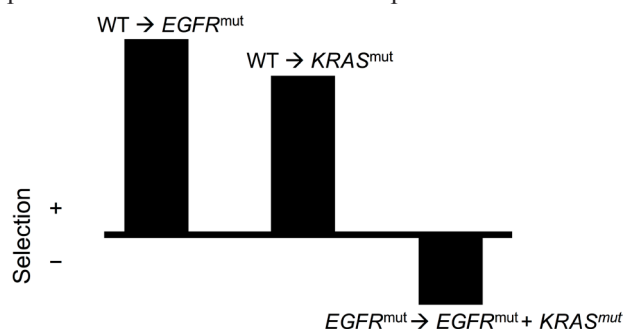
Figure 2: Illustration of negative epistasis between oncogenic EGFR and KRAS mutations in lung tissue. Individually, EGFR and KRAS mutations both confer a survival and/or reproductive benefit to lung cells, and are thus positively selected. However, if a cell contains both oncogenic mutations, there is a decline in survival and/or reproduction. Therefore, the double mutant is subject to negative selection—in this case, an epistatic effect of the possession of one mutation upon the selection imposed on the second mutation.

senescence [66], and negation of this senescence by other mutations in genes, such as p53, exacerbates cellular proliferation [67]. A similar synergistic mechanism involving p53 was also discovered for BRCA2 [68] and BRCA1 [69], FGFR2 [70] and for amplification of other genes [71] including ERBB2. This well-described mechanistic epistasis can give rise to a more abstract statistical epistasis: mutations in *TP53* are found nearly twice as often in ovarian cancers with *BRCA1* or *BRCA2* mutations compared with tumors with wild-type *BRCA1* and *BRCA2* [72, 73]. This pattern of co-occurrence is a consequence (and an indicator) of the fact that the selective advantage of an loss-of-function *TP53* mutation depends on the genotype at these other loci. We would expect that loss-of-function mutations in *TP53* and other apoptosis genes should be positively epistatic (synergistic) with activating mutations in oncogenes.

As was the case with antagonistic epistasis, the presence of synergistically epistatic interactions means that targeted therapies could provide greater clinical benefit to the appropriately targeted population than would be predicted on the basis of estimates of selection that assume no epistasis. An estimate of selection on *TP53* would average over cases with and without *BRCA* mutations. Therapies targeting *TP53*-mutant cells would be most effective in cases where those cells also carried mutations in *BRCA1* and/or *BRCA2*. Tumors exhibiting a *TP53* mutation would be enriched for exactly those mutations. The patient population toward which such a targeted treatment should be directed would constitute those who exhibit the common pattern of co-occurrence.

Synergistic epistasis is at the core of the multiple-hits model of cancer progression. Implicit in the model that you need a certain number of mutations before you get cancer progression is the idea that selection on individual variants is weak, and only becomes strong when multiple oncogenic variants are present simultaneously. The model also implies that targeting any one of causal variants should be sufficient to yield a beneficial response—although targeting multiple co-occurring variants might reduce the opportunity for the evolution of resistance to therapy. Similarly, synergistic epistasis may also play a role in "cancer fields", where collections of cells carry potentially oncogenic variants, but do not develop into full-blown cancer until one or more additional mutations arise. In this case, weak selection on the pre-cancerous field cells may contribute modestly to their expansion, but selection does not become strong until an epistatic partner arises through mutation.

Antagonistic and synergistic epistasis not only have distinct implications for the efficacy of therapy, but also often reflect different sorts of biochemical relationships in the cell. Antagonistic epistasis can arise between genes in linear or linear-like pathways with upstream-downstream relationships, whereas synergistic epistasis can arise between genes with partially overlapping or redundant functionality. Nevertheless, the two forms of epistasis have similar implications for how we should interpret the relationship between estimates of the strength of selection favoring an oncogenic variant and the potential clinical benefit of therapies targeting that variant. Due to limitations imposed by small sample sizes, it is often viewed as unavoidable in analyses of genome-scale data to assume that there is no epistasis. However, for variants that are subject to strong epistasis (antagonistic or synergistic), the selective advantage in the clinically relevant genetic context will be greater than a no-epistasis estimate would indicate. Thus, the potential clinical benefit of therapeutic targeting of genes that have epistatic partners is higher than would be predicted by an analysis that assumes the effects of therapeutic targeting of each gene are independent.

CONCLUSIONS

The development of increasing numbers of cancer therapeutics designed for specific molecular targets can provide significant efficacy in appropriately selected populations, as well as a reduction in side effects. However, this burgeoning multiplicity of treatment options simultaneously leads to a large number of potential combinations of therapies that each can be tested on a large number of classifications of tumor types often where hypothesis testing may not be feasible in the laboratory setting. When sufficient preclinical evidence exists for efficacy with a useful agent, each novel therapy or combination of therapies will need to be validated in clinical trials before general adoption, putting high strain on the resources required for drug development and for conducting clinical trials. Ethical considerations for the well-being of patients and financial considerations arising from the high up-front costs of a well-powered clinical trial demand that we develop systematic approaches that have predictive power to identify those therapies and therapeutic combinations that are most likely to be successful.

When treatments site-specifically target and inhibit a gain of function caused by an oncogenic mutation the potential of a targeted therapy is linked to the capacity of a targeted variant to drive tumor growth. That capacity is related most directly not to the frequency of that variant among tumors, but rather to the selective advantage associated with the oncogenic variant. Provided that an effective and tolerable pharmacological agent can be developed, those variants that provide the greatest selective advantage provide the greatest potential as promising therapeutic targets. Disruption of these targets will have the greatest impact on the growth and survival of tumor cells. To accurately estimate the selective advantage of oncogenic mutations requires accounting not only for the variant frequency across tumors, but also for variation in the underlying oncogenic mutation rates.

Mutation rates vary not only across the genome and but also among different types of mutation. Moreover, genes vary in the number and types of mutations that will give rise to oncogenic variants. Accurately accounting for these sources of variation will substantially enhance our ability to identify and target causal oncogenic variants.

Attempts to estimate selection are also complicated by antagonistic and synergistic epistasis, which are likely widespread among oncogenes and tumor suppressors. Due to limited sample sizes, epistasis is often challenging to identify conclusively on purely statistical grounds. However, knowledge of the biochemical relationships among gene products of interest can help to identify likely epistatic interactions. Where those biochemical relationships suggest epistasis, the selective advantage of oncogenic variants and the therapeutic opportunity is likely to be much greater than would be expected by genomic analyses of underlying mutation rate and variant frequency alone. Thus, all else being equal, a selectively favored molecular target—especially one suspected of having strong epistatic interactions—represents a promising candidate. Quantitative approaches need further development that use the known genetic architectures of cancer types to provide rigorous guidance regarding investment in prospective clinical trials. By integrating evolutionary principles and detailed mechanistic knowledge into those approaches, we will be able to maximize our ability to identify those combinations of targeted therapies and cancer types most likely to yield substantial clinical benefit.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose associated with publication of this research.

FUNDING

REFERENCES

1. Martin SD, Coukos G, Holt RA, Nelson BH. Targeting the undruggable: immunotherapy meets personalized oncology in the genomic era. Ann Oncol. 2015; 26: 2367-74.

2. Shekarian T, Valsesia-Wittmann S, Caux C, Marabelle A. Paradigm shift in oncology: targeting the immune system rather than cancer cells. Mutagenesis. 2015; 30: 205-11.

3. Simon R, Roychowdhury S. Implementing personalized cancer genomics in clinical trials. Nat Rev Drug Discov. 2013; 12: 358-69.

4. Huang M, Shen A, Ding J, Geng M. Molecularly targeted cancer therapy: some lessons from the past decade. Trends Pharmacol Sci. 2014; 35: 41-50.

5. Scagliotti G, von Pawel J, Novello S, Ramlau R, Favaretto A, Barlesi F, Akerley W, Orlov S, Santoro A, Spigel D, Hirsh V, Shepherd FA, Sequist LV, et al. Phase III Multinational, Randomized, Double-Blind, Placebo-Controlled Study of Tivantinib (ARQ 197) Plus Erlotinib *Versus* Erlotinib Alone in Previously Treated Patients With Locally Advanced or Metastatic Nonsquamous Non-Small-Cell Lung Cancer. J Clin Oncol. 2015; 33: 2667-74.

6. Köhne CH. Successes and Limitations of Targeted Cancer Therapy in Colon Cancer. Progress in Tumor Research. 2014; pp. 36-50.

7. Casali PG. Successes and Limitations of Targeted Cancer Therapy in Gastrointestinal Stromal Tumors. Progress in Tumor Research. 2014; pp. 51-61.

8. Curigliano G, Criscitiello C. Successes and limitations of targeted cancer therapy in breast cancer. Prog Tumor Res. 2014; 41: 15-35.

9. Damia G, Sessa C. Successes and limitations of targeted cancer therapy in ovarian cancer. Prog Tumor Res. 2014; 41: 89-97.

10. Romano E, Michielin O. Successes and limitations of targeted cancer therapy in melanoma. Prog Tumor Res. 2014; 41: 78-88.

11. Suda K, Mitsudomi T. Successes and limitations of targeted cancer therapy in lung cancer. Prog Tumor Res. 2014; 41: 62-77.

12. Pracht M, Berthold D. Successes and Limitations of Targeted Thera pies in Renal Cell Carcinoma. Progress in Tumor Research. 2014; pp. 98-112.

13. Rubio-Perez C, Tamborero D, Schroeder MP, Antolín AA, Deu-Pons J, Perez-Llamas C, Mestres J, Gonzalez-Perez A, Lopez-Bigas N. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell. 2015; 27: 382-96.

14. Hollingsworth SJ. Precision medicine in oncology drug development: a pharma perspective. Drug Discov Today. 2015; 20: 1455-63.

15. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2014; 43: D805-11.

16. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499: 214-8.

17. Zhao ZM, Zhao B, Bai Y, Iamarino A, Gaffney SG, Schlessinger J, Lifton RP, Rimm DL, Townsend JP. Early and multiple origins of metastatic lineages within primary tumors. Proc Natl Acad Sci U S A. 2016; 113: 2140-5.

18. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013; 501: 338-45.

19. Rosenthal R, McGranahan N, Herrero J, Swanton C. Deciphering Genetic Intratumor Heterogeneity and Its Impact on Cancer Evolution. Annual Review of Cancer Biology. 2017; 1: 223-40.

20. Zhao B, Pritchard JR, Lauffenburger DA, Hemann MT. Addressing genetic tumor heterogeneity through computationally predictive combination therapy. Cancer Discov. 2014; 4: 166-74.

21. Zhao B, Hemann MT, Lauffenburger DA. Intratumor heterogeneity alters most effective drugs in designed combinations. Proc Natl Acad Sci U S A. 2014; 111: 10773-8.

22. Redig AJ, Jänne PA. Basket Trials and the Evolution of Clinical Trial Design in an Era of Genomic Medicine. J Clin Oncol. 2015; 33: 975-7.

23. Fischer HP. Towards quantitative biology: integration of biological information to elucidate disease pathways and to guide drug discovery. Biotechnol Annu Rev. 2005; 11: 1-68.

24. Conley BA, Doroshow JH. Molecular analysis for therapy choice: NCI MATCH. Semin Oncol. 2014; 41: 297-9.

25. Somarelli JA, Ware KE, Kostadinov R, Robinson JM, Amri H, Abu-Asab M, Fourie N, Diogo R, Swofford D, Townsend JP. PhyloOncology: Understanding cancer through phylogenetic analysis. Biochim Biophys Acta. 2017; 1867: 101-8.

26. Almendro V, Marusyk A, Polyak K. Cellular Heterogeneity and Molecular Evolution in Cancer. Annu Rev Pathol: Mech Dis. 2013; 8: 277-302.

27. Srivastava AK, Hollingshead MG, Govindharajulu JP, Covey JM, Liston D, Simpson MA, Peggins JO, Bottaro DP, Wright JJ, Kinders RJ, Doroshow JH, Parchment RE. Molecular Pharmacodynamics-Guided Scheduling of Biologically Effective Doses: A Drug Development Paradigm Applied to MET Tyrosine Kinase Inhibitors. Mol Cancer Ther. 2018 Feb 14. doi: 10.1158/1535-7163.MCT-17-0552. [Epub ahead of print]

28. Srivastava AK, Navas T, Herrick WG, Hollingshead MG, Bottaro DP, Doroshow JH, Parchment RE. Effective implementation of novel MET pharmacodynamic assays in translational studies. Ann Transl Med. 2017; 5: 3.

29. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, et al. Mutations of the BRAF gene in human cancer. Nature. 2002; 417: 949-54.

30. Würth R, Thellung S, Bajetto A, Mazzanti M, Florio T, Barbieri F. Drug-repositioning opportunities for cancer therapy: novel molecular targets for known compounds. Drug Discov Today. 2016; 21: 190-9.

31. Chen CL, Rappailles A, Duquenne L, Huvet M, Guilbaud G, Farinelli L, Audit B, d'Aubenton-Carafa Y, Arneodo A, Hyrien O, Thermes C. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. Genome Res. 2010; 20: 447-57.

32. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordóñez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010; 463: 191-6.

33. DeGregori J. Connecting Cancer to Its Causes Requires Incorporation of Effects on Tissue Microenvironments. Cancer Res. 2017; 77: 6065-8.

34. Misra N, Szczurek E, Vingron M. Inferring the paths of somatic evolution in cancer. Bioinformatics. 2014; 30: 2456-63.

35. Hainaut P, Pfeifer GP. Patterns of p53 G->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. Carcinogenesis. 2001; 22: 367-74.

36. Ascierto PA, Kirkwood JM, Jean-Jacques G, Ester S, Grimaldi AM, Michele M, Giuseppe P, Alessandro T, Marincola FM, Nicola M. The role of BRAF V600 mutation in melanoma. J Transl Med. 2012; 10: 85.

37. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43: 491-8.

38. Jakob JA, Bassett RL Jr, Ng CS, Curry JL, Joseph RW, Alvarado GC, Rohlfs ML, Richard J, Gershenwald JE, Kim KB, Lazar AJ, Hwu P, Davies MA. NRAS mutation status is an independent prognostic factor in metastatic melanoma. Cancer. 2012; 118: 4014-23.

39. Cannataro VL, Gaffney SG, Townsend JP. Effect sizes of somatic mutations in cancer (Internet). bioRxiv. 2018 (cited 2018 Feb 16). https://doi.org/10.1101/229724.

40. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med. 2011; 364: 2507-16.

41. Hanahan D, Douglas H, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011; 144: 646-74.

42. Cannataro VL, Gaffney SG, Stender C, Zhao Z, Philips M, Greenstein AE, Townsend JP. Heterogeneity and mutation in KRAS and associated oncogenes: evaluating the potential for the evolution of resistance to targeting of KRAS G12C. Oncogene. 2018 Feb 16. doi: 10.1038/s41388-017-0105-z. [Epub ahead of print]

43. Blair BG, Bardelli A, Park BH. Somatic alterations as the basis for resistance to targeted therapies. J Pathol. 2013; 232: 244-54.

44. Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature. 1985; 315: 550-4.

45. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified

by quinacrine fluorescence and Giemsa staining. Nature. 1973; 243: 290-3.

46. Biernaux C, Loos M, Sels A, Huez G, Stryckmans P. Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals. Blood. 1995; 86: 3118-22.

47. Henry CJ, Marusyk A, Zaberezhnyy V, Adane B, DeGregori J. Declining lymphoid progenitor fitness promotes aging-associated leukemogenesis. Proc Natl Acad Sci U S A. 2010; 107: 21713-8.

48. Marusyk A, DeGregori J. Declining cellular fitness with age promotes cancer initiation by selecting for adaptive oncogenic mutations. Biochim Biophys Acta. 2008; 1785: 1-11.

49. Gambacorti-Passerini C, Antolini L, Mahon FX, Guilhot F, Deininger M, Fava C, Nagler A, Della Casa CM, Morra E, Abruzzese E, D'Emilio A, Stagno F, le Coutre P, et al. Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib. J Natl Cancer Inst. 2011; 103: 553-61.

50. Kantarjian H, O'Brien S, Jabbour E, Garcia-Manero G, Quintas-Cardama A, Shan J, Rios MB, Ravandi F, Faderl S, Kadia T, Borthakur G, Huang X, Champlin R, et al. Improved survival in chronic myeloid leukemia since the introduction of imatinib therapy: a single-institution historical experience. Blood. 2012; 119: 1981-7.

51. Sieber OM, Tomlinson SR, Tomlinson IPM. Tissue, cell and stage specificity of (epi)mutations in cancers. Nat Rev Cancer. 2005; 5: 649-55.

52. Park S, Lehner B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. Mol Syst Biol. 2015; 11: 824.

53. Szczurek E, Beerenwinkel N. Modeling mutual exclusivity of cancer mutations. PLoS Comput Biol. 2014; 10: e1003503.

54. Yeang CH, McCormick F, Levine A. Combinatorial patterns of somatic gene mutations in cancer. FASEB J. 2008; 22: 2605-22.

55. Gaffney SG, Townsend JP. PathScore: a web tool for identifying altered pathways in cancer data. Bioinformatics. 2016; 32:3688-90.

56. Kim YA, Cho DY, Dao P, Przytycka TM. MEMCover: integrated analysis of mutual exclusivity and functional network reveals dysregulated pathways across multiple cancer types. Bioinformatics. 2015; 31: i284-92.

57. Williams JP, Stewart T, Li B, Mulloy R, Dimova D, Classon M. The retinoblastoma protein is required for Ras-induced oncogenic transformation. Mol Cell Biol. 2006; 26: 1170-82.

58. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511: 543-50.

59. Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, Zakowski MF, Heelan RT, Kris MG, Varmus HE. KRAS Mutations and Primary Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib. PLoS Med. 2005; 2: e17.

60. Unni AM, Lockwood WW, Zejnullahu K, Lee-Lin SQ, Varmus H. Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. Elife. 2015; 4: e06907.

61. Oikonomou E, Koustas E, Goulielmaki M, Pintzas A. BRAF vs RAS oncogenes: are mutations of the same pathway equal? Differential signalling and therapeutic implications. Oncotarget. 2014; 5: 11752-77. https://dx.doi.org/10.18632%2Foncotarget.2555

62. Mina M, Raynaud F, Tavernari D, Battistello E, Sungalee S, Saghafinia S, Laessle T, Sanchez-Vega F, Schultz N, Oricchio E, Ciriello G. Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. Cancer Cell. 2017; 32: 155-68.e6.

63. Misale S, Di Nicolantonio F, Sartore-Bianchi A, Siena S, Bardelli A. Resistance to anti-EGFR therapy in colorectal cancer: from heterogeneity to convergent evolution. Cancer Discov. 2014; 4: 1269-80.

64. Land H, Parada LF, Weinberg RA. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature. 1983; 304: 596-602.

65. Ashworth A, Alan A, Lord CJ, Reis-Filho JS. Genetic Interactions in Cancer Progression and Treatment. Cell. 2011; 145: 30-8.

66. Di Micco R, Marzia F, Angelo C, Sara P, Patrizia G, Chiara L, Catherine S, Massimiliano G, Nuciforo PG, Aaron B, Roberta M, Pelicci PG, di Fagagna FD. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature. 2006; 444: 638-42.

67. Serrano M, Lin AW, McCurrach ME, Beach D, Lowe SW. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell. 1997; 88: 593-602.

68. Connor F, Frances C, David B, Joseph Mee P, Ross GM, Sally S, Elena G, Tybulewicz VLJ, Alan A. Tumorigenesis and a DNA repair defect in mice with a truncating Brca2 mutation. Nat Genet. 1997; 17: 423-30.

69. Liu X, Holstege H, van der Gulden H, Treur-Mulder M, Zevenhoven J, Velds A, Kerkhoven RM, van Vliet MH, Wessels LFA, Peterse JL, Berns A, Jonkers J. Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. Proc Natl Acad Sci U S A. 2007; 104: 12111-6.

70. Cherdyntseva NV, Denisov EV, Litviakov NV, Maksimov VN, Malinovskaya EA, Babyshkina NN, Slonimskaya EM, Voevoda MI, Choinzonov EL. Crosstalk between the FGFR2 and TP53 genes in breast cancer: data from an association study and epistatic interaction analysis. DNA Cell Biol. 2012; 31: 306-16.

71. Nikolsky Y, Sviridov E, Yao J, Dosymbekov D, Ustyansky

V, Kaznacheev V, Dezso Z, Mulvey L, Macconaill LE, Winckler W, Serebryiskaya T, Nikolskaya T, Polyak K. Genome-wide functional synergy between amplified and mutated genes in human breast cancer. Cancer Res. 2008; 68: 9532-40.

72. Ramus SJ, Bobrow LG, Pharoah PD, Finnigan DS, Fishman A, Altaras M, Harrington PA, Gayther SA, Ponder BA, Friedman LS. Increased frequency of TP53 mutations in BRCA1 and BRCA2 ovarian tumours. Genes Chromosomes Cancer. 1999; 25: 91-6.

73. Narod SA, Foulkes WD. BRCA1 and BRCA2: 1994 and beyond. Nat Rev Cancer. 2004; 4: 665-76.