Synthetic sickness or lethality points at candidate combination therapy targets in glioblastoma

Ewa Szczurek, Navodit Misra and Martin Vingron

Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestr. 63–73, 14195 Berlin, Germany

Synthetic lethal interactions in cancer hold the potential for successful combined therapies, which would avoid the difficulties of single molecule-targeted treatment. Identification of interactions that are specific for human tumors is an open problem in cancer research. This work aims at deciphering synthetic sick or lethal interactions directly from somatic alteration, expression and survival data of cancer patients. To this end, we look for pairs of genes and their alterations or expression levels that are "avoided" by tumors and "beneficial" for patients. Thus, candidates for synthetic sickness or lethality (SSL) interaction are identified as such gene pairs whose combination of states is under-represented in the data. Our main methodological contribution is a quantitative score that allows ranking of the candidate SSL interactions according to evidence found in patient survival. Applying this analysis to glioblastoma data, we collect 1,956 synthetic sick or lethal partners for 85 abundantly altered genes, most of which show extensive copy number variation across the patient cohort. We rediscover and interpret known interaction between *TP53* and *PLK1*, as well as provide insight into the mechanism behind *EGFR* interacting with *AKT2*, but not *AKT1* nor *AKT3*. Cox model analysis determines 274 of identified interactions as having significant impact on overall survival in glioblastoma, which is more informative than a standard survival predictor based on patient's age.

Single molecule-targeted therapies, the dominant tool for cancer treatment, have limited efficacy due to toxicity[1] and rapid development of drug resistance.[2–4] Combination therapies based on synthetic sickness or lethality (SSL) are hoped to overcome these difficulties[5] and promise successful treatment strategies.[6,7] The mechanism behind SSL-based therapy is that while targeting individual genes in a given interacting pair has a moderate effect, targeting both either kills, or significantly decreases tumor viability.

Compared to the comprehensive collection of synthetic lethal gene pairs in yeast,[8] the set of known SSL interactions in human cancer is disappointingly small[9] and their identification remains an open problem. Experimental approaches are overwhelmed by the quadratic number of possible pairs, and can only be applied to cell lines.[7] High-throughput studies focus on single, abundantly altered genes (called primary genes), such as *KRAS*,[10] or *PI3K*,[11] and screen through their possible partner genes. Alternatively, a small set of plausible genes is selected for testing, for example, based upon their function in DNA repair.[12,13] Existing predictive computational methods[14–17] require large training datasets of known genetic interactions, that are only available for few simple model organisms.[18] Genome-wide association studies[19–21] are limited to estimating cancer risk associated with certain single-nucleotide polymorphisms in the germline. Conde-Pueyo et al.[22] identify SSL interactions in humans based on evolutionary conservation to yeast, which are likely incomplete, since not all SSL interactions are conserved in such distant species.

Traditionally, the notion of synthetic interaction is based on a comparison of observed to expected fitness.[23] From a general, disease-oriented perspective, we are less concerned with tumor fitness, but rather with how well the patients perform in dealing with cancer and survive. Therefore, here, tumor *performance* is taken as inverse performance of their carrier patients and serves as the basis for detection of SSL interactions. Such an approach is also dictated by the data: at our hand are patient survival information, together with collective measurements across all cells in tumor samples and direct assessment of tumor fitness is not available to us. Although loosening the rigorous, fitness-based notion of synthetic interaction, we gain access to the more realistic tumor context and advantage over studies performed on cell lines.

Cancer Genetics

What's new?

While targeting one gene in cancer therapy has often only moderate effects, targeting a gene pair can lead to synthetic sickness or lethality (SSL), significantly decreasing the tumor's viability and defining new targets of combination therapy. However, the huge number of candidate gene pairs renders experimental testing of their interactions extremely difficult. Here, the authors decipher synthetic interactions directly from integrated multi-level data derived from glioblastoma patients. They search for gene pairs whose alterations are beneficial for the patient's survival and prioritize candidate interactions based on a quantitative score linked to patient survival. This approach provides a useful pre-selection for experimental verification, an important step towards finding new and effective combination therapies.

Our approach follows an intuitive principle that what is "avoided" by tumors may be "beneficial" for patients. We search large collections of tumor data for such pairs of genes and their states that are under-represented in the data given their individual prevalence, and that together coincide with better carrier patient performance, more than expected from their individual occurrences. Identifying SSL interactions therefore requires an integrated analysis of somatic alteration, expression and survival data. The key to finding them lies in particular patterns of genomic alterations and expression that have been observed in tumors.[24–27] For example, somatic mutations of two genes can be *mutually exclusive*, and do not occur simultaneously in the same tumor. Alternatively, expression of one gene can be *concurrent* with a genomic alteration of another, and be either high or low only in those tumors that carry this alteration. Existence of such tumor-specific patterns suggests that they may improve tumor viability. Consequently, states of genes that violate those patterns, like mutually exclusive genes being mutated together, or knock-down of a gene that is highly expressed concurrently with mutation of another, may decrease performance of tumors and thus increase performance of carrier patients.

Our analysis identifies candidate SSL interactions as such pairs of genes that follow either a concurrence or mutual exclusivity pattern, and both have a specific alteration or expression level, which, when occurring together, violates this pattern. Our main contribution is a score that allows ranking of candidate SSL interactions according to how strongly having both genes in their specific state or level (referred to as SSL level) decreases performance of tumors more than expected from having either of them alone at its respective level. We further use Cox modeling to identify SSL interactions likely to be of greatest therapeutic utility, describe their mechanism, and provide the associated SSL levels for independent experimental validation. We make sure that those SSL levels that need to be induced externally can be reached by manipulating gene expression. In this way, our analysis delivers a disease-specific collection of SSL interactions, intended as a carefully preselected input for subsequent experimental verification.

The proposed approach predicted SSL interactions in glioblastoma multiforme, the most common and the most lethal brain tumor.[28] Data stems from the Cancer Genome Atlas (TCGA),[29] a large, stable research network effort that spans the process of cancer sample collection, comprehensive laboratory analysis and database deposition. The short, 1 year median survival of newly diagnosed patients makes the glioblastoma dataset more amenable to the kind of statistical survival analysis performed here. We integrated clinical information available for 508 carrier patients with somatic point mutations of 424 genes in 145 glioblastomas, copy number variation (CNV) for 18,966 genes in 501 glioblastomas, and gene expression of 17,591 genes in 500 glioblastomas, relative to normal tissue. This resulted in a collection of 1,956 plausible SSL interactions, 274 of which were indicated by Cox analysis to have significant impact on overall survival, comparable to or even stronger than the established predictor based on patient's age (Fig. 1a). The analysis predicted the mechanism underlying two known SSL interactions in cancer: TP53 with PLK1 and EGFR with AKT2. Notably, our approach is successful in identifying SSL interactions between genes altered by CNV. To the best of our knowledge, this work is the first computational analysis that identifies SSL interactions from cancer patient data.

Material and Methods

Integrated multilevel data for a total of 577 glioblastomas[29] was accessed from TCGA in a processed form using the cgdsr package in R[30] (dataset version download as of April 2012). After identification of candidate SSL interactions, we ranked them based on interaction support found in patient survival (Fig. 1a). This analysis is explained in detail below.

Collection of candidate SSL interactions

Preselecting primary genes. The analysis starts with collecting genes of primary importance (shortly, primary genes), as assessed by abundances of their alterations in the tumors. To this end, we selected 85 genes that showed a consistent type of somatic point mutations (shortly, PMs in this Section) or CNV in at least 20 glioblastomas from the cohort. We considered only PMs that were non-silent, and applied a statistical approach called Gistic[31] to identify genes targeted by high-level homozygous amplifications or deletions, more frequently than expected by chance. Additionally, we restricted all primary genes identified as altered by CNV to be concordant with their own expression, that is, to have decreased or elevated expression that is consistent with their alteration (Supporting Information).

Concurrence and mutual exclusivity. Partners for the primary genes in the candidate SSL interactions are found by
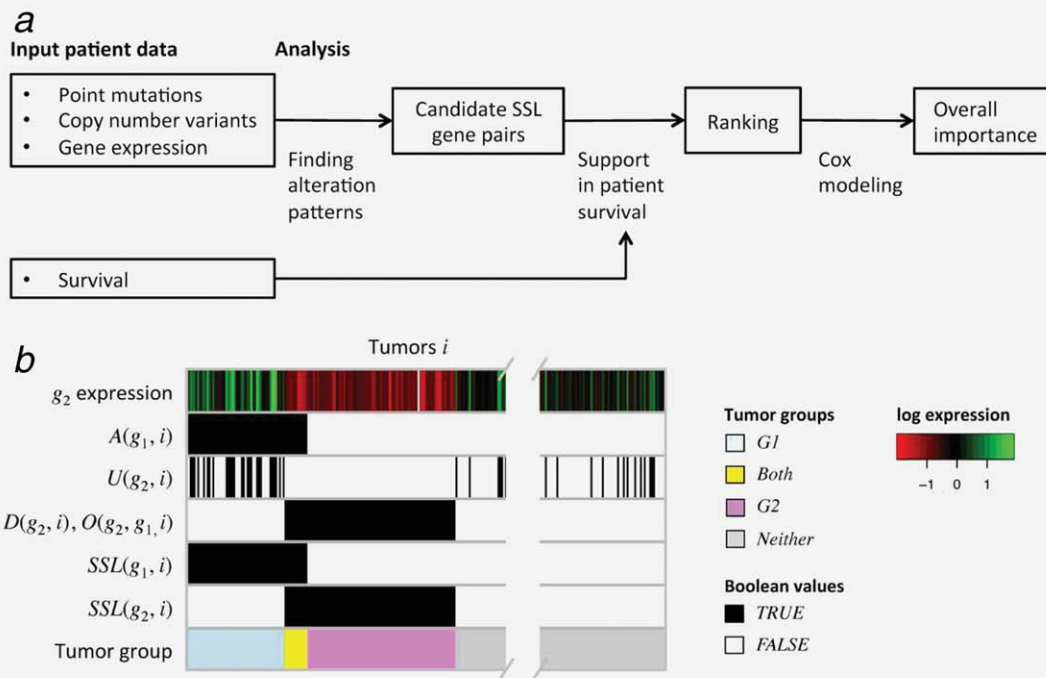
Figure 1. (*a*) Flow chart of our analysis. (*b*) Mutation and expression states of genes, SSL levels and tumor groups. Example of *TP53* (g_1) and *PLK1* (g_2) interaction, fitting scenario I. Columns of the matrix stand for tumors, rows stand for real-valued expression of g_2 (first row), Boolean-valued functions indicating whether one of the genes is on a particular level in each tumor *i* (rows 2–6), and tumor groups (last row). In this example, expression of *PLK1* and somatic point mutations of *TP53* are concurrent: *PLK1* tends to be elevated, and is often *up* ($U(g_2, i)$ is true; third row) in those tumors *i* where *TP53* is mutated (where $A(g_1, i)$ is true, second row). The SSL level of TP53 (fifth row) is equivalent to its *altered* level, and SSL(g_1, i) is equivalent to $A(g_1, i)$. The SSL level of *PLK1* (sixth row) is equivalent to its *opposite*, *down* level (fourth row). Here, the *Both* group (yellow) gathers those tumors where the concurrence pattern observed for *TP53* and *PLK1* is violated, both genes are on their SSL levels, and where *TP53* is altered, but *PLK1* is down.

identifying genomic alteration and expression patterns relating primary genes to each other, or primary genes to other genes, whose expression was measured in the analyzed dataset. First, for each primary gene, we screen all other genes for expression changes concurrent with alteration of this gene. Concurrence in the data is demonstrated by significant increase or decrease of gene expression levels (*i.e.*, concordance) exclusively in those tumors that have either PM or CNV of the altered gene (Supporting Information). Next, we utilize expression of the genes that are concurrent with each primary gene to impute its missing alterations (existence of either PMs, or CNV; Supporting Information). In this way, for the 85 primary genes we obtained a dataset of either true or imputed alteration values in together 447 tumor samples, for which also both expression and clinical data was measured.

Finally, we test each pair of primary genes for mutual exclusivity. We apply a lower-tail hypergeometric test for depletion of intersection between the known or imputed alterations across the set of patients. To select significantly mutually exclusive pairs, we use a Bonferroni-corrected *p*-value threshold 0.05.

Mutation and expression states of genes in tumors. For each primary gene *g* and tumor *i* we introduce a Boolean-

valued function $A(g, i)$. $A(g, i)$ is true, and we say *g* is on its *altered* level in tumor *i*, if and only if (shortly, iff) *g* is altered or imputed to be altered in *i*. Recall that, unlike for genes altered by PMs, we require the expression of primary genes altered by CNV to be consistently elevated or decreased upon their own alteration. As it is easier to manipulate gene expression rather than mutation, for each CNV-altered primary gene *g* and tumor *i*, we determined such a Boolean-valued function $A'(g, i)$ that depends on expression of *g* in *i*, and correlates with the genomic alteration-based attribute $A(g, i)$ (Supporting Information). We say *g* is on its *as altered* level in all tumors *i* for which $A'(g, i)$ holds. Supporting Information Figure S1 shows good general agreement of the expression-based attribute with genomic alterations. 20 cases of primary genes for which this agreement was not satisfactory were left out of the analysis.

We next define that a Boolean-valued function $U(g, i)$ holds iff expression of *g* is *up* in tumor *i*, that is, has a value greater than the 80%-th quantile of overall expression distribution of *g* across all tumors. Note that definition of $U(g, i)$ is disease-specific; higher end of the expression range that is observed in tumors, might, for example, happen to be the base level in the normal tissue. Similarly, $D(g, i)$ holds iff *g* is *down* in *i*, that is, expression of *g* in *i* is lower than the 20%-th quantile of *g*'s expression across all tumors.

Table 1. All variants of scenarios and relevant SSL levels considered in this work

Pattern	Scenario	SSL primary		SSL partner	
		PM-altered	CNV-altered	PM-altered	CNV-altered
Concurrence	I	Altered	As altered	Opposite	Opposite
	II	Down	Opposite	Opposite	Opposite
Mutual exclusivity	III	Altered	As altered	Altered	As altered

We assume that a candidate SSL pair of genes follows one of three scenarios (column Scenario). According to the scenarios, a particular genomic alteration and expression pattern (Pattern) should be observed for the pair. SSL levels denote such status of the primary gene (SSL primary) and its partner (SSL partner), which violate their pattern and potentially decrease tumor performance. SSL levels are defined in different variants depending on whether the genes are altered by somatic point mutations (subcolumn PM-altered) or by CNV (CNV-altered).

Let g be a primary gene and g' be a gene whose expression was measured. Let $m(g', g)$ denote median g' expression across all tumors i for which $A(g, i)$ holds. We set a Boolean-valued function $O(g', g, i)$ to true iff expression of gene g' in i is on the *opposite* extreme (down or up) of its entire expression range than $m(g', g)$. Trivially, $O(g', g, i)$ implies either $D(g', i)$ or $U(g', i)$ (example in Fig. 1b).

SSL scenarios and determination of SSL levels. We focus on three scenarios in which SSL interactions may occur. Each scenario assumes that either concurrence or mutual exclusivity pattern is observed in the data for a given pair of genes g_1, g_2. Boolean-valued functions $SSL(g_1, i)$ and $SSL(g_2, i)$ are defined, based on mutation states or expression levels of g_1 and g_2 (SSL levels; summarized in Table 1), in such a way that in all tumors i for which both $SSL(g_1, i)$ and $SSL(g_2, i)$ hold, the assumed pattern is violated.

In scenario I, exemplified in Figure 1b, expression of the partner gene g_2 is concurrent with alteration of the primary gene g_1. This scenario assumes that it would suffice to manipulate the level of the partner gene to violate concurrence and impair performance of tumors. Thus, we set the SSL level of the primary gene g_1 to its *altered* or *as altered* level, and the SSL level of the partner gene to its *opposite* level. More formally, in case g_1 is altered by PMs, $SSL(g_1, i)$ is true if and only if $A(g, i)$ holds (shortly, we write $SSL(g_1, i) \equiv A(g_1, i)$. In case g_1 is altered by CNV, we instead define $SSL(g_1, i) \equiv A'(g_1, i)$, since $A'(g_1, i)$ is expression-based and easier to induce experimentally than $A(g_1, i)$. This corresponds to identifying the SSL level of g_1 with its alterations as observed in tumors. On the other hand, for the partner gene g_2 we set $SSL(g_2, i) \equiv O(g_2, g_1, i)$. This corresponds to identifying the SSL level of g_2 with its expression level that is opposite to the level acquired in tumors upon the alteration of g_1.

In scenario II, the primary gene g_1 and its partner g_2 are also concurrent, but here it is assumed that both g_1 and g_2 need to be manipulated to impair tumor performance. In the case the primary gene g_1 is altered by CNV, we know it is concordant with its own expression. In this case we set the level of g_1 to *opposite*, and define $SSL(g_1, i) \equiv O(g_1, g_1, i)$ for each tumor i. For g_1 altered by PMs, which are not correlated with expression changes, we set its SSL level to *down*, and

define $SSL(g_1, i) \equiv D(g_1, i)$. In either case, we set the SSL level of the partner gene g_2 to *opposite*, and define $SSL(g_2, i) \equiv O(g_2, g_1, i)$.

Finally, scenario III assumes mutual exclusivity of two alterations. Here, both genes in the pair are primary. For each gene g in the pair, its SSL level is set to its *altered*, or *as altered* level, and we define $SSL(g, i) \equiv A(g, i)$ or $SSL(g, i) \equiv A'(g, i)$, depending on whether g is altered by PMs or CNV, respectively. In the following, we assume that candidate SSL interactions are pairs of genes that follow the pattern assumed by, and have their SSL levels defined according to these three scenarios.

Survival analysis for evaluation and ranking of SSL interactions

We apply survival analysis[32] to develop scores for evaluating and ranking of the candidate SSL interactions. For each such SSL interaction, we divide the analyzed tumors into four disjoint groups, denoted *Both*, *G1*, *G2* and *Neither*, depending on the corresponding SSL levels (example in Fig. 1b). *G1* is defined as the set of tumors where the primary gene is on its SSL level but the partner gene is not. Formally, for the pair of genes g_1, g_2, $G1 = \{i | SSL(g_1, i) \wedge \neg SSL(g_2, i)\}$. *G2* is defined for the partner analogously. *Both* is defined as the group of tumors with both genes on their SSL levels at the same time, that is, $Both = \{i | SSL(g_1, i) \wedge SSL(g_2, i)\}$. *Neither* is the group of remaining tumors. Intuitively, scoring of SSL candidates is performed by treating the group *Neither* as the reference. The score assesses whether the difference in survival of carrier patients between *Both* and the *Neither* is larger than expected from differences between *G1* and *Neither* as well as between *G2* and *Neither*.

Quantifying SSL. Let t denote the time duration from the moment of cancer diagnosis, $0 \leq t \leq t_{max}$, where t_{max} is the maximum monitoring time across all patients in the analyzed cohort. Given a candidate SSL interaction, we are interested in survival of patients carrying tumors in the four groups *Both*, *G1*, *G2* and *Neither*, defined for this interaction. Namely, we estimate the survival function $S(t)$, which is the cumulative probability of survival up to time t (the probability that the patient will die after a time point t). The survival data for cancer patients is often right censored (*e.g.*, 23% of

the glioblastoma dataset): some patients are still living or stopped being monitored before death, and for them only the time that passed from the diagnosis to the last follow-up is known. We thus apply Kaplan–Meier estimation[33] of the survival function, denoted $\hat{S}(t)$, separately for the four groups *Both*, *G1*, *G2* and *Neither*. Next, for each patient group, we compute its restricted mean lifetime,[34] denoted μ. μ is an equivalent of expected lifespan that is limited to the monitored time interval, and is defined as an integral of the survival function up to the maximum monitoring time *T*, that is common for all groups compared in the analysis, $T \leq t_{max}$. Here, it can be approximated by summing the Kaplan-Meier estimates $\hat{S}(t)$ of the survival function:

$$\mu = \int_0^T S(t) \approx \sum_0^T \hat{S}(t).$$

We approximate the *performance* of patients from group *G* in the set *Both*, *G1*, *G2* with a ratio $\mu^G/\mu^{Neither}$, where μ^G is the restricted mean computed for group *G*. Consequently, the performance f^G of tumors in group *G* is assessed as an inverse of the carrier patients' performance, so $f^G = \mu^{Neither}/\mu^G$. Finally, we propose a SSL measure, which we call *S*-score. For a given SSL interaction, *S* compares the performance of tumors having both genes on their SSL levels defined for this interaction, to the product of performances of tumors having exclusively one of the genes on its SSL level. Formally, *S* is defined as:

$$S = \log\left(\frac{f^{Both}}{f^{G1}f^{G2}}\right) = \log\left(\frac{\mu^{G1}\mu^{G2}}{\mu^{Both}\mu^{Neither}}\right)$$

S is negative when the performance of the tumors in group *Both* is lower than the expected product of the individual performances of tumors in *G1* and *G2*. Thus, negative *S* indicates a SSL interaction in the tumors, which is beneficial for the patients. Conversely, the score is positive when the tumors in *Both* are better off than expected from this product, and indicates a synthetic healthy or viable interaction in tumors, which is detrimental for patients.

To define the set of *plausible SSL interactions* for the glioblastoma cohort we applied a *S*-score threshold −0.4, and discarded all interactions with the *S*-score higher than this threshold. The threshold corresponds to a 1.5-fold decrease of tumor performance in the *Both* group as compared to the expected performance. Additionally, we conservatively filtered out candidate interactions with the *Both* group smaller than 10 tumors and a small subset of interactions that could potentially result in survival estimations of the *Both* group that are over-pessimistic for the tumors (Supporting Information).

Results

Here, we detail the results of each analysis step for the glioblastoma dataset (Fig. 1*a*). Out of 85 primary genes we

initially selected (Methods), 82 were altered by CNV, concordant with their own expression, and three genes were altered by point mutations. The prevalence of CNV-altered primary genes shows their critical importance in glioblastoma, and thus the great potential of therapies aimed at those alterations. As selection was separate for point mutations and for CNV alterations, a gene could be selected twice (*e.g.*, *EGFR*). This, however, is desired, since the same gene could be in SSL interactions with different partners depending on the type of its alterations. Next, the dataset was scanned for genes that are in candidate SSL interactions with the primary genes (Methods). Identification of alteration and expression patterns resulted in 17,390 concurrent and 96 mutually exclusive gene pairs. Determination of SSL levels according to three SSL scenarios delivered 14,045 candidate SSL interactions fitting scenario I, 17,390 interactions fitting scenario II, and 49 fitting scenario III (altogether 31,484 interactions).

The candidate SSL interactions were ranked by their *S*-scores, prioritizing interactions with the most prominent SSL effect according to carrier patient survival (see Supporting Information Fig. S2 for histogram of the *S*-scores). For each candidate interaction, its *S*-score compares the performance of the tumors with both genes on their SSL level (*i.e.*, in the *Both* group) to what is expected from having only one gene on its SSL level. The performance of tumors was estimated using the reverse restricted mean lifetime of their carrier patients, and treating patients with neither of the genes on their SSL levels as a reference (Methods). Further thresholding and filtering resulted in selection of 1,956 plausible SSL interactions (Supporting Information Table S1; figures available online at www.molgen.mpg.de/~szczurek/SLcancer/). A number of 867 plausible interactions were identified in scenario I, 1,084 in scenario II, and five interactions in scenario III. There are 74 unique genes altered by CNV either on the first or the second components of the interacting gene pairs, and only two altered by point mutations (no plausible pairs with *PTEN* as component were identified). Our analysis provides together 1917 plausible SSL interactions containing CNV-altered genes (either on the first or second component). These interactions hold the potential for SSL-based combined therapy that would directly target the widespread CNV alterations in glioblastoma.

Hints for known synthetic lethal interactions in glioblastoma

To inspect our results for the known SSL interactions, we collected eleven experimentally verified SSL partners of three primary glioblastoma genes: *TP53*, *EGFR* and *PTEN*, as reviewed by Weidle *et al.*[9] We are not aware of any experimentally verified synthetic lethal interaction with genes that are altered by CNV.

First, we found *PLK1* to have expression significantly and exclusively concurrent with point mutations of *TP53*. *PLK1* was shown to be consistently up-regulated under mutations of primary gene *TP53*, and its inhibition significantly

Table 2. Insights into known pairs

Primary	Partner	Cor. comb. p	S	SSL primary	SSL partner	Scenario
TP53	*PLK1*	5.7e−03	−0.6	*Altered*	*Opposite (down)*	I
EGFR	*AKT2*	1.8e−05	−0.58	*Opposite (down)*	*Opposite (down)*	II
	AKT1	1	–	–	–	–
	AKT3	1	–	–	–	–

Results for two known SSL pairs (white background), and two AKT kinases, known *not* to be SSL with EGFR (gray background). Column Cor. comb. *p*—Bonferroni-corrected combined *p*-value, serving to signify the concurrence of the partner's expression with the alteration. The *S*-score (column *S*), the SSL levels (columns SSL primary and partner), and the fitting scenario are reported where available. Bold marks values that meet our criteria for selection of SSL interaction candidates (using the *p*-values) and selection of plausible interactions (using the *S*-score). For the two known pairs, we identify the correct scenario and correct SSL levels. Neither *AKT1* not *AKT3* satisfy our criteria for candidate SSL interactions and thus for them the *S*-score and scenario are not available.

decreased viability of *TP53* mutant cancer cells.[35] As expected, disturbance of this concurrence (scenario I) delivered a very low, plausible *S*-score (Table 2). Another SSL interaction identified as plausible, *EGFR* and *AKT2*, is known to interact genetically in glioblastoma: combinatorial knock-down of *EGFR* and *AKT2* resulted in tumor-specific apoptosis and led to significantly increased survival in intracerebral mouse models.[36] As its discovery, the phenomenon of SSL interaction between *EGFR* and *AKT2*, but not *AKT1* nor *AKT3* (Akt family kinases with a similar function), remained unexplained. Our results give a highly suggestive clue for this phenomenon: as shown in Table 2, from this family of kinases only expression of *AKT2* is concurrent with *EGFR* mutation, fitting scenario II.

Supporting Information Table S2 lists the remaining nine known SSL interactions, which do not match any genomic alteration and expression pattern. Thus, those interactions cannot be explained by our scenarios, and were not preselected as input candidate SSL interactions for our analysis.

Plausible SSL interactions with overall influence on survival in glioblastoma

By definition, our *S*-score is helpful in identifying plausible SSL interactions, for which tumors having both genes on their SSL levels (tumors in the *Both* group; Methods) perform worse than expected from having only one gene on its SSL level (*G1* or *G2* group). We now ask a different question, and aim at such a subset of the plausible SSL interactions for which survival of patients with tumors in the *Both* group improves as compared to all other patients (in *G1*, *G2* or *Neither*). These interactions have an overall influence on patient survival and as such are of the most clinical interest.

To this end, we applied Cox modeling,[37] and tested which of the interactions are good survival predictors. Intuitively, Cox models estimate how strongly predictors in a given set relate to patient survival. The influence of each predictor is evaluated using hazard ratio[38] and its significance is estimated using the Wald′s test.[39] This approach allowed us to compare the interactions to a predictor based on patient′s age, with younger patients expected to have a significantly better outcome in glioblastoma.[40]

Testing was performed on a preselected subset of interactions. First, we chose such 1,901 of all plausible SSL interactions (97%), for which tumor performance in the *Both* group is lower than one, indicating that survival of patients in the *Both* group is better than of those in *Neither*. Interestingly, for 916 (48%) of these interactions, performance is lower than one *exclusively* for the *Both* group. In those cases the synthetic effect is very profound: the two individual SSL levels are advantageous for tumors, but combining them together decreases tumor performance and reinforces patients. Second, out of the 1,901 interactions we further selected a small set of 440, which satisfy the proportional hazard assumption[31,32,41] (required for Cox models; Supporting Information), and which showed a significant survival difference between patients in the *Both* group and all other patients (log-rank test *p*-value < 0.05).

For each of the 440 SSL interactions, we made a predictor variable with value 1 for patients in the *Both* group, and 0 otherwise. Next, we fitted bi-variate Cox models, using this together with a predictor based on discretized patient′s age (with a threshold of 40 years) as a reference. 274 of those SSL-based predictors (62%) have a profound influence on survival in glioblastoma (Wald′s test *p*-value < 0.05, Supporting Information Table S3). All of them are contained in a larger set of 418 predictors (95%), for which the fitted fold decrease of hazard was larger than the decrease of hazard associated with younger age. The hazard decrease caused by age was on average equal 1.4, and consistent across all models (standard deviation 0.03). For the SSL-based predictors, the mean hazard decrease was stronger, and equal to 1.8 (standard deviation 0.35). Taken together, Cox modeling identifies SSL interactions with very strong evidence for influence on overall glioblastoma survival.

Finally, we selected 40 (out of 274) such interactions that are the most feasible to verify experimentally, and have SSL levels that clearly correspond to knock-down or over-expression of genes as compared to the healthy control (removing interactions with SSL levels that effectively correspond to the level in healthy tissue; for example, see Supporting Information Fig. S3). Table 3 lists 20 such interactions underlying predictors with most impact on hazard in glioblastoma,

Table 3. Top plausible SSL pairs with significant impact on overall survival in GBM

Primary	Partner	HR	Wald's *p*	*S*	EM primary	EM partner	Alt.	Sc.
EGFR	IFIH1	1.97	0.001	−0.83	–	Knock down	CNV	I
EGFR	TRIM21	1.86	0.007	−0.69	–	Knock down	CNV	I
PDGFRA	OIP5	2.21	0.015	−0.73	–	Knock down	CNV	I
EGFR	NOLC1	1.96	0.016	−0.83	–	Over-expression	CNV	I
EGFR	INA	1.87	0.016	−0.86	–	Over-expression	CNV	I
TP53	SLC1A5	2.98	0.016	−0.65	–	Knock down	PM	I
EGFR	SUPV3L1	2.09	0.019	−0.87	–	Over-expression	CNV	I
EGFR	ART3	1.75	0.03	−0.42	–	Knock down	CNV	I
EGFR	TRIM5	1.87	0.03	−0.79	–	Knock down	CNV	I
EGFR	MMS19	1.78	0.031	−0.65	–	Over-expression	CNV	I
FKBP9L	MOSC2	2.16	0.001	−0.87	Knock down	Knock down	CNV	II
FKBP9L	GALNT13	2.4	0.001	−0.85	Knock down	Over-expression	CNV	II
FKBP9L	MEOX2	2.12	0.002	−0.78	Knock down	Knock down	CNV	II
FKBP9L	VAV3	1.87	0.006	−0.88	Knock down	Knock down	CNV	II
SRD5A3	FAM83D	2.66	0.008	−0.88	Knock down	Knock down	CNV	II
FKBP9L	SHOX2	2.12	0.01	−0.85	Knock down	Knock down	CNV	II
B4GALNT1	TMEM196	1.9	0.011	−0.84	Knock down	Knock down	CNV	II
FKBP9L	P2RY1	2.3	0.012	−1	Knock down	Knock down	CNV	II
B4GALNT1	MOV10	1.94	0.013	−0.95	Knock down	Over-expression	CNV	II
FKBP9L	GMPR	2.03	0.019	−0.77	Knock down	Knock down	CNV	II

SSL pairs with most impact on hazard in glioblastoma, verifiable in the lab by over-expression or knock down of genes, fitting scenario I and scenario II. For each pair of primary gene and its partner, significance of a predictor based on their SSL levels was assessed using two-variate Cox modeling, together with a predictor based on patient's age. The estimated fold-decrease of hazard ratio (column HR) and *p*-value in the Wald's test (Walds's *p*) are reported. In addition, for each pair we list the *S*-score (*S*), experimental manipulation required for validation (EM primary and EM partner), type of the alteration of the altered gene (Alt.) and scenario (Sc.).

fitting scenario I and scenario II (top ten each; see Supporting Information Table S4 for the full list). Figure 2 presents the Kaplan–Meier plots and SSL levels for two top selected interactions. The survival curves clearly illustrate the advantage of having both interacting genes on their SSL levels with respect to the rest of the patients.

SSL networks
Next, we constructed a network of SSL interactions over the set of 274 plausible SSL interactions that were indicated *via* Cox modeling to have a positive impact on overall patient survival in glioblastoma. The largest part of the network spans 213 interactions that fit scenario II and have the primary gene altered by CNV (Supporting Information Fig. S4A). Remarkably, the network has a visible hub structure. There are several genes for which multiple SSL partners with overall survival importance have been identified, predominantly *SEC61G* (55 partners), *FKBP9L* (21), *MTAP* (18), *CHIC2* (16), *CDKN2B* (11) and *PSPH* (11). Another subnetwork spans 58 interactions that are also altered by CNV, but fit scenario I (Supporting Information Fig. S4B). Here, the dominating hubs are *EGFR* (with 40 partners), *MTAP* (6), and *PDGFRA* (5). Only three interactions make up the small

subnetwork that fits scenario I, with all edges connecting *TP53* altered by point mutations to its partners (Supporting Information Fig. S4C).

Cox modeling validates the S-scores
Finally, we utilized Cox modeling for validation of the *S*-scores, verifying that interactions with extreme *S*-scores are enriched in predictors with high impact on overall survival. In contrast to the *S*-score, neither the hazard ratio nor the Wald′s *p*-value estimated for the Cox models are based on restricted mean, and thus they provide independent significance measures. Here, we repeated selection of plausible SSL interactions and subsequent Cox modeling for all 7,753 candidate SSL interactions with negative *S*-scores, which meet our filtering criteria for validity, as, for example, minimal size of the smallest patient group, but do not necessarily meet the stringent *S*-score threshold −0.4 (see Methods). The set of 1,956 interactions selected in our main analysis ("Results" Section) as plausible is the subset of those 7,753 interactions that satisfies this threshold. From the set of 7,753 interactions we further selected 852 as predictors for consequent bi-variate Cox modeling, together with a predictor based on age. Again, the 440 predictors analyzed in "Plausible SSL
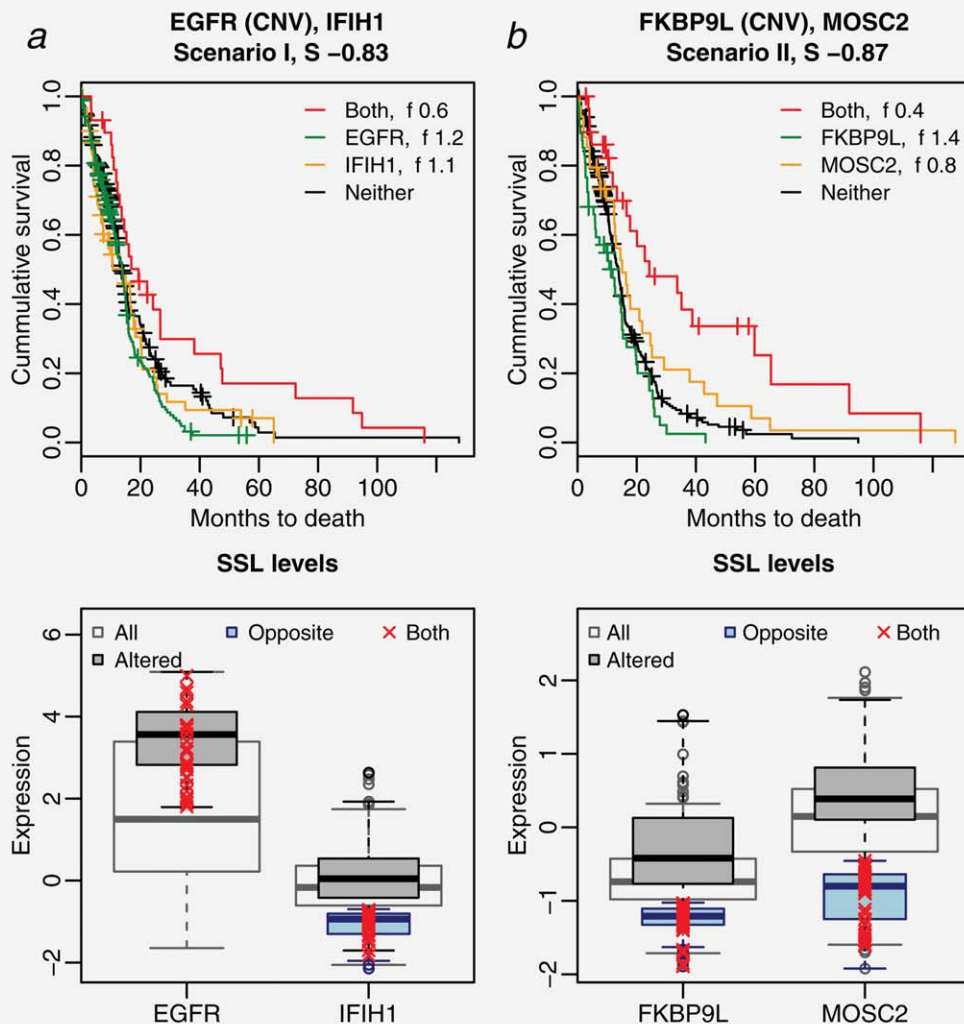
Cancer Genetics

Figure 2. Two most plausible SSL interactions with significant impact on overall survival in glioblastoma. Kaplan–Meier plots in the first row show survival curves for four patients groups, *Both* (having both genes on their SSL levels in tumors; plotted in red), *G1* (having only the first gene on its SSL level; blue), *G2* (having only SSL level of the second gene; orange), and *Neither* (black). *f* denotes performance of each group. On all plots the area under the survival curve for the *Both* group is significantly larger than under the reminding three curves. Similarly, the estimated performance of tumors in the *Both* group is much smaller than expected from performances in the groups *G1* and *G2*. Boxplots in the second row show expression value distributions for the two interacting genes (g_1, g_2). White boxplots: distribution of expression values across all tumors. Gray: expression in those tumors, that have gene g_1 on its *altered/as altered* level (all tumors *i* that satisfy $A(g_1, i)$ or $A'(g_1, i)$). Blue boxplots correspond to the *opposite* levels. Blue boxplots, left: expression in tumors *i* that have g_1 on the *opposite* level and satisfy $O(g_1, g_1, i)$; right: in tumors *i* that satisfy $O(g_2, g_1, i)$. Red: expression values in those patients that have both genes on their respective SSL levels. (a) Interaction *EGFR*, *IFIH1* fits scenario I, *EGFR* is altered by CNV, and *IFIH1* has increased expression values upon the alteration of *EGFR*. Thus, the SSL level of *EGFR* is set to the *as altered* level, while the SSL level of *IFIH1* is set to *opposite* (in this case, *opposite* is equivalent to *down*). (b) Interaction *FKBP9L*, *MOSC2* fits scenario II. Expression of *FKBP9L* is elevated compared to the range of all its expression values upon its own alteration. The level of *MOSC2* expression in those patients that have *FKBP9L* altered is also increased. The SSL level of *FKBP9L* is set to *opposite* (here equivalent to *down*), as is the SSL level of *MOSC2*.

interactions with overall influence on survival in glioblastoma" Section are subset of these 852, based on the *S*-score threshold −0.4.

Figure 3 shows that the *S*-scores for SSL interactions resulting in significant predictors are significantly enriched on the low end of the entire negative *S*-score range (Wilcoxon′s test enrichment *p*-value 6.7e−29; Supporting Information), similar to the interactions that resulted in hazard

decrease stronger than the decrease associated with age (*p*-value 4e−35). Supporting Information Figure S5 shows high correlation of *S*-scores with hazard ratio (Pearson correlation −0.64) and with Wald′s *p*-values (0.44). These results also suggest that those interactions for which the Cox models cannot be constructed (for which the proportional hazard assumption does not hold), but which have low *S*-scores, are potentially of high survival importance.
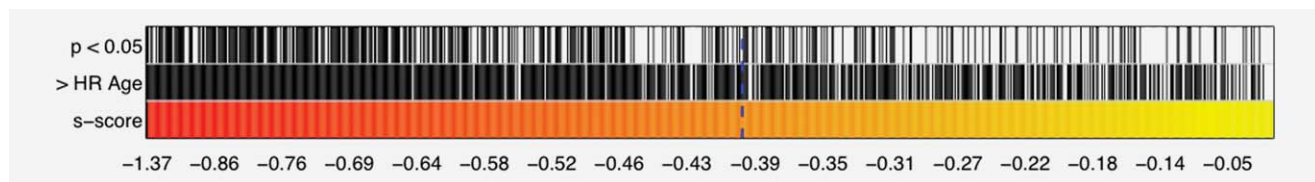
Figure 3. Independent evaluation validates the *S*-scores. Bottom row presents the *S*-score range for 852 interactions with negative scores. Middle row marks with black stripes which of them result in survival predictors that have a more profound influence on hazard decrease than age, and top row marks which of them have significant Wald's *p*-values. Strikingly, the significant results are strongly enriched on the low end of the *S*-score range. Blue dashed line indicates the *S*-score threshold −0.4 applied in our main analysis.

Discussion

This work presents statistical analysis that combines point mutations, CNV and gene expression together with carrier patient survival for a cohort of glioblastomas. We identify candidate SSL interactions based on mutual exclusivity and concurrence patterns, and propose the *S*-scores for their evaluation and ranking. The methodological advantage of the *S*-scores is that they are available for a large fraction of all interactions for which the proportional hazard assumption does not hold and for which significance tests that depend on this assumption would not be valid. This decision is supported by Royston and Parmar,[42] who advocated the use of restricted mean survival time in cases when this assumption is doubtful. Moreover, the *S*-score, constructed specifically to measure SSL interaction, combines survival comparisons between four tumor groups at the same time (*Both*, *G1*, *G2* and *Neither*).

Mutual exclusivity between genes has been extensively studied before[26,27,43,44] and was indicated[24] to either point to their synthetic lethality or to their involvement in the same pathway. Unlike the previous approaches, our analysis is able to distinguish these two cases, since only the former will be reflected in patient survival and our *S*-scores.

Our analysis does not exclude the fact that the combined treatment based on the selected plausible SSL interactions may also harm the normal cells, and this should be investigated experimentally. As shown in "Plausible SSL interactions with overall influence on survival in glioblastoma" Section, our approach can easily be adjusted to consider only such SSL levels that are practically verifiable in the lab. Experimen-

tal verification of a given SSL interaction should involve setting both genes to their SSL levels, for example in cancer cell lines or in mouse xenografts, and monitoring cell/tumor viability as compared to only either of the genes on its SSL level. Clearly, there is a long way from a verified SSL interaction to the actual drug discovery. To resolve how the required SSL levels could be induced for treatment of human tumors, their protein product localization, post-transcriptional modifications, and turnover should be assessed in further experimental rounds.

We foresee that enlarged sample datasets that are currently being generated for various cancers will increase the power of our approach. For example, the fact that a given interaction was not identified as SSL does not necessarily imply that it is not, since we discarded a large number of candidate interactions with small sized patient subgroups.

We note that there is room for future research, for example a deepened analysis of interactions with positive *S*-scores, which in the current work were discarded. Those interactions may be of diagnostic relevance since they point at gene states (their relevant SSL levels) that are together associated with unexpectedly bad outcome. Planned improvements include data model-driven definition of the SSL levels, and significance assessment of the *S*-scores. Still, our results with the current approach already very clearly indicate that we are able to decipher traces of SSL interactions hidden in the tumor genomic data.

References

1. Force T, Kolaja KL. Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. *Nat Rev Drug Discov* 2011;10:111–26.
2. Diaz LA Jr, Williams RT, Wu J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 2012;486:537–40.
3. Misale S, Yaeger R, Hobor S, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* 2012;486:1476–4687.
4. Ellis LM, Hicklin DJ. Resistance to targeted therapies: refining anticancer therapy in the era of molecular oncology. *Clin Cancer Res* 2009;15:7471–8.
5. Kaelin WG. The concept of synthetic lethality in the context of anticancer Therapy. *Nature Reviews Cancer* 2005;5:689–98.
6. Ashworth A, Lord CJ, Reis-Filho JS. Genetic interactions in cancer progression and treatment. *Cell* 2011;145:30–8.
7. Chan DA, Giaccia AJ. Harnessing synthetic lethal interactions in anticancer drug discovery. *Nat Rev Drug Discov* 2011;10:351–64.
8. Costanzo M, Baryshnikova A, Bellay J, et al. The Genetic Landscape of a Cell. *Science* 2010;327:425–31.
9. Weidle UH, Maisel D, Eick D. Synthetic lethality-based targets for discovery of new cancer therapeutics. *Cancer Genomics Proteomics* 2011;8:159–71.
10. Luo J, Emanuele MJ, Li D, et al. A Genome-wide RNAi Screen Identifies Multiple Synthetic Lethal Interactions with the Ras oncogene. *Cell* 2009;137:835–48.
11. Kim Y-W, Liu TJ, Koul D, et al. Identification of novel synergistic targets for rational drug combinations with PI3 kinase inhibitors using siRNA synthetic lethality screening against GBM. *Neuro Oncol* 2011;13:367–75.

Cancer Genetics

12. Guo GS, Zhang FM, Gao RJ, et al. DNA repair and synthetic lethality. *Int J Oral Sci* 2011;3:176–9.

13. Moeller BJ, Arap W, Pasqualini R. Targeting synthetic lethality in DNA damage repair pathways as an anti-cancer strategy. *Curr Drug Targets* 2010;11:1336–40.

14. Qi Y, Suhail Y, Lin Y-Y, et al. Finding friends and enemies in an enemies-only network: a graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. *Genome Res* 2008;18:1991–2004.

15. Paladugu S, Zhao S, Ray A, et al. Mining protein networks for synthetic genetic interactions. *BMC Bioinformatics* 2008;9:426.

16. Pandey G, Zhang B, Chang AN, et al. An integrative multi-network and multi-classifier approach to predict genetic interactions. *PLoS Comput Biol* 2010;6:e1000928.

17. Wong SL, Zhang LV, Tong AHY, et al. Combining biological networks to predict genetic interactions. *Proc Natl Acad Sci USA* 2004;101:15682–7.

18. Dixon SJ, Costanzo M, Baryshnikova A, et al. Systematic mapping of genetic interaction networks. *Annu Rev Genet* 2009;43:601–25.

19. Jiang X, Neapolitan R, Barmada MM, et al. Learning genetic epistasis using Bayesian network scoring criteria. *BMC Bioinformatics* 2011;12:89.

20. Li J, Horstman B, Chen Y. Detecting epistatic effects in association studies at a genomic level based on an ensemble approach. *Bioinformatics* 2011;27:i222–i9.

21. Kam-Thong T, Pütz B, Karbalai N, et al. Epistasis detection on quantitative phenotypes by exhaustive enumeration using GPUs. *Bioinformatics* 2011;27:i214–i21.

22. Conde-Pueyo N, Munteanu A, Sole R, et al. Human synthetic lethal inference as potential anti-cancer target gene detection. *BMC Syst Biol* 2009;3:116.

23. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. *Nat Rev Genet* 2007;8:437–49.

24. Ciriello G, Cerami EG, Sander C, et al. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res* 2012;22:398–406.

25. Ding L, Getz G, Wheeler DA, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;455:1069–75.

26. Masica DL, Karchin R. Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. *Cancer Res* 2011;71:4550–61.

27. Miller C, Settle S, Sulman E, et al. Discovering functional modules by identifying recurrent and mutually exclusive mutational patterns in tumors. *BMC Med Genom* 2011;4:34.

28. Mischel P, Cloughesy T. Targeted molecular therapy of GBM. *Brain Pathol* 2003;13:52–61.

29. TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061–8.

30. Jacobsen A. cgdsr: *R-Based API for accessing the MSKCC Cancer Genomics Data Server (CGDS)*. R package version 1.1.13. Available at: http://CRAN.R-project.org/package=cgdsr, 2011.

31. Beroukhim R, Getz G, Nghiemphu L, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci USA* 2007;104:20007–12.

32. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data, second ed. New York: Springer, 2003.

33. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457–81.

34. Irwin JO. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *J Hyg* 1949;47:188–9.

35. Sur S, Pagliarini R, Bunz F, et al. A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53. *Proc Natl Acad Sci USA* 2009;106:3964–9.

36. Michiue H, Eguchi A, Scadeng M, et al. Induction of in vivo synthetic lethal RNAi responses to treat glioblastoma. *Cancer Biol Ther* 2009;8:2306–13.

37. Cox DR. Regression models and life-tables. *J R Stat Soc Series B (Methodological)* 1972;32:187–220.

38. Spotswood L, Spruance JER, Michael G, et al. Hazard ratio in clinical trials. *Antimicrob Agents Chemother* 2004;48:2787–92.

39. Agresti A. Categorical data analysis, Second edn., New York: Wiley, 2002.

40. Batchelor TT, Betensky RA, Esposito JM, et al. Age-dependent prognostic effects of genetic alterations in glioblastoma. *Clin Cancer Res* 2004;10:228–33.

41. Breslow NE. Analysis of survival data under the proportional hazards model. *Int Stat Rev* 1975;43:45–57.

42. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med* 2011;30:2409–21.

43. Vandin F, Upfal E, Raphael BJ. De novo discovery of mutated driver pathways in cancer. *Genome Res* 2012;22:375–85.

44. Vandin F, Upfal E, Raphael BJ. Algorithms for detecting significantly mutated pathways in cancer. *J Comput Biol* 2011;18:507–22.