



HAL
open science

Between compensatory mutations and synthetic lethals: genetic mutations, a new challenge for tomorrow's medicine

Claude Bazin, Rémi Coupaye, Sandrine Middendorp, Anne Vanet

► **To cite this version:**

Claude Bazin, Rémi Coupaye, Sandrine Middendorp, Anne Vanet. Between compensatory mutations and synthetic lethals: genetic mutations, a new challenge for tomorrow's medicine. *Science Postprint*, 2014, 1 (1), pp.e00035. 10.14340/spp.2014.10r0002 . hal-01898724

HAL Id: hal-01898724

<https://hal.science/hal-01898724v1>

Submitted on 18 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Between compensatory mutations and synthetic lethals: genetic mutations, a new challenge for tomorrow's medicine

Claude Bazin^{1,2}, Rémi Coupaye^{1,2}, Sandrine Middendorp^{1,3}, Anne Vanet^{1,2,4}

Abstract

In the macromolecular world, the evolution of two building blocks (two nucleotides or two amino acids) can be interdependent in various ways, including: (i) a mutation at one site compensates a deleterious mutation at another site or (ii) mutations at two different sites are lethal only when they co-occur in the same genome. These two situations are known as “compensatory mutations” and “synthetic lethals,” respectively. Although the first category has been studied extensively, especially since the 1970s—a period of time during which prokaryotic genetics grew by leaps and bounds—the second remained unstudied until the late 1990s. Studies on yeast first placed synthetic lethals at the forefront; at the beginning of the new century, therapies against cancers relied on such relationships. Finally, in recent years, synthetic lethals were used to develop stable therapies against RNA viruses, and these studies revealed a promising method for developing vaccines against these viruses. Here, we review the current understanding of these two situations and the implications of both compensatory mutations and synthetic lethals for the elucidation of biological pathways, cancer research, evolution, and gene expression.

Keywords: covariation, compensatory mutations, synthetic lethals, drug targets, RNA viruses

Author info:

1 Paris Diderot University, Sorbonne Paris Cité University, Paris, France

2 Jacques Monod Institute, CNRS, UMR7592, Paris, France

3 Department of Functional and Adaptive Biology, CNRS, UMR8251, Paris, France

4 Atelier de Bio Informatique, 75005, Paris, France

Received: Jul 10 2014 **Accepted:** Sep 26 2014 **Published:** Oct 22 2014

Citation: Bazin C, Coupaye R, Middendorp S, Vanet A (2014) Between compensatory mutations and synthetic lethals: genetic mutations, a new challenge for tomorrow's medicine. *Science Postprint* 1(1): e00035. doi: 10.14340/spp.2014.10R0002

Copyright: ©2014 The Authors. Science Postprint published by General Healthcare Inc. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 2.1 Japan (CC BY-NC-ND 2.1 JP) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Funding: Our research is not supported by any funding.

Competing interest: No relevant competing interest were disclosed.

Donation message: Our team is working in silico to find a new strategy against RNA-viruses as HIV, flu, ebola virus or Hepatitis C virus. These viruses are very dangerous because they accumulate mutations that can allow the drug to

escape. To bypass this problem, we have described a bioinformatic method which compute new therapeutic targets using viral protein alignments. We now would like to confirm this concept by biological experiments and use chemistry studies to define the future drug which will bind to the viral protein.

Patents: Vanet A, Muller-Trutwin M, Valère T, Brouillet S, Ollivier E, Marsan L, inventor. Method for identifying combinations of motifs that do not mutate simultaneously In a set of viral polypeptide sequences comprising a putative drug binding site. Patent US 7917303 B2. Mar 29 2011. Available from: http://www.lens.org/lens/patent/US_7917303_B2.

Vanet A, Muller-Trutwin M, Valère T, inventor. Method for identifying motifs and/or combinations of motifs having a boolean state of predetermined mutation in a set of sequences and its applications. Patent US 7734421 B2, Jun 8 2010. Available from: http://www.lens.org/lens/patent/US_7734421_B2.

Vanet A, Muller-Trutwin M, Valère T. Method for identifying motifs and/or combinations of motifs having a boolean state of predetermined mutation in a set of sequences and its application. Patent US 8032309 B2, Oct 4 2011. Available from: http://www.lens.org/lens/patent/US_8032309_B2.

Corresponding author: Anne Vanet

Address: Paris Diderot University, Sorbonne Paris Cité University, F-75013, Paris, France

E-mail: anne.vanet@univ-paris-diderot.fr

Introduction

During the twentieth century, geneticists used logical inference and macroscopic methods to facilitate the discovery of microscopic interactions. The reductionist paradigm was, and still is, that a function must be performed by a structure. When this function is so complex that it requires several molecules to be driven, these various molecules should cooperate, either directly, by interacting to form a protein complex, or indirectly, by sharing the same metabolic or signaling pathway. In the case of protein complexes, it is easy to conceive that the residues mediating the interactions are not independent of each other: they form either pairs or groups of covariant residues, also known as epistatic interactions. We will focus on two major covariant relationships: compensatory mutations (CM) and synthetic lethals (SL). Classically, the mutation that compensates a deleterious mutation has been uncovered by the appearance of revertants. Jarvik and Botstein [1] explain that “if a revertant is not a true wild type, they may have acquired suppressors—new mutations that act so as to correct, replace, or bypass the original defect.” Figure 1 shows a physical analogy involving an electrical circuit to explain the effects of covariant relationships: a mutation causes an increase in electrical resistance, which decreases the current flow. A

second mutation decreases the resistance of the other resistor and the current intensity is restored (when the light is on, the combination has a wild-type phenotype, and when it is off, the combination is deleterious). In the case of metabolic or signaling pathways, according to Jarvik and Botstein, reversion can imply a restoration of the original sequence (Figure 1B). Otherwise, a first deleterious mutation appears at random in the genome followed by an additional mutation elsewhere in the genome (second-site reversion), which will remedy the defect caused by the first mutation (Figure 1C). The latter mutation is known as a CM. Both mutations can be located in the same gene (intragenic CM) or in different genes (intergenic CM). In the latter case, they often involve actors in the same metabolic pathway or protein complex. Interacting proteins within complexes are studied by detecting covariant pairs, as follows: a first mutation abolishes a function, which is restored by a second mutation. Thus, the return to a wild-type phenotype is made possible either by the restoration of the mutated factor's functionality itself, or through a change elsewhere that compensates for the abolition of the first activity. Structural CMs are also powerful tools for defining spatial interactions. All of these scenarios will be treated in *extenso* below.

For their part, the SL pairs were finally recognized mainly through studies of *Saccharomyces cerevisiae*. Giaever's team [2] has shown that no more than a thousand proteins (out of more than six thousand) perform the so-called essential functions: only 18.2% of yeast genes are absolutely essential. However, the number of supposed essential functions is larger than this, which pleads in favor of the existence of functions performed by protein complexes or by alternative pathways. These complexes and/or pathways can be borne by covariant phenomena involving SL covariant pairs. Synthetic lethality describes a genetic interaction in which the combination of two separately non-lethal mutations results in lethality (Figure 1D). The central question involves identifying residues of several proteins that either cooperate to perform these essential functions (in the case of CMs) or that cannot co-occur due to synthetic lethality.

Originally CM and SL were studied using genetic, biochemical, and evolutionary biology methods. In the genomic era, bioinformatics can be used to detect them in interacting networks. These historical fields share a rich vocabulary behind which are hidden the intricacies related to biological covariation: epistasis, coevolved residues, covariants, CMs, extragenic suppressors, revertants, correlated mutations, etc. We focused on these two concepts because they initially represented a tool for studying biological functions and they have now become a powerful therapeutic tool. We also advance the vision that the landscape of epistasis is richer than previously supposed because of the existence of networks involving CM and SL.

Extragenic CMs

CMs: the stars of the geneticists

Originally, prokaryotic organisms were the reference model for studying metabolic pathways using CM. One of the first famous revertants was identified by Beckwith. It restored the sensitivity to the Lactose operon repressor [3] and then the wild-type phenotype. Following the tremendous work by Jacob and Monod, the notion of CM expanded beyond the questions of phenotype and genetic loci by introducing the concept of individual nucleotides in the DNA sequence. This moment marks the transition from classical genetics to molecular biology, which subsequently opened the way to genomics. CM could henceforth be identified with all possible molecular precision.

After decades of work on prokaryotic models, the CMs had been analyzed in haploid eukaryotes, yeast in particular, in order to dissect the analyzable pathways. In a covariant pair, the second mutation is called the suppressor, in reference to the revertant phenotype it causes. The second mutation can be located either in the originally mutated gene (intragenic suppressor) or in another gene (extragenic suppressor).

Physical interactions

How should functional reversion by an extragenic suppressor be interpreted? Figure 2 illustrates the simplest case, in which two interacting proteins are necessary for a function [4]. In the case of proteins that physically interact, as in a "key/lock" mechanism, if the lock is changed, the key must also be changed to retain the function. This implies that CMs have physical and/or structural interactions, and the second mutation then belongs to the class of conformational suppressors. The suppressor reflects compensating interactions between gene products that interact physically via a "lock and key" mechanism (Figure 2). The understanding of this mechanism opened various domains of research, especially in the comprehension of functional biological pathways [4], as CMs between proteins in the same pathway in the same organism [5-7], pathway links which were not *a priori* related [8], or links between two different organisms such as a host and a pathogen [9, 10], facilitated a tremendous step forward in comprehension of the pathological mechanism.

CMs also exist in the RNA domain [11, 12] mostly between snoRNA and rRNA, and are important in understanding ribosome structure (Figure 3A). Suppressors of ribosomal frame shifts allow a P-site realignment during translation [13]. Some CMs have even been described between a protein and an RNA [14-16].

The concept of CM is very broad and includes all types of protein relationships within the cell. An array approach pioneered in *S. cerevisiae* has facilitated the development of high-resolution CM genetic mapping techniques; it has shown that CM relationships can be numerous and complex [17].

At another level, important biological fields such as antibiotic resistance have greatly benefited from our growing understanding of such compensatory mechanisms [16, 18, 19] involved in the infectivity of a pathogen [19], which have provided significant openings for development of new therapies.

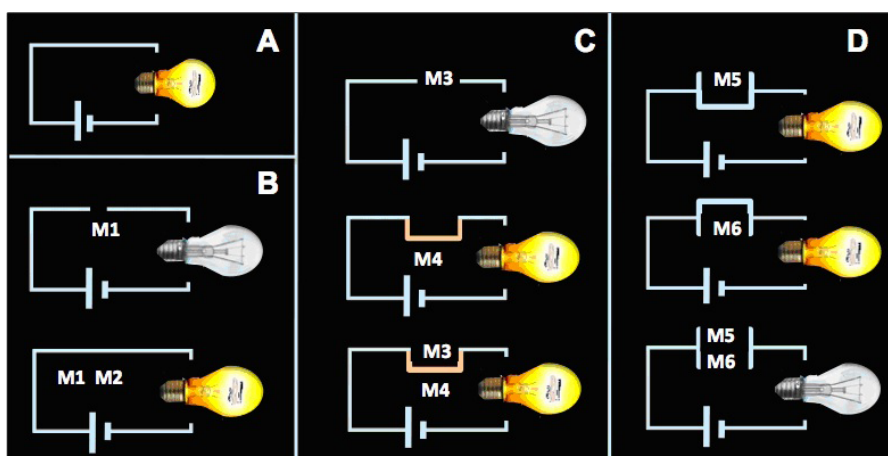


Figure 1: Genetics visualized as an electrical circuit

We illustrate all possible scenarios that compare double mutations to the wild type: when the light is on, the combination produces a wild-type phenotype, when it is off, the combination is deleterious. M1 to M6 are point mutations.

A: wild-type circuit configuration.

B: revertant at the same position circuit configuration. M1 is a deleterious mutation at position X, M2 is a mutation at the same position, which restores the wild type sequence.

C: compensatory mutation pairs circuit configuration. The first mutation, M3, is deleterious alone but is compensated by the second mutation, M4, which has no effect alone and is not located at the same position.

D: synthetic lethal pair circuit configuration. M5 and M6 have no effect alone, but together are lethal.

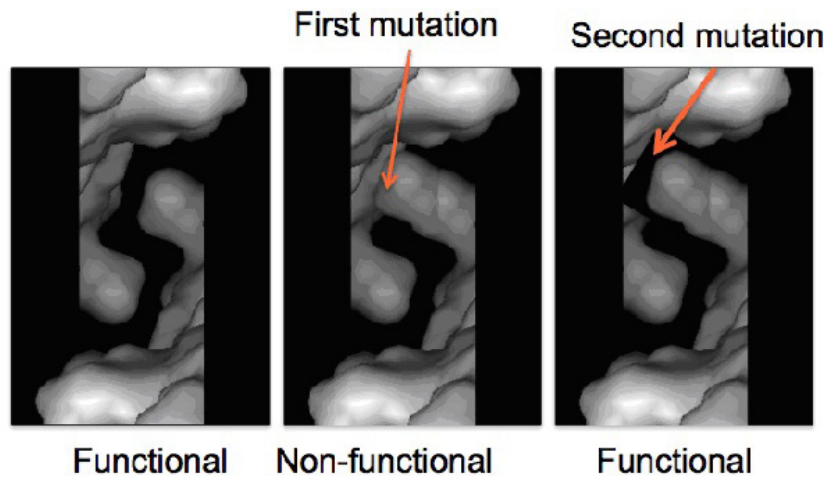


Figure 2: A key/lock model, where the second mutation restores a functional active site

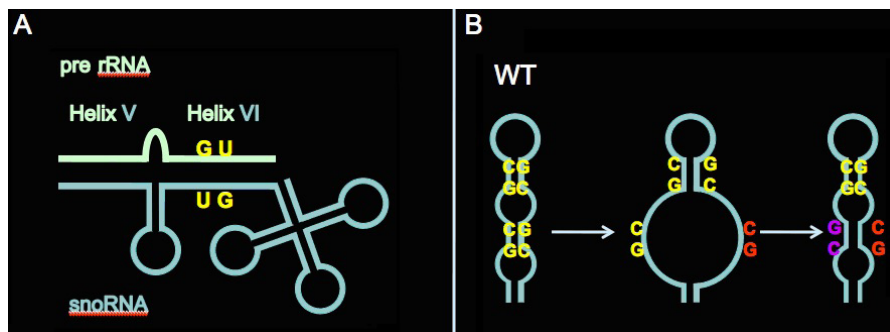


Figure 3: Covariant positions in the RNA world; A: Interaction of pre-rRNA and snoRNA in *Saccharomyces cerevisiae*; B: RNA secondary structure that can be detected and described by detecting CM pairs

A: As explained in [11], compensatory mutations in one-half of helix VI fully restored cell growth.

B: Here two point mutations alter the secondary structure, which is restored by two compensatory mutations. WT: Wild Type RNA

Functional interactions

Not all CMs are necessarily located in close physical proximity to the compensated mutation [20–22]. When they are far from each other, the covariant pair cannot be explained by a physical interaction between residues, but rather by a functional relationship participating in the establishment and maintenance of the relationship between secondary structural elements that are distant in the primary sequence and in the tridimensional structure of the protein [23]. For example, intergenic compensation can help restore proper peptide synthesis following an initial mutation that has created a stop codon in the open reading frame. One such example was described for the first time in the 1950s, a tRNA-aminoacyl tRNA synthetase charged with an amino acid that did not correspond to the anticodon it carried [24]. Subsequent to this first example, many other teams have identified other mutations in the translational apparatus [25, 26].

The formation of a biological bypass is one way to overcome a deleterious mutation. Thus, the deletion of the gene for T7 ligase is compensated by other mutations in different metabolic pathways for DNA formation to recover the original fitness [27]. Astonishingly, a novel approach has been proposed that can identify a CM pair using SL genetic interaction analysis [28]. This method identifies pairs of genes rarely bound by the SL relationship, but each taken separately is connected by many SL interactions with other genes. This approach selects pairs of CMs that repair the deleterious effect of one of the two genes, if it was mutated, without the need for physical interaction.

Intragenic CM

The definition of an intragenic CM is the following: it is a mutation that has a beneficial effect on the fitness of an organism that contains a deleterious mutation in the same gene. This secondary mutation (which compensates the first mutation) does not seem to appear randomly in the gene sequence, but rather close to the deleterious mutation, making this system a predictable mechanism [29]. Considering all revertants, the overall frequency of CM was approximately 70%, against 30% reverse mutation. Half of all CMs are located intragenically [30]. Although this percentage is rather high, it is undervalued because it does not include deletion-related mutations. Taking into account this bias, the rate of intragenic reversion is estimated at 78%, compared to 22% for extragenic reversion. Moreover, Poon et al. [31], based on a compilation of data from several genes in different organisms, estimate that among deleterious mutations, 20% cannot be compensated, 43% can be compensated by some CMs, and 37% are associated with at least 10 CMs [31], which illustrates the extreme complexity of the interaction networks and demonstrates that few proteins are essential by themselves. Chen et al. [32] focused on tandem-based substitution mutations and showed that CMs appear in 0.4% of single nucleotide substitutions in human germline cells. Bazykin et al. [33], who are interested in multiple amino acid variation, have shown that this phenomenon occurs rapidly via a succession of amino acid replacements at many sites, instead of via relaxed negative selection. Indeed, many different mutational trajectories can be described, but only few are possible [34].

Determination of intracompensatory pairs of mutations has also been used to determine the location of non-coding RNA, which can function as structural, regulatory, or even catalytic RNA. Indeed, if the position of a gene in a DNA sequence can be predicted by bioinformatic analyses using translation signals and open reading frames, these specific signals cannot be used to identify non-coding RNA.

However, even if these RNAs contain no translation signals, they have a particular secondary structure that must be preserved for them to perform their role [35]. Polymorphism analysis of RNA sequences and their secondary structures will highlight the maintenance of these structures via study of CM pairs [36] (Figure 3B). Several software packages have been developed to analyze deleterious-compensatory pairs of allosteric ribozymes and riboswitches [37].

Mutations and fitness

In evolutionary biology, the ability of individuals to reproduce can be summarized as their fitness. All organisms that are able to reproduce have a specific fitness. The appearance of a mutation is usually deleterious and sometimes neutral. When the mutation is deleterious, the organism's fitness decreases. More often than not, a secondary mutation is selected when its appearance compensates the first mutation to restore the original fitness. External pressure can require selection for a mutation that allows an organism to survive, even if this mutation decreases the general fitness of the organism. Indeed, for 60 years—since humans have been using antibiotics to treat infectious diseases—a strong selective pressure has been applied to infectious organisms such as bacteria; later this pressure was extended to viruses, fungi, insects, and plants because of the use of antiviral and antifungal therapies, insecticides, and herbicides. A first mutation, which allows these organisms to resist the treatment, quickly appeared, but in most cases showed decreased fitness. Pathogenic organisms possess several mechanisms for recovering their original fitness [38]: intragenic compensatory mechanisms have been described above, but we would like to emphasize a study on rifampicin resistance mutations in *Pseudomonas aeruginosa*. In this work, it was shown that two deleterious mutations, when associated in the same genomic background, often yielded improved fitness, and compensated for each other's costs [39]. The same conclusion was reached in studies of other organisms such as HIV [40] and *Candida glabrata* [41]. In these cases, CMs allowed recovery of higher fitness. Corbett-Detig [42] explained that fitness epistasis is widespread within natural populations in which many different alleles are present in the same species. The material necessary for reproduction is segregated within species, and does not necessarily require the emergence of genetically incompatible mutations, as previously thought. This segregation pressure is only released when divergent lineages hybridize.

The impact on evolution

What impact do these interactions have on the evolution of species? Phylogeneticists describe coevolution as an example of microadaptation: a CM is an adaptation to the first mutation to appear. This adaptation will be positively selected if the final fitness of the double-mutated organism is higher than the fitness of the wild-type organism [43]. If so, how does natural selection influence genetic equilibrium and how can scientists better understand species' genetic polymorphisms? Do only neutral mutations accumulate, or can we identify mutations selected to compensate original deleterious mutations? Coevolution can result from parallel or convergent changes with a very slow substitution rate. They were first described as neutral mutations [44], but were also presented as deleterious mutations, revealing a more highly constrained evolutionary landscape [45, 46]. Unfortunately, this controversy remains unresolved [47]. We are of the opinion that if a mutation appears that is deleterious in a given genetic background, the organism disappears. However, in another background the same mutation might not be deleterious and would appear neutral because its CM already exists in this genomic context [48]. Genomic screening can differentiate neutral intragenic polymorphism from CM by selecting some revertants, as discussed by Jarvik and Botstein [1]. Kondrashov's team [48] have shown that about 90% of amino acid substitutions have a neutral or beneficial impact only in the genetic background in which they occur, and could be deleterious elsewhere.

SLs: from a foggy story to anticancer therapy

Epistasis is not just about compensation, however complex it may be. In "The Genetics of Natural Populations," published in 1946, Dobzhansky [49] revealed an SL phenotype via crossing and recombination of chromosomes in *Drosophila pseudoobscura*. When detected by crossbreeding, this phenomenon is termed "synthetic" because it cannot appear in nature, but can only be detected in the laboratory. Except Lucchesi in 1967 [50], who discovered an SL system similar to the one discovered in 1922 by Bridges in *Drosophila melanogaster*, studies of SL essentially vanished, partially because of the unclear importance of the results, and partially because of the difficulty of detecting SL pairs. In addition, SL can be defined in multiple ways, though with essentially the same meaning. First, when both parts of an SL pair are detected in two different genes, they are called extragenic or intergenic SLs. Second, a group of mutations belonging to the same invariance group. In other words, alone, these positions are not invariant sites, but become so when grouped. In this situation, the essential function is not performed by a single amino acid but rather by several (Figure 1D and Figure 4).

Yeast, a launching pad for SL

In the 1990s, Kayser et al. [51] discovered SL gene pairs by studying the genes involved in yeast secretory pathways. Subsequently, several different techniques were developed, such as colony-sectoring assays [52], cell synchronization [53], SL screens [54], and studies of double mutants [55], allowing the discovery of new functions not performed by a single gene but by a pair; these techniques were developed almost exclusively in yeast. Indeed, as already explained, only 18.5% of *S. cerevisiae* genes were shown to be essential, but essential functions are far more numerous in eukaryotic cells. The experimental accessibility of budding yeast makes it a test-bed for technology development and application. There are hundreds of yeast examples where these techniques showed the importance of SL pairs in specific biological pathways, including morphogenesis [52, 56, 57], the N-end rule pathway [58], a new subclass of nucleoporins [59, 60], the spliceosome assembly pathway [61], replication [62–64], protein translocation [65–67], the cell cycle [53], the cytoskeleton [55], telomere size regulation [68], transcription [69], translation [70], mitosis [71], fission in *Schizosaccharomyces pombe* [72–74], and budding of *S. cerevisiae* [75].

SL pairs have also been described in the RNA realm, such as the U2 and U6 snRNA pair, which are involved in selection of pre-mRNA splice sites [76].

In the 2000s, a turning point was reached, when the complete sequence of the yeast genome [77] was made available, allowing an international consortium to generate a collection of 6,000 deletions [78] in the 6,000 predicted genes of the yeast genome, opening up new horizons for global analysis of SL genetic interactions, called a synthetic genetic array (SGA). In this screen, the query mutation is compared to an ordered array of ~5,000 viable gene deletion mutants. Following a series of replica-pinning steps, the meiotic progeny harboring both mutations can be scored for fitness defects [79, 80]. This new technique allows the discovery of genes that are in a SL relationship with a specific query gene. Next, the entire SL network for kinetochores [81], chromosome segregation [82], *S. cerevisiae* budding [83], and many others were compiled in the DRYGIN database [84]. The same SL screening method was developed for *S. pombe* [85]. This new technology allowed the development of the reporter synthetic genetic array (R-SGA), where the mutated query gene is replaced by a reporter gene downstream of a promoter of interest [86, 87] introduced in the same SGA gene deletion collection. This technique allowed the discovery of new promoter trans-acting regulators, where a mutated regulator can regulate the query promoter differently than the WT regulator. The results of *S. pombe* SGA were compared to the results for *S. cerevisiae*, generating the conclusion that 29% of the genetic interactions are common to both species [88]. An alternative technique is synthetic lethality analysis by microarray (SLAM), in which the mutated query gene is directly introduced into the yeast knockout mutant collection. The quantitative performance of the microarray allows ranking of candidates [89]. Using colony size as a proxy for fitness, another quantitative SGA has been

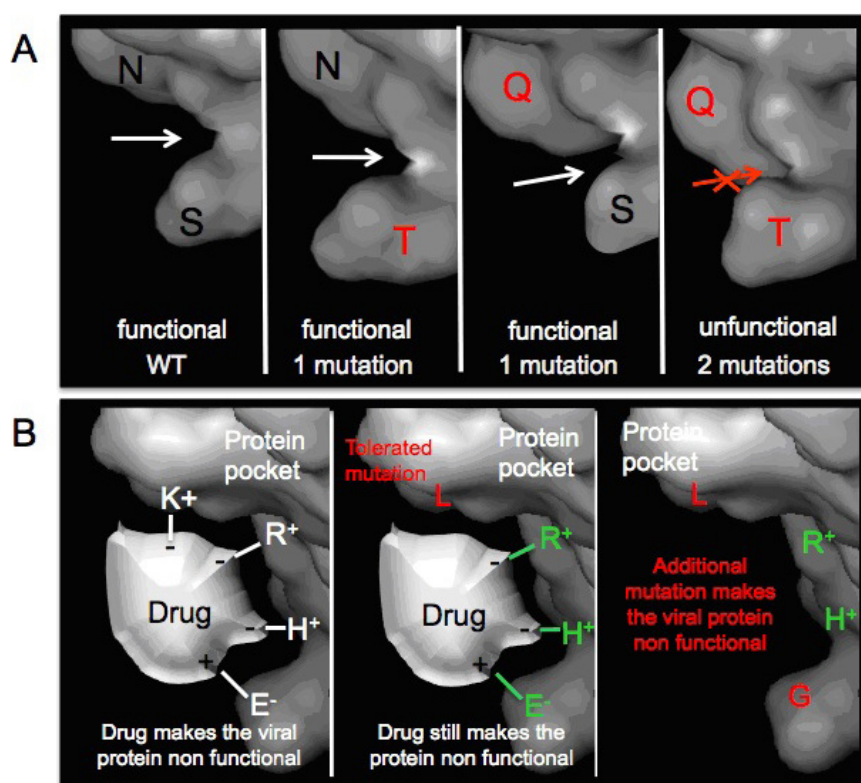


Figure 4: SL behavior ; A: A pair of SLs; B: A group of SLs

The model explains that if binding and function are located on the same amino acids, this site makes an excellent therapeutic target.

N, S, T, Q, K⁺, R⁺, -H⁺, E⁻, L, and G: amino acids

described by Baryshnikova [85]. An elegant model for inferring compensatory pathways from SGA results in yeast was described by MA et al. [27].

Cancer

At the beginning of the 21st century, most SL interactions had been discovered in yeast and *Caenorhabditis elegans*, but few had been described in mammalian cells, likely owing to the lack of availability of efficient tools to identify such interactions in these cells. This new field of discovery of SL pairs provides new perspectives on cellular functions and facilitates discovery of new therapeutic approaches. Indeed, drugs act on pathogenic targets and on any proteins in the host body that have a function related to the target, producing significant side effects. Study of SL allows identification of new targets that may provoke less damage in cells. This method has largely been applied to cancer therapy during the first decade of the 21st century. In an SL pair, both proteins must be non-functional to produce a lethal phenotype. In a cancer cell, the first non-functional protein is the one responsible for the cancer, and if another protein in an SL relationship with this first protein can be identified, this new protein would make an excellent target for an anticancer drug. If it were so, both proteins would be non-functional, leading to cancer cell death. However, normal cells would have a wild-type first protein and a non-functional version of the second protein (the protein targeted by the drug). Thus, the drug's effect on normal cells, to render one protein non-functional, would not change the phenotype of the cell, and thus should not induce any secondary effects. The use of SL as a target provides a therapeutic window for the treatment of cancer. In other words, if protein X is SL with a mutated protein Y that causes cancer, a proposed solution is to inhibit X, which will cause the death of the cancer cells but little or no harm to normal cells.

Thus, the objective is double: targeting proteins on cancer cells to destroy them and minimizing the effects of the drugs on non-cancerous cells (Figure 5).

Some cancers develop because of loss-of-function [90, 91]

mutations in tumor suppressor genes. Mammalian genes in an SL relationship with such mutated tumor suppressor genes could provide an excellent therapeutic target [91, 92]. Indeed, a good example for TP53 is described in [93]. As an example, Dicer1 targeting prevents retinoblastoma formation in mice by SL with combined inactivation of p53 and Rb [94]. Blocking protein kinase targets produced selective SL when combined with phosphatidylinositol-3-kinase inhibitors, a molecule that is a malignant glioma signature [95]. MYC oncogene family members have been broadly implicated in human cancers, yet are considered "undruggable" because they encode transcription factors. Experiments by a different research group [96–99] have revealed a rich therapeutic space comprising more than 48 genes in SL relationship with MYC. SL also shows promise for identifying drugs against colorectal cancer, involving the tumor suppressor BRCA1/2 [100] and the DNA repair pathway poly-(ADP-ribose) polymerase 1 (PARP) [101, 102]. The same *BRCA* gene has been studied as a target in breast cancer (review: [103], where it has been shown that it shares a SL relationship with PARP [104]). SL approaches using an siRNA library have shown that NF- κ B signaling pathways were significantly enriched, making this protein a good target in combination with topotecan, an anticancer molecule [105] in neuroblastoma.

Intragenic SLs

Extragenic SLs were discovered first. However, a few intragenic SLs have also been detected [106–109]. The intragenic SLs have been so little studied that they have not yet been named. Rather, they are defined by the negation of the covariance concept: negative coevolved pairs, negative co-variants, and negative epistasis. The principal reason that they have not been investigated is that no role has been assigned to them. Nevertheless, the same role as the one described for extragenic SL in cancer research can be exploited. Indeed, if a pair of SL mutations has the following definition: when both mutations are non-lethal when alone but lethal when combined in one genome, the identical definition can be used when both mutations are

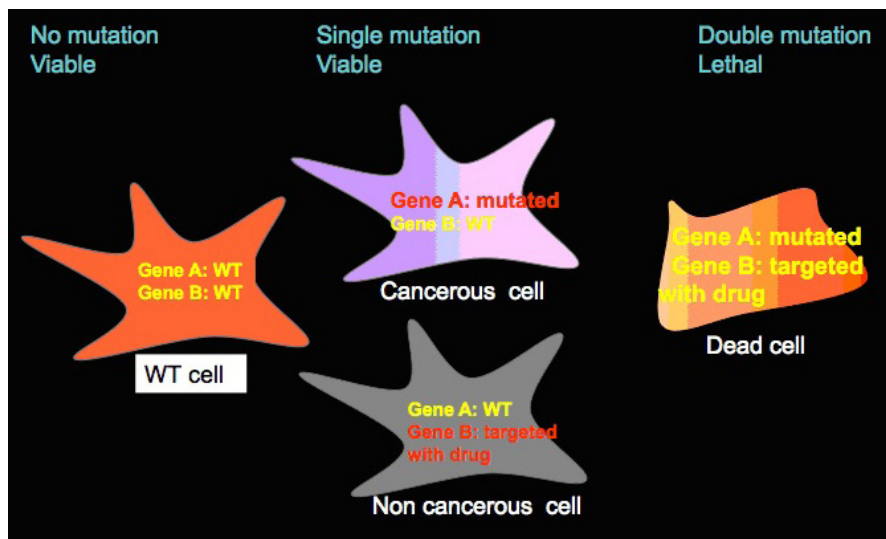


Figure 5: Use of synthetic lethality to target cancerous cells

In this example, gene B is in a SL relationship with gene A, which causes cancer. The goal is to target the product of gene B with a drug. This drug will cause no damage to wild-type (WT) cells, but will kill cancerous cells.

located in the same gene. A functional example is shown in Figure 4A. The mutations could be referred to as intra-SL pairs, and their role may be very similar to the one mentioned in the cancer section. Indeed, to identify good protein targets in cancer research, researchers do not directly target the query protein but a protein, which is in a SL relationship with it. To make a parallel argument, an escape-proof protein target should be comprised of a group of SL positions. If the same group of positions must be mutated to resist a drug and to perform an essential function, this group defines a new “therapeutic” pocket of choices for a protein target (Figure 4B).

This approach was used by Brouillet et al. [110] to define the region of the HIV protease to which it might be appropriate to target potential drugs. On RNA viruses, the main challenge in drug development is the rapid accumulation of resistant mutations. SL can be used to identify new binding sites. Even if resistant mutations appear on these sites, these mutations render the protein inactive [110–112]. These authors used this method to define new therapeutic targets on the HIV protease. Interestingly, Perryman’s team, using molecular dynamics simulations, achieved the same results [113]. The positions involved in SL relationships were then deeply studied, to describe the exact amino acid pairs that are implicated. Indeed, it is often possible to identify several SL pairs composed of various pairs of amino acids, but localized to the same pair of positions. Moreover, to define a complete schematic of dependent positions, an entire network involving SL and invariant positions has been constructed. All proximate pairs of SL positions that localize to the surface of proteins, as well as invariant positions in their vicinity, are shown in the graph in Figure 6 (personal data). This graph is composed of subgraphs containing positions that can be displayed on the three-dimensional structure of the protein of interest (Figure 6), at which point two binding sites for a potential drug can be drawn. Such SL pairs can now be simulated using VIRAPOPS, a forward RNA virus sequence population simulator [114, 115].

The same concept can be applied to the development of vaccines. Indeed, the development of an anti-RNA virus vaccine is an annual public health issue. A vaccine is usually composed of different antigenic epitopes against which the human body develops antibodies that should protect against a future viral infection. Vaccine production is time consuming, time during which the virus can continue to mutate. Thus, the patient may become infected with a virus with different mutations than the one from which the vaccine was developed, and this virus will not necessarily be recognized by the patient’s antibodies. The development of stable peptide vaccines not sensitive to this phenomenon can be achieved using SLs. An escape-proof vaccine

should be composed of different peptides containing positions that are in a SL relationship [116, 117]. In these conditions, the vaccine will be effective against the non-mutated virus and not effective against the mutated virus, but the latter will not be able to reproduce because it will have mutations in positions that perform essential functions.

Towards a bioinformatics analysis

The constant growth of sequence databases makes it easy to identify in silico covariant sites. The correlation between two sites (nucleotides or AAs) has been formalized by Atchley, W.R., et al. [118], who explain that global correlation is due to five different correlations: structure, function, phylogeny, interactions, and stochastic correlations. These first two types, structural and functional, were illustrated above. Phylogenic relationships describe correlations due to a common ancestor and they must be treated differently than the first two. Interaction correlations describe relationships between these three previously mentioned types. Finally, random covariation, due to uneven sequences or unknown random effects is represented by stochastic correlation.

Different methods have been used to describe structural and functional correlations. Some have used different correlation coefficients such as Fisher’s exact test, Lewontin’s D’ coefficient, and the chi squared test. Other methods have used information theory, probabilistic methods, and machine learning approaches [119]. Others have used spectral clustering [120]. Computational predictions could also be used to infer viral fitness from viral sequences by targeting regions that are vulnerable to selection pressure [121].

In the 2000s, an interesting technique was developed to identify the difference between functional covariation and covariation due to a common ancestor (phylogenic correlations). A mutation giving rise to synonymous codon (S) cannot be selected, because selection acts at the protein level, and in the case of a synonymous mutation, the synthesized protein is the same with or without this mutation. However, a non-synonymous mutation (A) could be selected. Thus, if the number of synonymous versus non-synonymous mutations at a given codon is the same, it has not been more selected than not selected. Concerning covariation, the same work can be performed on two codons, showing if this pair of codons gives rise to (A,A) or (S,S) codon pairs [122]. Other algorithms have employed reconstruction of phylogenic trees [123] to determine the difference between functional and common ancestor covariation. Unfortunately, although this type of work can efficiently select covariant pairs of positions, it cannot reveal whether they are CM or SL. However, a special dissimilarity coefficient can be used to determine this.

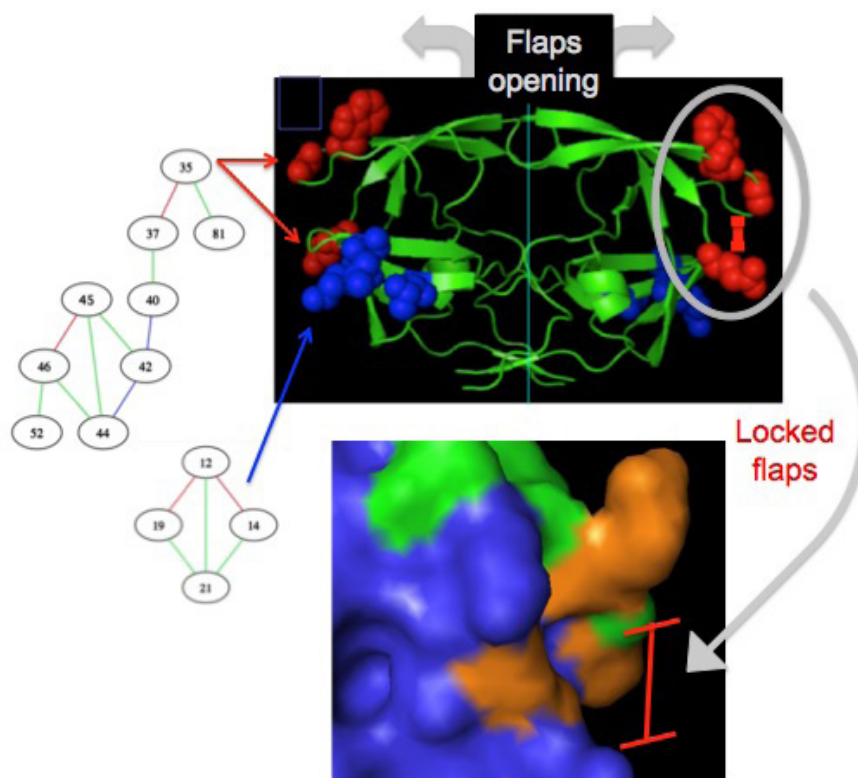


Figure 6: Graphic representation of synthetic lethal (SL) and invariant interactions in HIV protease and a 3D view of potential target sites

The graph is composed of 2 subgraphs derived from our computational analyses. A link between two positions indicates that these two positions are at the surface of the protein and near each other. Moreover, a red link binds two SL positions. A green link binds an SL position to an invariant position. Finally, a blue link binds two invariant positions. The numbers correspond to the HIV protease positions.

The different target sites are shown on a 3D representation of the HIV-1 homodimer (pdb:1HSG) protease. On the ribbon scheme: the red target is lining a pocket, which is necessary for flap opening. The blue target is located at the fulcrum of the protein.

Several bioinformatics methods can be used to detect CM and intergenic SL, but are ineffective for detecting intragenic SL groups. However, as explained above, the concept used to describe new therapeutic targets against cancer could be used to identify targets for rapidly evolving organisms (e.g. RNA viruses). One or several molecules perform functions. For each of these functions, different positions are involved which can be involved in CM and SL relationships. All these data will require further study using large-network bioinformatics analysis. These data will involve CM and SL in the same covariant network, interacting with each other. Large genomic sequences will yield sufficient information on mutation to discover new functional and regulatory pathways, which will give rise to new and precisely targeted therapies.

Conclusions

A very long story about epistasis was told during the 20th century, a story that has not yet ended. These pairs of positions allowed the discovery of whole functional pathways, structural and functional relationships in protein and RNA studies, development of new targets in cancer research, and detection of regulators of genomic expression. They also provided many defenses against pathogenic organisms, without which they could have bypassed medical treatments. These pairs have been found in most of the organisms studied so far. The list of the scientific advances based upon studies of covariant positions is so long that we could believe that the debate is closed. However, in most of the field described in this work one can imagine new perspectives. Much has been done in the study of SGA to find functional pathways using SL pairs, but very few studies have used the same system to detect CM pairs using the SGAM process [16]. Several software

packages employ knowledge of protein structural constraints, to infer evolutionary relationships between several organisms (i.e. HIV) [124]. In fact, covariant positions impose other constraints on protein structure, and can be used to infer evolutionary relationships. The double role of CMs, first in recovering fitness from a deleterious but necessary first mutation, and second, as a way to facilitate gene evolution, remains unclear. Finally, the use of intragenic SLs in the discovery of new stable pockets in protein targets to prevent resistance to drugs and vaccines shows great promise. This development will be most important in the RNA virus realm, where multitherapies rapidly end with failure.

Acknowledgements

We thank Anne-Rozenn Jouble for proofreading this manuscript and P. Le Chien for his encouragement since 1993. This work was conducted for Lou, 5 years old, who died of AIDS in 1997.

Author Contributions

Bazin C: Analyzed and interpreted the data and drafted the work.
 Coupaye R, Middendorp S: Acquired the data and revised and approved the work.
 Vanet A: Conceived and designed the work, analyzed and interpreted the data, drafted the work, revised and approved the work.

References

- Jarvik J, Botstein D (1975) Conditional-lethal mutations that suppress genetic defects in morphogenesis by altering structural proteins. *PNAS* **72**(7): pp. 2738–2742. doi:10.1073/pnas.72.7.2738.
- Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*. **418**(6896): pp. 387–391. doi: 10.1038/nature00935.
- Beckwith JR (1964) A deletion analysis of the Lac operator region in *Escherichia Coli*. *J. Mol. Biol.* **8**: pp. 427–430. doi: 10.1016/S0022-2836(64)80206-0.
- Lehner B (2011) Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27**(8): pp. 323–331. doi: 10.1016/j.tig.2011.05.007.
- Roelants FM, Torrance PD, Bezman N, Thorner J (2002) Pkh1 and Pkh2 differentially phosphorylate and activate Ypk1 and Ykr2 and define protein kinase modules required for maintenance of cell wall integrity. *Mol Biol. Cell* **13**(9): pp. 3005–3028. doi: 10.1091/mbc.E02-04-0201.
- Black WP, Xu Q, Cadieux CL, Suh SJ, Shi W, Yang Z (2009) Isolation and characterization of a suppressor mutation that restores *Myxococcus xanthus* exopolysaccharide production. *Microbiology* **155**(Pt 11): pp. 3599–3610. doi: 10.1099/mic.0.031070-0.
- Krishnakumar S, Oppenheimer DG (1999) Extragenic suppressors of the arabidopsis *zwi-3* mutation identify new genes that function in trichome branch formation and pollen tube growth. *Development* **126**(14): pp. 3079–3088.
- Fagundes MR, Fernandes L, Savoldi M, Harris SD, Goldman MH, Goldman GH (2003) Identification of a topoisomerase I mutant, *scsA1*, as an extragenic suppressor of a mutation in *scaA*(NBS1), the apparent homolog of human nibrin in *Aspergillus nidulans*. *Genetics* **164**(3): pp. 935–945.
- Martinez FP Tang Q (2012) Leucine zipper domain is required for Kaposi sarcoma-associated herpesvirus (KSHV) K-BZIP protein to interact with histone deacetylase and is important for KSHV replication. *J. Biol. Chem.* **287**(19): pp. 15622–15634. doi: 10.1074/jbc.M111.315861.
- Vandeputte P, Ischer F, Sanglard D, Coste AT (2011) In vivo systematic analysis of *Candida albicans* Zn2-Cys6 transcription factors mutants for mice organ colonization. *PLoS One* **6**(10): p. e26962. doi: 10.1371/journal.pone.0026962.
- Marmier-Gourrier N, Cléry A, Schlotter F, Senty-Ségault V, Branlant C (2011) A second base pair interaction between U3 small nucleolar RNA and the 5'-ETS region is required for early cleavage of the yeast pre-ribosomal RNA. *Nucleic. Acids. Res.* **39**(22): pp. 9731–9745. doi: 10.1093/nar/gkr675.
- Chen CL, Perasso R, Qu LH, Amar L (2007) Exploration of pairing constraints identifies a 9 base-pair core within box C/D snoRNA-rRNA duplexes. *J. Mol. Biol.* **369**(3): pp. 771–783. doi: 10.1016/j.jmb.2007.03.052.
- Atkins JF, Bjork GR (2009) A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight P-site realignment. *Microbiol. Mol. Biol. Rev.* **73**(1): pp. 178–210. doi: 10.1128/MMBR.00010-08.
- Sun HY, Ou NY, Wang SW, Liu WC, Cheng TF, Shr SJ, et al. (2011) Novel nucleotide and amino acid covariation between the 5'UTR and the NS2/NS3 proteins of hepatitis C virus: bioinformatic and functional analyses. *PLoS One.* **6**(9): p. e25530. doi: 10.1371/journal.pone.0025530.
- Brandman R, Brandman Y, Pande VS (2012) Sequence coevolution between RNA and protein characterized by mutual information between residue triplets. *PLoS One* **7**(1): p. e30022. doi: 10.1371/journal.pone.0030022.
- Normark BH, Normark S (2002) Evolution and spread of antibiotic resistance. *J. Intern. Med.* **252**(2): pp. 91–106. doi: 10.1046/j.1365-2796.2002.01026.x.
- Costanzo, M. and C. Boone (2009) SGAM: an array-based approach for high-resolution genetic mapping in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **548**: pp. 37–53. doi: 10.1007/978-1-59745-540-4_3.
- McCallum N, Berger-Bachi B, Senn MM (2009) Regulation of antibiotic resistance in *Staphylococcus aureus*. *Int. J. Med. Microbiol.* **300**(2–3): pp. 118–129. doi: 10.1016/j.ijmm.2009.08.015.
- Zhang Q, Sahin O, McDermott PF, Payot S (2006) Fitness of antimicrobial-resistant *Campylobacter* and *Salmonella*. *Microbes. Infect.* **8**(7): pp. 1972–1978. doi: 10.1016/j.micinf.2005.12.031.
- Lunzer, M, Golding GB, Dean AM (2010) Pervasive cryptic epistasis in molecular evolution. *PLoS Genet.* **6**(10): p. e1001162. doi: 10.1371/journal.pgen.1001162.
- Perez-Jimenez R, Wiita A, Rodriguez-Larrea D, Kosuri P, Gavira JA, Sanchez-Ruiz JM, et al. (2008) Force-clamp spectroscopy detects residue co-evolution in enzyme catalysis. *J. Biol. Chem.* **283**(40): pp. 27121–27129. doi: 10.1074/jbc.M803746200.
- Fares MA, Travers SA (2006) A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics* **173**(1): pp. 9–23. doi: 10.1534/genetics.105.053249.
- Gloor G.B, Martin LC, Wahl LM, Dunn SD (2005) Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. *Biochemistry* **44**(19): pp. 7156–7165. doi: 10.1021/bi050293e.
- Murgola EJ (1985) tRNA, suppression, and the code. *Annu. Rev. Genet.* **19**: pp. 57–80. doi: 10.1146/annurev.ge.19.120185.000421.
- Kurland CG, Hughes D, Ehrenberg M (1996) Limitations of translational accuracy. In: F. Neidhardt (ed.) *Escherichia coli ans salmonella: cellular and molecular biology, second Ed.* Washington DC: American society for microbiology. pp. 979–1004. ISBN: 1-55581-084-5.
- Nilsson AI, Zorzet A, Kanth A, Dahlström S, Berg OG, Andersson DI (2006) Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *PNAS* **103**(18): pp. 6976–6981. doi: 10.1073/pnas.0602171103.
- Rokyta D, Badgett MR, Molineux IJ, Bull JJ (2002) Experimental genomic evolution: extensive compensation for loss of DNA ligase activity in a virus. *Mol. Biol. Evol.* **19**(3): pp. 230–238. doi: 10.1093/oxfordjournals.molbev.a004076.
- Ma X, Tarone AM, Li W, (2008) Mapping genetically compensatory pathways from synthetic lethal interactions in yeast. *PLoS One* **3**(4): p. e1922. doi: 10.1371/journal.pone.0001922.
- Davis BH, Poon AF, Whitlock MC (2009) Compensatory mutations are repeatable and clustered within proteins. *Proc. Biol. Sci.* **276**(1663): pp. 1823–1827. doi: 10.1098/rspb.2008.1846.
- Poon A, Chao L (2005) The rate of compensatory mutation in the DNA bacteriophage phiX174. *Genetics* **170**(3): pp. 989–999. doi: 10.1534/genetics.104.039438.
- Poon A, BH Davis, Chao L (2005) The coupon collector and the suppressor mutation: estimating the number of compensatory mutations by maximum likelihood. *Genetics* **170**(3): pp. 1323–1332. doi: 10.1534/genetics.104.037259.
- Chen JM, Cooper DN, Férec C (2014) A new and more accurate estimate of the rate of concurrent tandem-base substitution mutations in the human germline: approximately 0.4% of the single-nucleotide substitution mutation rate. *Hum. Mutat.* **35**(3): pp. 392–394. doi: 10.1002/humu.22501.
- Bazykin GA, Kondrashov FA, Oqurtshov AY, Sunyaev S, Kondrashov AS (2004) Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. *Nature* **429**(6991): pp. 558–562. doi: 10.1038/nature02601.
- Weinreich DM, Dalaney NF, Depristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**(5770): pp. 111–114. doi: 10.1126/science.1123539.
- Gardner PP (2009) The use of covariance models to annotate RNAs in whole genomes. *Brief Funct. Genomic. Proteomic.* **8**(6): pp. 444–450. doi: 10.1093/bfpg/elp042.
- Rivas E, Klein RJ, Jones TA, Eddy SR (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr. Biol.* **11**(17): pp. 1369–1373. doi: 10.1016/S0960-9822(01)00401-8.
- Barash D, Churkin A (2011) Mutational analysis in RNAs: comparing programs for RNA deleterious mutation prediction. *Brief Bioinform.* **12**(2): pp. 104–114. doi: 10.1093/bib/bbq059.
- Maisnier-Patin S, Andersson DI (2004) Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Res. Microbiol.* **155**(5): pp. 360–369. doi: 10.1016/j.resmic.2004.01.019.

39. Hall AR, MacLean RC (2011) Epistasis buffers the fitness effects of rifampicin- resistance mutations in *Pseudomonas aeruginosa*. *Evolution* **65**(8): pp. 2370–2379. doi: 10.1111/j.1558-5646.2011.01302.x.
40. Martinez-Picado J, Martinez MA (2008) HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo. *Virus Res.* **134**(1-2): pp. 104–123. doi: 10.1016/j.virusres.2007.12.021.
41. Singh-Babak SD, Babak T, Diezmann S, Hill JA, Xie JL, Chen YL, et al. (2012) Global analysis of the evolution and mechanism of echinocandin resistance in *Candida glabrata*. *PLoS Pathog.* **8**(5): p. e1002718. doi: 10.1371/journal.ppat.1002718.
42. Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF (2013) Genetic incompatibilities are widespread within species. *Nature* **504**(7478): pp. 135–137. doi:10.1038/nature12678.
43. Pollock DD (2002) Genomic biodiversity, phylogenetics and coevolution in proteins. *Appl. Bioinformatics* **1**(2): pp. 81–92.
44. Yanofsky C, Horn V, Thorpe D (1964) Protein Structure Relationships Revealed by Mutational Analysis. *Science* **146**(3651): pp. 1593–1594. doi: 10.1126/science.146.3651.1593.
45. Camps M, Herman A, Loh E, Loeb LA (2007) Genetic constraints on protein evolution. *Crit. Rev. Biochem. Mol. Biol.* **42**(5): pp. 313–326. doi: 10.1080/10409230701597642.
46. Harris EE (2010) Nonadaptive processes in primate and human evolution. *Am. J. Phys. Anthropol.* **143** (S51): pp. 13–45. doi: 10.1002/ajpa.21439.
47. Gloor GB, Gloor GB, Tyagi G, Abrassart DM, Kingston AJ, Fernandes AD, Dunn SD et al. (2010) Functionally compensating coevolving positions are neither homoplastic nor conserved in clades. *Mol. Biol. Evol.* **27**(5): pp. 1181–1191. doi: 10.1093/molbev/msq004.
48. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA (2012) Epistasis as the primary factor in molecular evolution. *Nature* **490**(7421): pp. 535–538. doi: 10.1038/nature11510.
49. Dobzhansky T (1946) Genetics of natural populations. XIII. Recombination and variability in populations of *Drosophila pseudoobscura*. *Genetics* **31**: pp. 270–289.
50. Lucchesi JC (1967) Synthetic lethality and semi-lethality among functionally related mutants of *Drosophila Melanogaster*. *Genetics* **59**(1): pp. 37–44.
51. Kaiser CA, Schekman R (1990) Distinct sets of SEC genes govern transport vesicle formation and fusion early in the secretory pathway. *Cell* **61**(4): pp. 723–733. doi: 10.1016/0092-8674(90)90483-U.
52. Bender A, Pringle JR (1991) Use of a screen for synthetic lethal and multicopy suppressor mutants to identify two new genes involved in morphogenesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**(3): pp. 1295–1305.
53. Igual JC, Johnson AL, Johnston LH (1996) Coordinated regulation of gene expression by the cell cycle transcription factor Swi4 and the protein kinase C MAP kinase pathway for yeast cell integrity. *Embo. J.* **15**(18): pp. 5001–5013.
54. Singer-Kruger B, Ferro-Novick S (1997) Use of a synthetic lethal screen to identify yeast mutants impaired in endocytosis, vacuolar protein sorting and the organization of the cytoskeleton. *Eur. J. Cell. Biol.* **74**(4): pp. 365–375.
55. Lila T, Drubin DG (1997) Evidence for physical and functional interactions among two *Saccharomyces cerevisiae* SH3 domain proteins, an adenylyl cyclase-associated protein and the actin cytoskeleton. *Mol. Biol. Cell* **8**(2): pp. 367–385.
56. Costigan C, Gehrung S, Snyder M (1992) A synthetic lethal screen identifies SLK1, a novel protein kinase homolog implicated in yeast cell morphogenesis and cell growth. *Mol. Cell. Biol.* **12**(3): pp. 1162–1178. doi: 10.1128/MCB.12.3.1162.
57. Kim YJ, Francisco L, Chen GC, Marcotte E, Chan CS (1994) Control of cellular morphogenesis by the Ip12/Bem2 GTPase-activating protein: possible role of protein phosphorylation. *J. Cell. Biol.* **127**(5): pp. 1381–1394. doi: 10.1083/jcb.127.5.1381.
58. Ota IM, Varshavsky A (1992) A gene encoding a putative tyrosine phosphatase suppresses lethality of an N-end rule-dependent mutant. *PNAS* **89**(6): pp. 2355–2359. doi: 10.1073/pnas.89.6.2355.
59. Wimmer C, Doye V, Grandi P, Nehrbass U, Hurt EC (1992) A new subclass of nucleoporins that functionally interact with nuclear pore protein NSP1. *Embo. J.* **11**(13): pp. 5051–5061.
60. Doye V, Wepf R, Hurt EC (1994) A novel nuclear pore protein Nup133p with distinct roles in poly(A)⁺ RNA transport and nuclear pore distribution. *Embo. J.* **13**(24): pp. 6062–6075.
61. Legrain P, Chapon C (1993) Interaction between PRP11 and SPP91 yeast splicing factors and characterization of a PRP9-PRP11-SPP91 complex. *Science* **262**(5130): pp. 108–110. doi: 10.1101/gad.7.7b.1390.
62. Araki H, Leem SH, Phongdara A, Sugino A (1995) Dpb11, which interacts with DNA polymerase II(epsilon) in *Saccharomyces cerevisiae*, has a dual role in S-phase progression and at a cell cycle checkpoint. *PNAS* **92**(25): pp. 11791–11795. doi: 10.1073/pnas.92.25.11791.
63. Hardy CF (1996) Characterization of an essential Orc2p-associated factor that plays a role in DNA replication. *Mol. Cell. Biol.* **16**(4): pp. 1832–1841.
64. Merrill BJ, Holm C, (1998) The RAD52 recombinational repair pathway is essential in pol30 (PCNA) mutants that accumulate small single-stranded DNA fragments during DNA synthesis. *Genetics* **148**(2): pp. 611–624.
65. Boisrime A, Beckerich JM, Gaillardin C (1996) Sls1p, an endoplasmic reticulum component, is involved in the protein translocation process in the yeast *Yarrowia lipolytica*. *J. Biol. Chem.* **271**(20): pp. 11668–11675. doi: 10.1074/jbc.271.20.11668.
66. Sapperstein SK, Lupashin VV, Schmitt HD, Waters MG (1996) Assembly of the ER to Golgi SNARE complex requires Uso1p. *J. Cell. Biol.* **132**(5): pp. 755–767. doi: 10.1083/jcb.132.5.755.
67. Chen CY, Graham TR (1998) An arf1Delta synthetic lethal screen identifies a new clathrin heavy chain conditional allele that perturbs vacuolar protein transport in *Saccharomyces cerevisiae*. *Genetics* **150**(2): pp. 577–589.
68. Grandin N, Reed SI, Charbonneau M (1997) Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with Cdc13. *Genes. Dev.* **11**(4): pp. 512–527. doi: 10.1101/gad.11.4.512.
69. Igual JC, Toone WM, Johnston LH (1997) A genetic screen reveals a role for the late G1-specific transcription factor Swi4p in diverse cellular functions including cytokinesis. *J. Cell. Sci.* **110** (14): pp. 1647–1654.
70. Tarun SZ Jr, Wells SE, Deardorff JA, Sachs AB (1997) Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. *PNAS* **94**(17): pp. 9046–9051. doi: 10.1073/pnas.94.17.9046.
71. Efimov VP, Morris NR (1998) A screen for dynein synthetic lethals in *Aspergillus nidulans* identifies spindle assembly checkpoint genes and other genes involved in mitosis. *Genetics* **149**(1): pp. 101–116.
72. Munoz MJ, Jimenez J (1999) Genetic interactions between Hsp90 and the Cdc2 mitotic machinery in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* **261**(2): pp. 242–250. doi: 10.1007/s004380050963.
73. Kano H, Russell P (1998) The protein kinase Cdr2, related to Nim1/Cdr1 mitotic inducer, regulates the onset of mitosis in fission yeast. *Mol. Biol. Cell* **9**(12): pp. 3321–3334. doi: 10.1091/mbc.9.12.3321.
74. Chang F, Drubin D, Nurse P (1997) cdc12p, a protein required for cytokinesis in fission yeast, is a component of the cell division ring and interacts with profilin. *J. Cell. Biol.* **137**(1): pp. 169–182. doi: 10.1083/jcb.137.1.169.
75. Haarer BK, Corbett A, Kweon Y, Petzold AS, Silver P, Brown SS (1996) SEC3 mutations are synthetically lethal with profilin mutations and cause defects in diploid-specific bud-site selection. *Genetics* **144**(2): pp. 495–510.
76. Luukkonen BG, Seraphin B (1998) A role for U2/U6 helix Ib in 5' splice site selection. *RNA* **4**(8): pp. 915–927. doi: 10.1017/S1355838298980591.
77. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. (1996) Life with 6000 genes. *Science* **274**(5287): pp. 546, 563–567. doi: 10.1126/science.274.5287.546.
78. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**(5429): pp. 901–906. doi: 10.1126/science.285.5429.901.
79. Tong AH, Boone C (2006) Synthetic genetic array analysis in *Saccharomyces cerevisiae*. In: Xiao W (ed.) *Yeast Protocol*. Totowa, NJ: Humana Press, ISSN: 978-1588294731.

80. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**(5550): pp. 2364–2368. doi: 10.1126/science.1065810.
81. Baetz K, Measday V, Andrews B (2006) Revealing hidden relationships among yeast genes involved in chromosome segregation using systematic synthetic lethal and synthetic dosage lethal screens. *Cell Cycle* **5**(6): pp. 592–595. doi: 10.4161/cc.5.6.2583.
82. Measday V, Baetz K, Guzzo J, Yuen K, Kwok T, Sheikh B, et al. (2005) Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. *PNAS* **102**(39): pp. 13956–13961. doi: 10.1073/pnas.0503504102.
83. Ragni E, Piberger H, Neupert C, Garcia-Cantalejo J, Popolo L, Arroyo J, et al. (2011) The genetic interaction network of CCW12, a *Saccharomyces cerevisiae* gene required for cell wall integrity during budding and formation of mating projections. *BMC Genomics* **12**: p. 107. doi: 10.1186/1471-2164-12-107.
84. Koh JL, Ding H, Costanzo M, Baryshnikova A, Toufighi K, Bader GD, et al. (2009) DRYGIN: a database of quantitative genetic interaction networks in yeast. *Nucleic Acids Res.* **38**(Database issue): pp. D502–D507. doi: 10.1093/nar/gkp820.
85. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, et al. (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* **7**(12): pp. 1017–1024. doi: 10.1038/nmeth.1534.
86. Kainth P, Andrews B (2009) Quantitative cell array screening to identify regulators of gene expression. *Brief Funct. Genomics* **9**(1): pp. 13–23. doi: 10.1093/bfpg/elp047.
87. Kainth P, Sassi HE, Peña-Castillo L, Chua G, Hughes TR, Andrews B (2009) Comprehensive genetic analysis of transcription factor pathways using a dual reporter gene system in budding yeast. *Methods* **48**(3): pp. 258–264. doi: 10.1016/j.ymeth.2009.02.015.
88. Dixon SJ, Fedyshyn Y, Koh JL, Prasad TS, Chahwan C, Chua G, et al. (2008) Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *PNAS* **105**(43): pp. 16653–16658. doi: 10.1073/pnas.0806261105.
89. Ooi SL, Shoemaker DD, Boeke JD (2003) DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. *Nat. Genet.* **35**(3): pp. 277–286. doi: 10.1038/ng1258.
90. Kaelin WG Jr (1999) Choosing anticancer drug targets in the postgenomic era. *J. Clin. Invest.* **104**(11): pp. 1503–1506. doi: 10.1172/JCI8888.
91. Reddy A, Kaelin WG Jr (2002) Using cancer genetics to guide the selection of anticancer drug targets. *Curr. Opin. Pharmacol.* **2**(4): pp. 366–373. doi: 10.1016/S1471-4892(02)00178-9.
92. Friend SH, Oliff A (1998) Emerging uses for genomic information in drug discovery. *N. Engl. J. Med.* **338**(2): pp. 125–126. doi: 10.1056/NEJM199801083380211.
93. Brummelkamp TR, Bernards R (2003) New tools for functional mammalian cancer genetics. *Nat. Rev. Cancer.* **3**(10): pp. 781–789. doi: 10.1038/nrc1191.
94. Nittner D, Lambert I, Clermont F, Mestdagh P, Köhler C, Nielsen SJ, et al. (2012) Synthetic lethality between Rb, p53 and Dicer or miR-17-92 in retinal progenitors suppresses retinoblastoma formation. *Nat. Cell Biol.* **14**(9): pp. 958–965. doi: 10.1038/ncb2556.
95. Cheng CK, Gustafson WC, Charron E, Houseman BT, Zunder E, Goga A, et al. (2012) Dual blockade of lipid and cyclin-dependent kinases induces synthetic lethality in malignant glioma. *PNAS* **109**(31): pp. 12722–12727. doi: 10.1073/pnas.1202492109.
96. Toyoshima M, Howie HL, Imakura M, Walsh RM, Annis JE, Chang AN, Frazier J, et al. (2012) Functional genomics identifies therapeutic targets for MYC-driven cancer. *PNAS* **109**(24): pp. 9545–9550. doi: 10.1073/pnas.1121119109.
97. Kessler JD, Kahle KT, Sun T, Meerbrey KL, Schlabach MR, Schmitt EM, et al. (2011) A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science* **335**(6066): pp. 348–353. doi: 10.1126/science.1212728.
98. Yang D, Liu H, Goga A, Kim S, Yuneva M, Bishop JM (2010) Therapeutic potential of a synthetic lethal interaction between the MYC proto-oncogene and inhibition of aurora-B kinase. *PNAS* **107**(31): pp. 13836–13841. doi: 10.1073/pnas.1008366107.
99. Kang J, Sergio CM, Sutherland RL, Musgrove EA (2014) Targeting cyclin-dependent kinase 1 (CDK1) but not CDK4/6 or CDK2 is selectively lethal to MYC-dependent human breast cancer cells. *BMC Cancer* **14**(1): p. 32. doi:10.1186/1471-2407-14-32.
100. Rosen EM, Pishvaian MJ (2014) Targeting the BRCA1/2 Tumor Suppressors. *Curr. Drug Targets* **15**(1): pp. 17–31.
101. Soncini D, Caffa I, Patrone F, Ballestrero A, Nencioni A (2012) Synthetic lethality-based therapeutics: perspectives for applications in colorectal cancer. *Curr. Cancer Drug Targets* **12**(4): pp. 329–338. doi: 10.2174/156800912800190938.
102. Liu X (2013) PARP inhibition as a prototype for synthetic lethal screens. *Methods Mol. Biol.* **986**: pp. 123–137. doi: 10.1007/978-1-62703-311-4_9.
103. Rehman FL, Lord CJ, Ashworth A (2010) Synthetic lethal approaches to breast cancer therapy. *Nat. Rev. Clin. Oncol.* **7**(12): pp. 718–724. doi: 10.1038/nrclinonc.2010.172.
104. Pessetto ZY, Yan Y, Bessho T, Natarajan A (2012) Inhibition of BRCT(BRCA1)-phosphoprotein interaction enhances the cytotoxic effect of olaparib in breast cancer cells: a proof of concept study for synthetic lethal therapeutic option. *Breast. Cancer Res. Treat.* **134**(2): pp. 511–517. doi: 10.1007/s10549-012-2079-4.
105. Tsang PS, Cheuk AT, Chen QR, Song YK, Badgett TC, Wei JS, et al. (2012) Synthetic lethal screen identifies NF-kappaB as a target for combination therapy with topotecan for patients with neuroblastoma. *BMC Cancer* **12**: p. 101. doi: 10.1186/1471-2407-12-101.
106. Hoffman NG, Schiffer CA, Swanstrom R (2003) Covariation of amino acid positions in HIV-1 protease. *Virology* **314**(2): pp. 536–548. doi: 10.1016/S0042-6822(03)00484-7.
107. Rhee SY, Liu TF, Holmes SP, Shafer RW (2007) HIV-1 subtype B protease and reverse transcriptase amino acid covariation. *PLoS Comput. Biol.* **3**(5): p. e87. doi: 10.1371/journal.pcbi.0030087.
108. Haq O, Levy RM, Morozov AV, Andreic M (2009) Pairwise and higher-order correlations among drug-resistance mutations in HIV-1 subtype B protease. *BMC Bioinformatics* **10** (S8): p. S10. doi:10.1186/1471-2105-10-S8-S10.
109. Wu TD, Schiffer CA, Gonzales MJ, Taylor J, Kantor R, Chou S (2003) Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J. Virol.* **77**(8): pp. 4836–4847. doi: 10.1128/JVI.77.8.4836-4847.2003.
110. Brouillet S, Valere T, Ollivier E, Marsan L, Vanet A (2010) Co-lethality studied as an asset against viral drug escape: the HIV protease case. *Biol. Direct.* **5**: p. 40. doi:10.1186/1745-6150-5-40.
111. Vanet A, Muller-Trutwin M, Valère T, Brouillet S, Ollivier E, Marsan L, inventor. Method for identifying combinations of motifs that do not mutate simultaneously In a set of viral polypeptide sequences comprising a putative drug binding site. Patent US 7917303 B2. Mar 29 2011. Available from: http://www.lens.org/lens/patent/US_7917303_B2.
112. Vanet A, Muller-Trutwin M, Valère T, inventor. Method for identifying motifs and/or combinations of motifs having a boolean state of predetermined mutation in a set of sequences and its applications. Patent US 7734421 B2, Jun 8 2010. Available from: http://www.lens.org/lens/patent/US_7734421_B2.
113. Perryman AL, Lin JH, McCammon JA (2006) Restrained molecular dynamics simulations of HIV-1 protease: the first step in validating a new target for drug design. *Biopolymers* **82**(3): pp. 272–284. doi: 10.1002/bip.20497.
114. Petitjean M, Vanet A (2014) VIRAPOPS: A forward simulator dedicated to rapidly evolved viral population. *Bioinformatics* **30**(4): pp. 578–580. doi: 10.1093/bioinformatics/btt724.
115. Petitjean M, Vanet A (2014) VIRAPOPS2 supports the influenza virus reassortments. *Source Code Biol. Med.* **9**(1): 18.
116. Vanet A, Muller-Trutwin M, Valère T. Method for identifying motifs and/or combinations of motifs having a boolean state of predetermined mutation in a set of sequences and its application. Patent US 8032309 B2, Oct 4 2011. Available from: http://www.lens.org/lens/patent/US_8032309_B2.
117. Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, Talsania S, et al. (2011) Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. *PNAS* **108**(28): pp. 11530–11535. doi: 10.1073/pnas.1105315108.

118. Atchley WR, Wollenberg KR, Fitch WM, Terhalle W, Dress AW (2000) Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. *Mol. Biol. Evol.* **17**(1): pp. 164–178. doi: 10.1093/oxfordjournals.molbev.a026229.
119. Horner DS, Pirovano W, Pesole G (2008) Correlated substitution analysis and the prediction of amino acid structural contacts. *Brief Bioinform.* **9**(1): pp. 46–56. doi: 10.1093/bib/bbm052.
120. Liu Y, Eyal E, Bahar I (2008) Analysis of correlated mutations in HIV-1 protease using spectral clustering. *Bioinformatics* **24**(10): pp. 1243–1250. doi: 10.1093/bioinformatics/btn110.
121. Ferguson AL, Mann JK, Omarjee S, Ndung'u T, Walker BD, Chakraborty AK (2013) Translating HIV Sequences into Quantitative Fitness Landscapes Predicts Viral Vulnerabilities for Rational Immunogen Design. *Immunity* **38**(3): pp. 606–617. doi: 10.1016/j.immuni.2012.11.022.
122. Wang Q, Lee C (2007) Distinguishing functional amino acid covariation from background linkage disequilibrium in HIV protease and reverse transcriptase. *PLoS One* **2**(8): p. e814. doi: 10.1371/journal.pone.0000814.
123. Noivirt O, Eisenstein M, Horovitz A (2005) Detection and reduction of evolutionary noise in correlated mutation analysis. *Protein. Eng. Des. Sel.* **18**(5): pp. 247–253. doi: 10.1093/protein/gzi029.
124. Williams SG, Madan R, Norris MG, Archer J, Mizuguchi K, Robertson DL (2011) Using knowledge of protein structural constraints to predict the evolution of HIV-1. *J. Mol. Biol.* **410**(5): pp. 1023–1034. doi: 10.1016/j.jmb.2011.04.037.