Mapping genetic interaction networks in yeast

by

Anastasia Baryshnikova

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Molecular Genetics
University of Toronto

Abstract

Mapping genetic interaction networks in yeast

Anastasia Baryshnikova

Doctor of Philosophy, 2013

Department of Molecular Genetics

University of Toronto

Global quantitative analysis of genetic interactions provides a powerful approach for deciphering the roles of genes and mapping functional relationships amongst pathways. Using colony size as a proxy for fitness, I developed a method for measuring genetic interactions from high-density arrays of yeast double mutants generated by synthetic genetic array (SGA) technology. I identified several experimental sources of systematic variation and developed normalization strategies to obtain accurate fitness measurements. I used this scoring method to map quantitative genetic interactions among 5.4 million yeast double mutants and generated the first functionally unbiased genetic interaction map of a eukaryotic cell. My map produced an unprecedented view of the cell in which genes of similar biological processes cluster together in coherent subsets and functionally interconnected bioprocesses map next to each other. We discovered several physiological and evolutionary gene features that are characteristic of genetic interaction hubs, and explored the relationship between genetic and protein-protein interaction networks. In particular, by comparing quantitative single and double mutant phenotypes, we identified specific cases of positive genetic interactions, termed genetic suppression, and constructed a global network of suppression interactions among protein complexes. I also demonstrated that an extensive and unbiased mapping of genetic interactions provides a key for interpreting chemical-genetic interactions and identi-

fying drug targets. In addition, I used genome-wide SGA data to map profiles of genetic linkage along all sixteen yeast chromosomes. These linkage profiles recapitulated previously identified recombination patterns and uncovered an unexpected correlation between chromosome length and the extent of centromere-related recombination repression. These findings suggest a chromosome size-dependent mechanism for ensuring proper chromosome segregation and highlight the SGA methodology as a unique approach for systematic analysis of yeast meiotic recombination.

Acknowledgements

First and foremost, I would like to thank my supervisor Charlie Boone for being a great mentor, in science and in life, and for making things happen. I am very proud of being his student.

I was incredibly lucky to join Charlie's laboratory right when a group of great scientists was about to start a great project. As a new member of that team, I worked side by side with Michael Costanzo and Chad Myers, two of the smartest and most dedicated people I have ever met, who changed forever the way I think and do. I would like to thank them both for guiding me through these incredible six years, for never letting me give up and showing me how much I am capable of.

It was a privilege to be around and work with all the great people who populated the Donnelly Centre over the years, especially my co-supervisor Gary Bader, my committee members Howard Lipshitz and Quaid Morris, past and present members of the Boone and Bader laboratories. A very special thanks goes to Elena Lissina, Iain Wallace and Magali Michaut, colleagues and friends, who brightened up the darkest days of grad school and helped me stay human.

I am also very grateful to the Costanzo family (Julie, Olivia, Sienna, Angela and Salvatore Costanzo) for inviting me into their wonderful house and making me feel at home even when home was thousands of kilometers away.

Finally, I thank my amazing family for giving me strength and having faith in me.

Table of Contents

1 Introduction 1
1.1 Defining genetic interactions 3
1.2 Negative and positive genetic interactions 5
1.3 Mapping genetic interactions in yeast 10
1.3.1 Mutant strain libraries 10
1.3.2 Genetic interaction mapping technologies 13
1.3.3 Phenotypes 19
1.4 Mapping genetic interactions in other model organisms 22
1.5 Mapping genetic interactions in mammalian cell lines 26
1.6 Mapping genetic interactions in human populations 28
1.7 Thesis rationale 30

2 Measuring quantitative fitness and genetic interactions in yeast on a genome scale 34
2.1 Abstract 35
2.2 Introduction 36
2.3 A quantitative fitness-based model for genetic interactions 37
2.4 SGA genetic interaction score 41
2.4.1 Plate effect 42
2.4.2 Row/column effect 45
2.4.3 Spatial effect 46
2.4.4 Competition effect 47
2.4.5 Batch effect 48
2.4.6 Linkage filter 50
2.4.7 Large colony filter 51
2.4.8 Jackknife filter 51
2.4.9 Estimation of fitness, genetic interactions and confidence measures 51
2.4.10 Post-interaction filters 54
2.5 A high-resolution catalog of yeast single mutant fitness 55
2.6 Evaluation of the reproducibility of genetic interaction measurements 58
2.7 Functional impact of systematic effects 60
2.8 Comparative analysis of function prediction capacity 67
2.9 SGA score application to other functional genomics datasets 69
2.10 Materials & methods 72
2.10.1 SGA query construction and screening 72
2.10.2 Precision/recall analysis against functional standards 72
2.10.3 Estimation of false negative and false discovery rates 73
2.10.4 Comparative analysis of function prediction capacity 75
2.10.5 Data access 78

3 The genetic landscape of a cell 79
3.1 Abstract 80
3.2 Introduction 81
3.3 A functional map of the cell 82

3.4 Predicting function and relations 84
3.5 Deciphering complex regulatory relations from the global genetic network 87
3.6 Genetic network connectivity 88
3.7 Distribution of genetic interactions by bioprocess 92
3.8 Overlap between genetic and protein-protein interaction network 96
3.9 Analysis of genetic interactions within and between protein complexes 97
3.10 Defining subclasses of positive genetic interactions 101
3.11 Integrating genetic and chemical-genetic interaction networks 109
3.12 Summary & conclusions 112
3.13 Materials & methods 114
3.13.1 Identifying subnetworks from the global map 114
3.13.2 GINECA clustering algorithm 115
3.13.3 Gap1 permease localization and activity 115
3.13.4 Identification of Sgt2 and Ubp15 physical interactors 116
3.13.5 Pex15 localization 118
3.13.6 Elp/Urm amino acid usage analysis 118
3.13.7 Identification of genetic interaction hubs and monochromatic genes 118
3.13.8 Functional enrichment analysis of genetic interaction hubs and monochromatic genes 119
3.13.9 Biological process annotations 119
3.13.10 Correlation analysis of genetic and physical interaction degree 120
3.13.11 Analysis of variation in genetic interactions number across bioprocesses 124
3.13.12 Analysis of genetic interactions within and between protein complexes 124
3.13.13 Protein complex suppression network 126
3.13.14 Suppression confirmation experiments 127
3.13.15 Integrating genetic and chemical-genetic interaction data 128
3.13.16 Ero1 *in vitro* activity and CPY processing assays 129

4 Mapping genome-wide recombination patterns using SGA 130
4.1 Abstract 131
4.2 Introduction 132
4.3 Mapping genetic linkage using SGA 137
4.4 Recovering genetic interactions from linked gene pairs 138
4.5 Estimating genetic distances from colony size data 144
4.6 Validation of recombination patterns 148
4.7 Characterization of genetic linkage patterns in pericentric regions 155
4.8 Discussion 158
4.9 Materials & methods 162
4.9.1 Overlap with previous genetic interactions reported in BioGRID 162
4.9.2 Recovering genetic interactions from genetically linked gene pairs 163
4.9.3 Calculation of genetic distances 163
4.9.4 Reproducibility of genetic linkage profiles 166
4.9.5 Construction of consolidated genetic linkage maps 166

4.9.6 Hotspot analysis 167

5 Conclusions, discussion & future directions 168

5.1 Thesis summary 169
5.2 Next steps in genetic interaction analysis 171
5.2.1 Decoding genetic redundancy 171
5.2.2 Mapping complex phenotypes in yeast and in higher organisms 175
5.2.3 Stochastic effects in genotype to phenotype mapping 176
5.2.4 The global functional map of *S. cerevisiae* 178
5.3 Yeast as a model for genome organization 179
5.4 Mechanisms of recombination inhibition within pericentric regions 183
5.5 Conclusions 186

6 Appendix 187

6.1 Theoretical evidence for adopting the multiplicative model of genetic interactions for fitness 187

List of figures

Figure 1.1. Defining genetic interactions. 4

Figure 1.2. Molecular mechanisms underlying genetic interactions. 7

Figure 1.3. The yeast deletion-mutant collection. 11

Figure 1.4. Synthetic Genetic Array (SGA) methodology. 15

Figure 1.5. Diploid Synthetic Lethal Analysis by Microarray (dSLAM) and Genetic Interaction Mapping (GIM) methodologies. 18

Figure 2.1. The SGA score for measuring quantitative genetic interactions. 38

Figure 2.2. Correction of systematic experimental effects. 40

Figure 2.3. Comparison of single mutant fitness measures. 56

Figure 2.4. Reproducibility of double mutant fitness measurements. 57

Figure 2.5. Reproducibility of quantitative genetic interactions. 59

Figure 2.6. Functional evaluation of negative and positive genetic interactions. 62

Figure 2.7. Functional evaluation of genetic interaction profile similarities. 63

Figure 2.8. Relative contributions of each systematic effect 64

Figure 2.9. Characterization of SGA genetic interaction score thresholds. 66

Figure 2.10. Functional evaluation of negative and positive genetic interaction profile similarities. 67

Figure 2.11. Comparative function prediction capacity of the genetic interaction dataset.69

Figure 3.1. A correlation-based network connecting genes with similar genetic interaction profiles. 83

Figure 3.2. Magnification of the functional map better resolves cellular processes. 85

Figure 3.3. Genetic interaction profiles reflect diverse types of functional relationship. 86

Figure 3.4. Genetic network connectivity. 89

Figure 3.5. Genetic interaction degree, fitness, multifunctionality, and pleiotropy. 90

Figure 3.6. Distribution of genetic and physical interactions by bioprocess. 94

Figure 3.7. Gene-specific factors explaining the variation in genetic interaction number across biological processes. 95

Figure 3.8. Overlap of genetic and physical interaction networks..... 97

Figure 3.9. Analysis of genetic interactions within protein complexes. 98

Figure 3.10. Analysis of genetic interactions between protein complexes. 100

Figure 3.11. Cross-complex genetic suppression network..... 102

Figure 3.12. Genetic suppression between Swr1 protein complex and *htz1*Δ 103

Figure 3.13. Genetic suppression between Rim101 signaling and MVB sorting pathways.104

Figure 3.14. Genetic suppression between FAR and TORC2 protein complexes. 106

Figure 3.15. Genetic suppression between Polδ and COG protein complex. 108

Figure 3.16. Integrating genetic and chemical-genetic interaction networks..... 111

Figure 4.1. Mechanism of meiotic recombination. 133

Figure 4.2. Segregation of unlinked and linked loci during meiosis..... 138

Figure 4.3. Recovering genetic interactions from genetically linked gene pairs. 141

Figure 4.4. Intra-complex negative genetic interactions involving linked gene pairs. .. 143

Figure 4.5. Relationship between genetic and physical distances..... 146

Figure 4.6. Reproducibility of genetic linkage profiles..... 148

Figure 4.7. Genetic linkage maps for all yeast chromosomes..... 149

Figure 4.8. Validation of known recombination hotspots. 150

Figure 4.9. Comparison of recombination profiles across various experimental approaches. 152

Figure 4.10. Overlap between recombination peaks across various experimental approaches. 153

Figure 4.11. Coldspots of recombination. 155

Figure 4.12. Relationship between linkage and chromosome size..... 157

Figure 4.13. Cohesin-dependent linkage formation around centromeres. 160

List of tables

Table 2.1. Sensitivity and precision of SGA genetic interaction scores. 75

Table 2.2. List of datasets used in comparative function prediction analysis. 77

List of equations

Eq. 2.1 41

Eq. 2.2 42

Eq. 2.3 43

Eq. 2.4 43

Eq. 2.5 44

Eq. 2.6 44

Eq. 2.7 46

Eq. 2.8 46

Eq. 2.9 47

Eq. 2.10 49

Eq. 2.11 49

Eq. 2.12 52

Eq. 2.13 52

Eq. 2.14 53

Eq. 2.15 53

Eq. 2.16 54
Eq. 2.17 54
Eq. 3.1 125
Eq. 3.2 126
Eq. 4.1 162
Eq. 4.2 163
Eq. 4.3 164
Eq. 4.4 165
Eq. 4.5 165
Eq. 4.6 165
Eq. 4.7 165
Eq. 4.8 166

1 Introduction

Portions of this chapter have been reprinted or adapted from:

- Costanzo M., Baryshnikova A., VanderSluis B., Andrews B., Myers C. L., Boone C., "Genetic Networks", in press in "Systems Biology, a Handbook", 2012, eds. Dekker J., Vidal M., Walhout A. J. M. (Costanzo *et al.*, 2012)
- Costanzo M., Baryshnikova A., Myers C. L., Andrews B., Boone C., "Charting the Genetic Landscape of a Cell", Current Opinion in Biotechnology, 22:66–74, 2011 (Costanzo *et al.*, 2011)
- Baryshnikova A., Costanzo M., Dixon S., Vizeacoumar F. J., Myers C. L., Andrews B., Boone C., "Synthetic Genetic Array (SGA) Analysis in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*", Methods in Enzymology, 470:145–179, 2010 (Baryshnikova *et al.*, 2010a)
- Dixon S., Costanzo M., Baryshnikova A., Andrews B., Boone C., "Systematic Mapping of Genetic Interaction Networks", Annual Review of Genetics, 43:601–625, 2009 (Dixon *et al.*, 2009)

I wrote or substantially edited all sections presented in this Chapter.

Following the sequencing of the human genome, one of the biggest challenges in human genetics is understanding the relationship between the genotype and the phenotype of an individual. Mendelian disorders, caused by single gene mutations, account only for a small fraction of rare human diseases, while the genetic causes of more common and complex conditions still remain largely unknown. The difficulty in understanding common diseases stems from the complexity of the human genome: each individual carries ~4 million genetic variants and polymorphisms (Frazer *et al.*, 2009), none of which can be pinpointed as the single cause for a given phenotype. Instead, genetic variants appear to interact with one another by enhancing or suppressing each other's effects on quantitative traits and susceptibility to disease. These interactions between genetic variants determine the phenotype arising from a given genotype and may provide the key for explaining the missing heritability of complex traits (Zuk *et al.*, 2012).

The study of genetic interaction networks in human populations is extremely challenging due to the impossibility to isolate the effect of each individual genetic variant. As noted by Richard Lewontin almost 40 years ago, "There is simply no way to make a large number of individuals identically homozygous or heterozygous at one locus while keeping the rest of the genome segregating at random" (Lewontin, 1974). This problem, however, can be approached in model organisms, where inbred populations and systematic mutagenesis projects have enabled researchers to analyze the effects of alleles alone and in combination. In this thesis, I summarize the state of the art of genetic interaction research in model systems and present my own work towards generating and analyzing the properties of quantitative genetic interaction networks in yeast *Saccharomyces cerevisiae*.

1.1 Defining genetic interactions

Geneticists have long recognized that genetic interactions are important for shaping the phenotypic landscape of a population. For example, William Bateson introduced the term "epistasis" (i.e. "standing upon") in 1909 to describe a specific type of genetic interaction whereby one mutation masks the effects of another mutation on fur color in rabbits and mice (Bateson, 1909). The same term was later adopted to describe genetic relationships that often occur among members of the same metabolic pathway, where the action of one enzyme depends on a substrate produced by another enzyme. In 1918 Ronald Fisher introduced a similar term, "epistacy", to define any multi-locus mutant effect that deviates from the sum of the effects of individual loci (Fisher, 1918). Today, the term "genetic interaction" incorporates both Bateson's and Fisher's definitions of epistasis and refers to any unexpected phenotype that cannot be explained by the combined effects of the individual mutations (Figure 1.1).

A genetic interaction between two genes can be experimentally measured by comparing the phenotype of the double mutant to the expected combination of the two single mutant phenotypes. While this may seem relatively straightforward, determining how mutations are expected to combine is not as obvious and has generated several mathematical theories, including the most widely adopted additive and multiplicative models of genetic interactions (Phillips *et al.*, 2000; Mani *et al.*, 2008). In the case of the additive model, each gene is expected to add a specific quantity to the phenotype, such that the phenotype of a double mutant would equal the sum of the two single mutants phenotypes. Alternatively, according to the multiplicative model, each gene is expected to change the phenotype by a specific fraction, and the phenotype of the double mutant would equal the product of the two single mu-

tants phenotypes. The choice of the most appropriate model is dependent on the particular phenotype and scale of measurement (e.g. linear vs. logarithmic) because genes may appear independent on one scale but show a genetic interaction on a different scale (Mani *et al.*, 2008). There exists a theoretical reason for adopting the multiplicative model when measuring interactions based fitness. According to our current understanding of evolutionary mechanisms, linkage disequilibrium will not arise in a population at equilibrium if genes are independent under the multiplicative model (for more details, see Appendix 6.1).

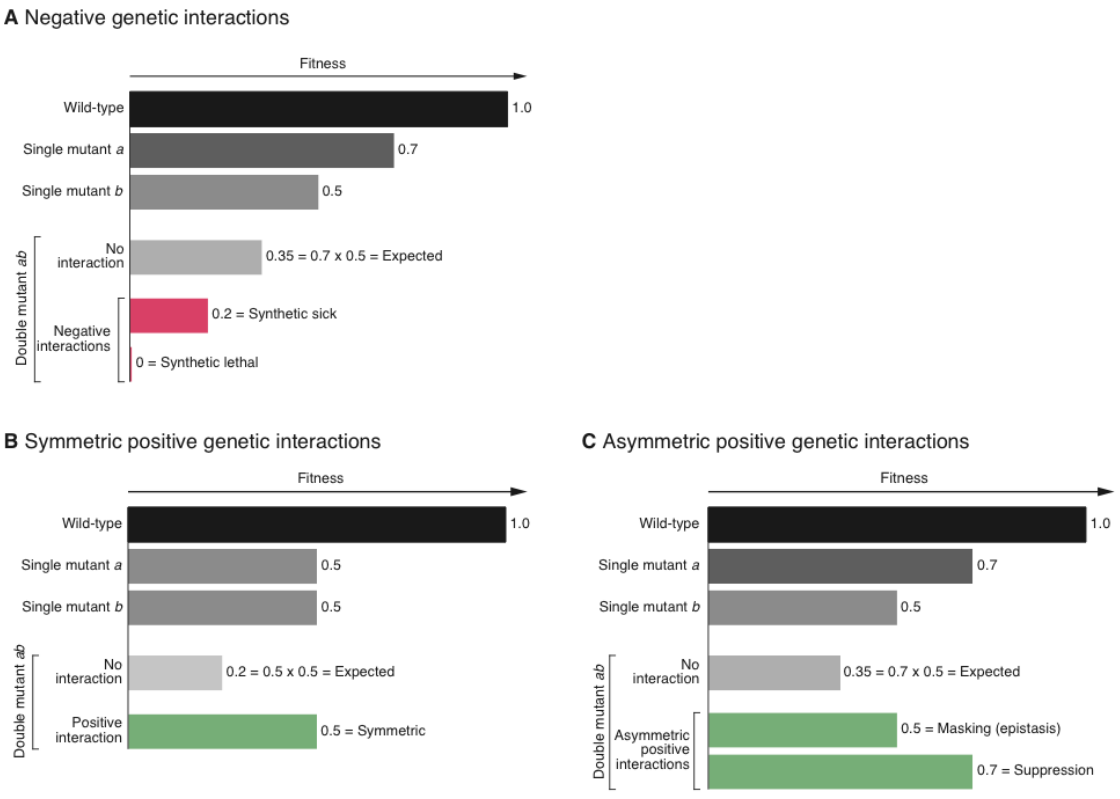


Figure 1.1. Defining genetic interactions.

A graphical representation of how genetic interactions are inferred from a measurable phenotype, such as growth.
(**A**) Negative genetic interactions. The wild-type strain and the single mutants *a* and *b* have fitness of 1.0, 0.7 and 0.5, respectively. Based on the multiplicative model, the expected fitness of the double mutant *ab* would be 0.35. Negative deviations from the expected fitness are scored as either synthetic sick or synthetic lethal interactions.
(**B**) Symmetric positive genetic interactions. Each single mutant (*a* and *b*) exhibits a twofold fitness defect (0.5)

relative to wild-type (1.0). The fitness of the resultant *ab* double mutant is greater than expected (0.5 > 0.25) and identical to the fitness of the two single mutants (0.5).
(**C**) Asymmetric positive genetic interactions. Single mutants *a* and *b* differ in fitness (0.7 and 0.5, respectively). Positive deviations from expectation, along with the comparison of single and double mutant fitness, allow classification of asymmetric positive interactions into masking and suppression subcategories.

1.2 Negative and positive genetic interactions

Based on the difference between the observed and the expected mutant phenotypes, genetic interactions can be broadly divided into two major classes, which we refer to as negative and positive genetic interactions (Baryshnikova *et al.*, 2010b).

Negative genetic interactions describe double mutants whose phenotype is stronger than expected (Figure 1.1) (Phillips *et al.*, 2000; Mani *et al.*, 2008), with the most extreme example being synthetic lethality where two mutations, each causing little fitness defect on their own, result in an inviable phenotype when combined in the same organism. This phenomenon, initially observed among the progeny of intercrosses between natural variants of *Drosophila pseudoobscura*, provided the first insight into the degree of genetic variability concealed within natural populations (Dobzhansky, 1946). Synthetic lethality has since been extensively explored in yeast and other organisms, and proved to be an extremely powerful genetic tool for identifying and characterizing genes in numerous biological processes (Novick and Botstein, 1985; Bender and Pringle, 1991; Guarente, 1993).

Synthetic lethality and its milder variant, synthetic sickness, are interesting because they often occur between genes acting in parallel pathways that impinge on a common essential biological function, such that the cell can survive the loss of either gene alone but not in combination (Figure 1.2A, left panel) (Guarente, 1993; Tong *et al.*, 2001). For example,

many genes involved in response to DNA damage are synthetic lethal with one another (Gurley and Kemp, 2001; Farmer *et al.*, 2005; Pan *et al.*, 2006), possibly reflecting the importance of compensatory systems in maintaining the integrity of the genetic material. Genes acting in the same essential pathway or protein complex may also exhibit synthetic lethality if each mutation is hypomorphic, having a partial inhibitory effect on pathway activity (Figure 1.2A, right panel) (Guarente, 1993; Bandyopadhyay *et al.*, 2008; Baryshnikova *et al.*, 2010b). For example, synthetic lethal interactions among partial loss-of-function mutations in a group of essential *SEC* genes helped elucidating the structure of the post-Golgi secretory pathway and its relationship to other secretion-related processes (Finger and Novick, 2000).

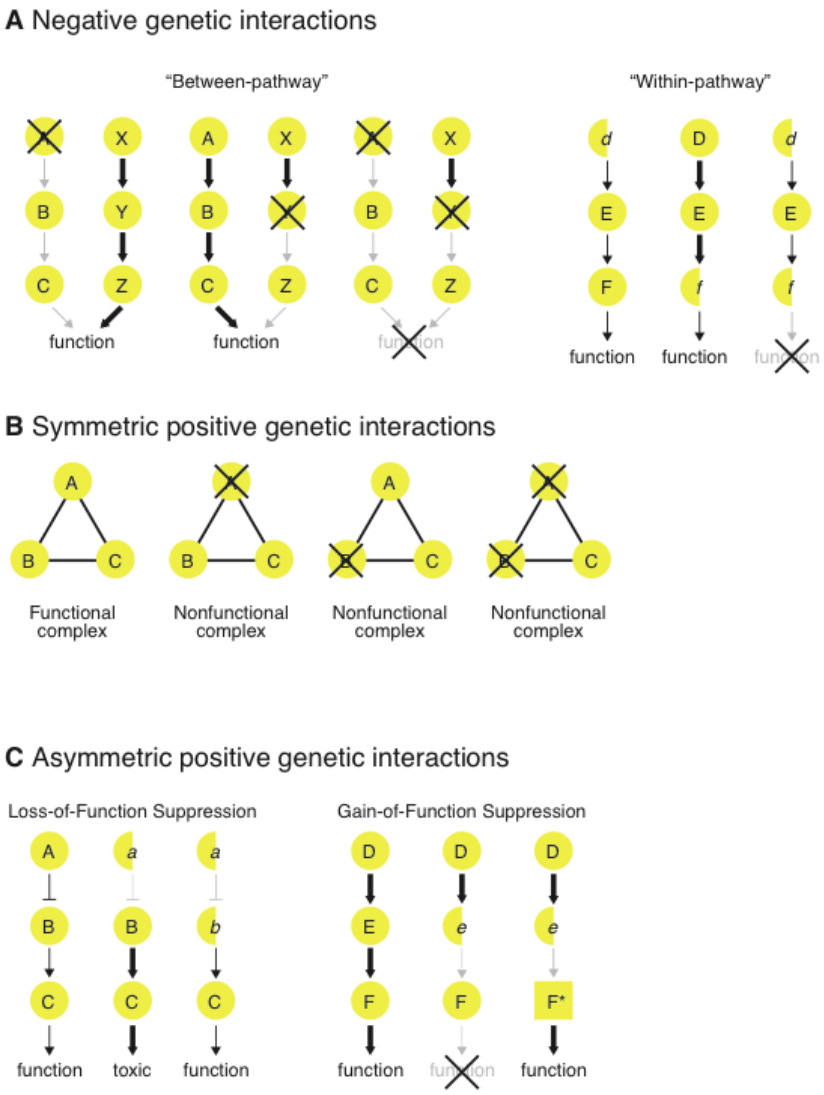


Figure 1.2. Molecular mechanisms underlying genetic interactions.

(**A**) Negative interactions can arise from the disruption of parallel pathways converging on a common process ("between-pathway" genetic interactions) or by decreasing the flux through the same essential pathway ("within-pathway" genetic interactions).
(**B**) Symmetric positive genetic interactions are enriched among members of the same nonessential protein complex, such that the disruption of any one or any pair of complex members results in the same effect on fitness.
(**C**) Asymmetric positive genetic interactions. Mutation of a negative regulator (*a*) leads to hyperactivation of the pathway and accumulation of a toxic gene product (C). Subsequent loss-of-function mutation of a downstream pathway component (*b*) reduces flux through the pathway, thereby suppressing the toxic effects caused by mutant *a*. Gain-of-function suppression may arise when a downstream or terminal pathway component acquires a mutation (F*) such that it is no longer dependent on upstream activation events.

Positive genetic interactions refer to double mutants exhibiting a less severe phenotype than expected, and can be further sub-classified into a variety of categories (Figure 1.1) (Drees *et al.*, 2005; St Onge *et al.*, 2007; Mani *et al.*, 2008). For example, the symmetric class of positive interactions describes cases where the phenotypes of the double mutant and the corresponding single mutants are quantitatively indistinguishable (Figure 1.1B). Conversely, the asymmetric class consists of interactions, such as genetic suppression, in which the phenotypes of the single mutants differ from each other and from the resultant double mutant, which shows a fitness greater than that of the sickest single mutant (Figure 1.1C) (Drees *et al.*, 2005; St Onge *et al.*, 2007).

Importantly, different types of positive genetic interactions are associated with different biological interpretations and, when measured accurately, offer the potential to infer biochemical relationships between gene products (Figure 1.2) (Drees *et al.*, 2005; St Onge *et al.*, 2007; Breslow *et al.*, 2008). For example, genetic suppression, whereby the double mutant fitness is higher than the fitness of the sickest single mutant, is an asymmetric positive genetic interaction that often links genes to their negative regulators, assuming both carry loss-of-function mutations (Figure 1.2C; Chapter 3.10) (Baryshnikova *et al.*, 2010b). Suppression interactions of this kind are often encountered among genes regulating general amino acid biosynthesis in *S. cerevisiae*. For example, growth defects associated with mutations in *GCD1*, which acts as a negative regulator of amino acid biosynthesis upon starvation, are suppressed by mutations in the *GCN4* locus, a tightly regulated transcriptional activator acting downstream of Gcd1 (Hinnebusch and Fink, 1983). The suppressor may also carry a gain-of-function mutation which renders the pathway independent from an upstream compo-

nent (Figure 1.2C). An example is the yeast pheromone response pathway, which is normally triggered by the binding of a/α factor to their corresponding cell surface receptor, which leads to activation of the coupled heterotrimeric G protein, composed by Gpa1, Ste18 and Ste4 (Marsh *et al.*, 1991). Constitutive pathway inactivation, observed in null *ste4* mutants, can be suppressed by dominant mutations in *STE11*, a protein kinase acting in downstream signal transmission (Stevenson *et al.*, 1992).

Symmetric positive interactions, also known as co-equal interactions, often connect genes encoding members of the same non-essential protein complex (Sung, 1997; Ulrich and Jentsch, 2000; Ito *et al.*, 2001; Kaliraman *et al.*, 2001; Shor *et al.*, 2005; St Onge *et al.*, 2007), presumably because the observed phenotype is due to loss of the complex function, which can be achieved by removing any one of its components, individually or in pairs (Figure 1.2B). A systematic analysis of 26 genes involved in DNA damage repair showed that nine out of ten symmetric positive interactions corresponded to physical interactions among the encoded proteins, including all four members of the well-characterized SHU complex (St Onge *et al.*, 2007).

The set of negative and positive genetic interactions involving a particular gene make up its genetic interaction profile, which was found to be an extremely accurate reflection of the gene's biological function (Tong *et al.*, 2004; Ye *et al.*, 2005; Costanzo *et al.*, 2010) (Chapter 3.3). Early genetic interaction studies recognized that genes acting in the same molecular pathway or protein complex tend to share similar genetic interaction partners (Figure 1.2) (Tong *et al.*, 2004). Further analysis showed that common patterns of genetic interactions are often more predictive of co-membership in a protein complex than direct genetic

interactions themselves (Ye *et al.*, 2005). Therefore, genetic interaction profiles, just like gene expression profiles and other multidimensional phenotypes, can be used to cluster genes into functional modules and predict function for previously uncharacterized genes using guilt-by-association approaches (Costanzo *et al.*, 2010) (Chapter 3.3).

1.3 Mapping genetic interactions in yeast

The vast majority of large-scale genetic interaction screens completed to date have used the budding yeast *Saccharomyces cerevisiae*. With its elegant and straightforward genetics, yeast is a powerful model system for dissecting fundamental properties of eukaryotic cells and has served as a primary test bed for development of numerous functional genomic technologies (Botstein and Fink, 2011). Specifically, genome-scale mapping of genetic interactions requires three fundamental tools. First, large collections of mutant strains, carrying either gain- or loss-of-function alleles, are required for systematically perturbing gene activity. Second, high-throughput methodologies must be available for combining mutations in a rapid, accurate and comprehensive manner. Third, phenotypes of interest must be scored easily, quantitatively and in parallel for a large number of samples. The development of all three of these fundamental genomic tools was pioneered in *S. cerevisiae* and has enabled early progress in mapping quantitative genetic interactions.

1.3.1 Mutant strain libraries

The yeast deletion collection is a library of strains, in which each known or suspected open reading frame was deleted and replaced with the dominant drug-resistance marker,

kanMX (Giaever *et al.*, 2002) (Figure 1.3). This collection contains deletion strains for ~4,800 nonessential genes, available as haploids or homozygous diploids, as well as ~1,000 essential genes that are required for viability under regular growth conditions and are maintained as heterozygous diploids. Molecular tags or barcodes, i.e. strain-specific 20-bp DNA sequences, were introduced at both ends of the deletion cassette and act as unique mutant strain identifiers (Figure 1.3). The presence of molecular barcodes enables the fitness of a particular mutant to be assessed within a pooled population using a barcode microarray (Giaever *et al.*, 1999) or, more recently, high-throughput barcode sequencing (Smith *et al.*, 2009).

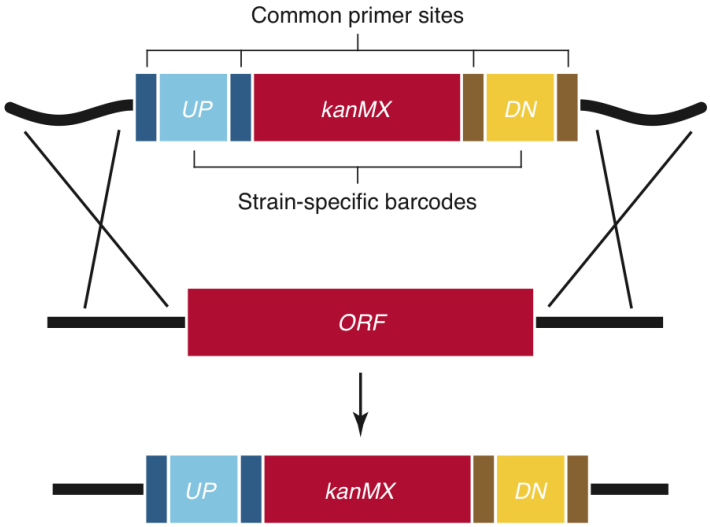


Figure 1.3. Construction of the yeast deletion-mutant collection.

Each yeast ORF is replaced with a "deletion cassette" that consists of an antibiotic-resistance marker, *kanMX* (which confers resistance to kanamycin), and two unique 20 nucleotide molecular barcodes, "uptag" (UP) and "downtag" (DN). Each barcode is flanked by common primer sites (indicated by darker blue and brown colors). Incorporation of the cassette into the yeast genome is accomplished through homologous recombination of 45-bp regions of homology upstream and downstream of the yeast ORF.

To enable the analysis of essential genes, the yeast community developed additional strain libraries that complement the non-essential deletion collection. These libraries consist of subsets of the ~1,000 essential genes that have been altered to produce either conditional alleles (Mnaimneh *et al.*, 2004; Davierwala *et al.*, 2005; Ben-Aroya *et al.*, 2008; Li *et al.*, 2011) or hypomorphic alleles that are compatible with viability (Schuldiner *et al.*, 2005; Breslow *et al.*, 2008).

Conditional expression of an essential gene can be obtained in several ways. In one of the first systematic approaches, 575 essential genes were placed under the control of a tetracycline-inducible promoter, which turns off in the presence of a tetracycline analog doxycycline (Mnaimneh *et al.*, 2004; Davierwala *et al.*, 2005). An alternative strategy is to target protein stability in response to high temperatures by introducing point mutations into the gene coding sequence such that the activity of the protein is normal at a permissive temperature, but is substantially reduced or abolished at a restrictive temperature (Ben-Aroya *et al.*, 2010). The largest collection of yeast conditional temperature-sensitive (ts) alleles in an isogenic background contains 747 mutant strains, covering ~45% of all essential genes, many of which represented by multiple alleles (Li *et al.*, 2011). This collection complements a non-overlapping array of 250 ts mutants constructed independently (Ben-Aroya *et al.*, 2008) to generate a combined dataset covering ~65% of all essential genes in yeast.

In addition to inducible promoters and point mutations, the activity of a gene can be partially reduced by destabilizing its mRNA transcript. This approach is exploited by the DAmP (decreased abundance by mRNA perturbation) collection, which contains 842 yeast

essential genes whose 3'-UTR regions were disrupted through integration of an antibiotic selectable marker (Schuldiner *et al.*, 2005; Breslow *et al.*, 2008).

1.3.2 Genetic interaction mapping technologies

The availability of large-scale mutant strain collections prompted the development of high-throughput technologies for combining mutations and studying genetic interactions. Synthetic Genetic Array (SGA) was among the first automated approaches for systematically constructing yeast double mutants from arrays of single mutants and analyzing their phenotypes (Tong *et al.*, 2001; Baryshnikova *et al.*, 2010a). In a typical SGA screen (Figure 1.4), a "query" mutant strain is crossed to the array of ~4,800 viable deletion mutants or conditional alleles of essential genes. The resulting heterozygous diploids are sequentially replica-pinned onto different selective media to induce sporulation, selection of haploid meiotic progeny and, finally, selection of haploid double mutants, which can then be scored for growth, measured as colony size, or other phenotypes of interest (Chapter 1.3.3).

The power of SGA stems from several genetic constructs, inserted into the query strain genome, which enable precise step-wise selection of double mutant cells and can be applied to any genetic element, including point mutants, gene-fusion systems and plasmids, for similar high-throughput manipulations (Baryshnikova *et al.*, 2010a). In the query strain genome, the *CAN1* locus is replaced with a *MAT***a** haploid specific reporter, *STE2pr*-Sp_*his5*, in which a *MAT***a**-specific promoter (*STE2pr*) drives the expression of the *S. pombe his5* gene, thus enabling *MAT***a** haploid cells to selectively grow on media lacking histidine. The *CAN1* locus encodes an arginine permease, which also imports canavanine, an arginine toxic ana-

logue. The *can1*Δ mutation allows survival in the presence of canavanine and thus enables counter-selection against unsporulated diploid mutants, which are heterozygous for the *can1*Δ mutation. To reinforce this selection, the query strain also carries a deletion of the *LYP1* gene, which encodes a lysine permease and allows mutant cells to grow in the presence of thialysine, a lysine toxic analog.

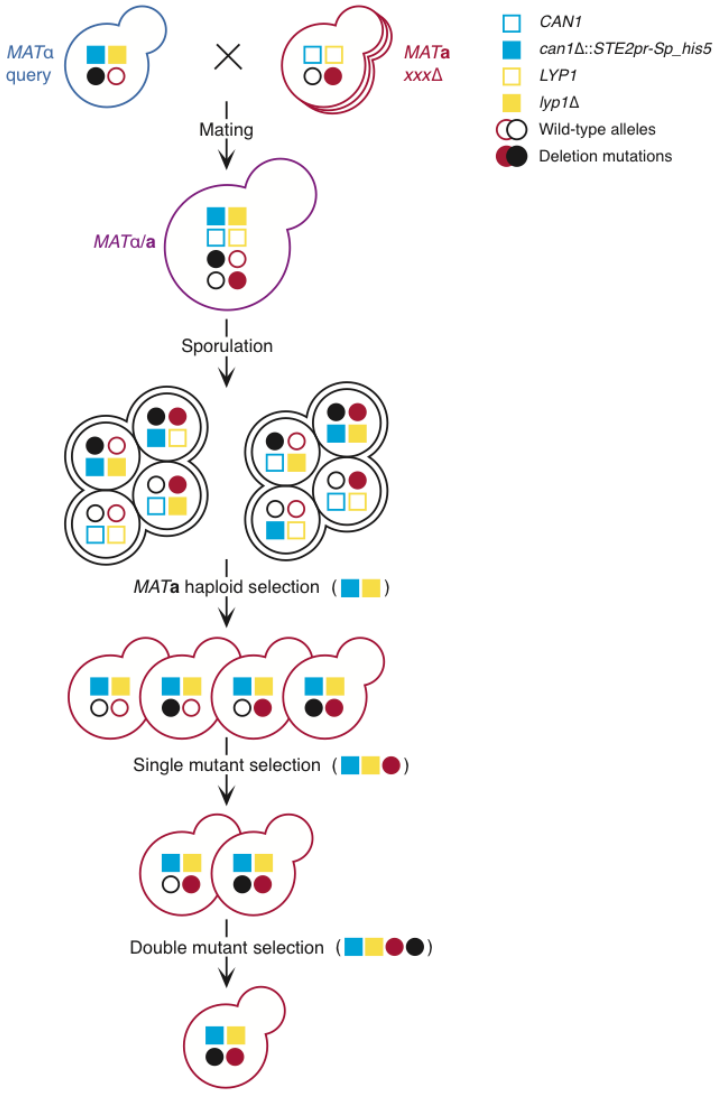


Figure 1.4. Synthetic Genetic Array (SGA) methodology.

In a typical SGA screen, a *MAT*α mutant strain carrying a "query" mutation, marked with the dominant drug-resistance marker *natMX4* (filled black circle), is crossed to an array of ~4,800 viable *MAT***a** deletion mutants or conditional alleles of essential genes. Each "array" mutation is marked with a *kanMX4* resistance cassette (filled red circle). The *CAN1* gene, which encodes an arginine permease, is replaced in the SGA query strain with a *MAT***a** haploid specific reporter, *STE2pr*-Sp_*his5* such as *can1*Δ::*STE2pr*-Sp_*his5*. The query strain also carries a deletion of the lysine permease gene, *LYP1*. Following mating, diploid selection and sporulation, meiotic progeny are grown on media containing the G418 and nourseothricin and lacking histidine to select *MAT***a** haploid double mutants. In addition to positive selection, the media is also supplemented with canavanine and thialysine (toxic analogs of arginine and lysine, respectively) in order to counter-select against unsporulated diploid mutants that are heterozygous for *CAN1* and *LYP1* deletion mutations. For a more detailed protocol and media composition, refer to Baryshnikova *et al.* (Baryshnikova *et al.*, 2010a)

In a pilot study, SGA was applied to examine synthetic lethality between eight query strains and the complete array of ~4,800 viable haploid deletion mutants (Tong *et al.*, 2001). This approach was later expanded to 132 query genes and resulted in the first large-scale genetic interaction network consisting of ~1,000 genes and ~4,000 synthetic lethal/sick interactions (Tong *et al.*, 2004). The initial analysis of this dataset provided important insights into fundamental properties of biological networks and their topology. For example, it was shown that, while generally rare, synthetic lethal interactions often connect genes with related biological functions and are extremely useful for uncovering novel functional relationships (Tong *et al.*, 2004). Similarly to other biological and technological networks, synthetic lethality follows a power-law degree distribution, with most genes having few interactions, while a few genes have many and act as network hubs (Tong *et al.*, 2004). Moreover, synthetic lethal partners of a gene tended to be synthetic lethal with each other, thus uncovering dense local neighborhoods in the genetic interaction network (Tong *et al.*, 2004).

To follow up on these early discoveries, SGA was applied to subsets of functionally related genes to uncover genetic interaction networks amongst genes known to participate in specific biological processes. In this targeted approach, select "query" strains were crossed to arrays composed of a biased set of several hundred deletion mutants involved in vesicle-mediated transport (Schuldiner *et al.*, 2005), chromosome biology (Collins *et al.*, 2007), RNA processing (Wilmes *et al.*, 2008), phosphorylation-mediated signaling (Fiedler *et al.*, 2009), transcription (Zheng *et al.*, 2010), plasma membrane related processes (Aguilar *et al.*, 2010) or mitochondrial functions (Hoppins *et al.*, 2011). Together with early genome-wide SGA experiments (Tong *et al.*, 2001; Tong *et al.*, 2004), these studies uncovered genetic in-

teraction networks underlying small subsets of gene-gene combinations. However, unbiased examination of genetic interactions on a global scale had not been performed until recently and is the focus of Chapters 2 and 3 of my thesis.

Methods complementary to SGA have also been developed. For example, dSLAM (diploid Synthetic Lethal Analysis by Microarray) takes advantage of the unique molecular barcodes associated with each deletion mutant to map synthetic lethal interactions by measuring the relative abundance of double mutants in a mixed population (Pan *et al.*, 2004) (Figure 1.5). Briefly, a pooled set of heterozygous deletion strains, each containing an SGA marker, is mass transformed with a marked query mutation. Using the same selection steps as in SGA, double mutant haploids are selected and their relative abundance is quantified by measuring barcode intensities compared to a non-selected control pool. dSLAM has been used to map synthetic genetic interactions between genes involved in DNA integrity and histone modification (Pan *et al.*, 2006; Lin *et al.*, 2009).

An SGA/dSLAM hybrid approach, called Genetic Interaction Mapping (GIM), was used to examine interactions between genes involved in mRNA processing (Decourty *et al.*, 2008). In a typical GIM experiment, double mutants are generated by mating and sporulation in a manner analogous to SGA (Figure 1.5). However, similar to dSLAM, all steps are performed in a pooled format, which involves competitive growth of double mutant meiotic progeny, and interactions are identified by comparing barcode microarray hybridization intensities between double mutants and a reference population (Decourty *et al.*, 2008).

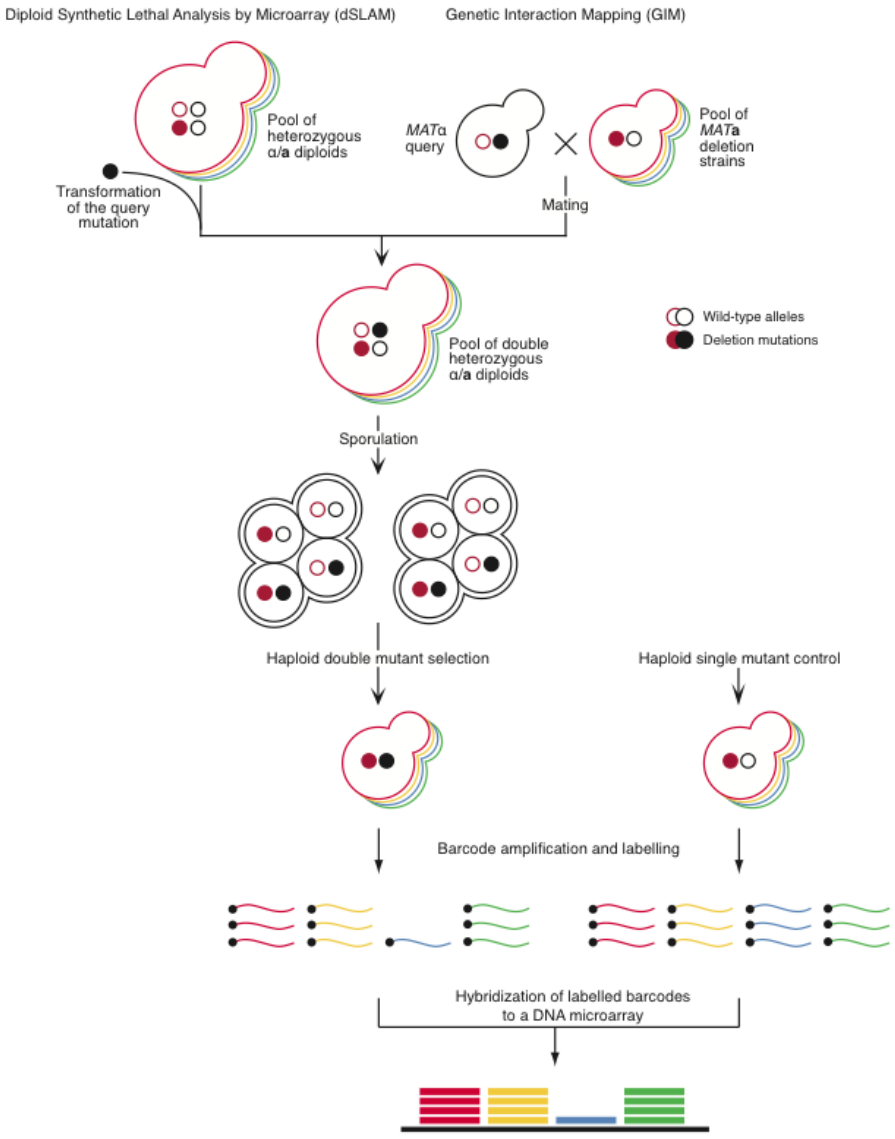


Figure 1.5. Diploid Synthetic Lethal Analysis by Microarray (dSLAM) and Genetic Interaction Mapping (GIM) methodologies.

The dSLAM and GIM genetic interaction mapping methods differ in the approach for constructing a pool of double heterozygous diploid mutants, but are similar in the subsequent selection steps and phenotypic readout. In dSLAM, a query mutation (filled black circle), linked to the *URA3* selectable marker, is introduced into the pool of haploid-convertible heterozygous diploid strains by high-efficiency integrative transformation. In GIM, a *MAT*α haploid query strain, in which a specific genomic locus is replaced with a nourseothricin resistance marker, is mated with a pool of viable *MAT*a deletion mutants. Haploid single-mutant (control) or double-mutant (experimental) pools are selected after sporulation, through germination of spores on a medium that lacks histidine and selection for the relevant alleles. Genomic DNA samples are isolated from both pools and used as templates for PCR amplification of the tags, during which they are labeled with fluorescent dyes. Microarray analysis of these dye-labeled tags reveals the synthetic interaction between each of the corresponding deletion alleles with the query mutation.

Global estimates, based on the relatively small data sample available at the time, indicated that the *S. cerevisiae* genome could harbor as many as ~200,000 synthetic sick and lethal interactions, involving both essential and non-essential genes (Tong *et al.*, 2004; Davierwala *et al.*, 2005). In light of the most recent findings, including our own work described in Chapter 3, this number may in reality be an underestimate of the extensive genetic complexity encoded within a single unicellular organism. As we move towards completion of the yeast pairwise genetic interaction network, with the spectrum and the resolution of our phenotypic assays progressively increasing, it appears that the network might be much denser than anticipated and comprise an order of magnitude more genetic interactions (Chapters 1.3.3, 2 and 3).

1.3.3 Phenotypes

All three major technologies for mapping genetic interactions in yeast (SGA, dSLAM and GIM) use cellular fitness as their primary phenotype. Fitness, intended broadly as growth in a given laboratory environment, is an excellent phenotypic readout as it can be measured easily and quantitatively, and it is well amenable for high-throughput applications. Different technologies measure fitness in different ways. In SGA, single and double mutant fitness estimates are obtained from measurements of colony sizes: plates corresponding to the last SGA selection step are digitally photographed and analyzed by an image processing algorithm that identifies the colonies and measures their areas in units of pixels (Chapter 2) (Baryshnikova *et al.*, 2010a). dSLAM and GIM estimate fitness by measuring the hybridization intensities of mutant-associated barcodes, which are indicative of the relative abundance of double mutants in a population (Pan *et al.*, 2006; Decourty *et al.*, 2008; Lin *et al.*, 2009).

On a smaller scale, fitness has been measured by monitoring the optical density of a growing yeast culture over time and calculating its exponential growth rate, duration of lag phase and saturation level (St Onge *et al.*, 2007).

Despite all the available measures of fitness, it is becoming increasingly evident that exploring more subtle and specific phenotypes is necessary for obtaining a more complete picture of a cell's functional organization. For example, nearly 50% of non-essential yeast deletion mutants, while having no fitness defect, exhibit a number of morphological problems (Ohya *et al.*, 2005), suggesting that complex phenotypes may uncover functional connections between genes that are not readily apparent from growth data alone.

Small-scale studies have explored a variety of high-resolution phenotypes for measuring genetic interactions. For example, a quantitative agar-invasion assay was developed to monitor yeast filamentous growth in a set of single and double deletion mutants (Drees *et al.*, 2005). In response to low ammonium availability, *S. cerevisiae* differentiates into a filamentous form that can invade agar and serve as a model for understanding the infective mechanisms of pathogenic yeasts, such as *Candida albicans* (Jin *et al.*, 2008). Genetic interactions derived from the agar-invasion phenotype demonstrated a complex relationship between different signaling pathways and regulatory components that impinge on filamentous growth and are likely to affect pathogenicity (Drees *et al.*, 2005).

A pioneering study in *Dictyostelium discoideum* used global gene expression profiles to map genetic interactions among members of the signal transduction pathways that govern the *Dictyostelium* multicellular developmental program (Van Driessche *et al.*, 2005). In another

study, arrays of single and double mutants were used to investigate the mechanisms of receptor endocytosis (Burston *et al.*, 2009). Efficiency of protein internalization in single and double mutants was quantified using an enzymatic assay based on the cell surface localization of Snc1, the yeast vesicle-associated membrane protein (VAMP)/synaptobrevin homologue. Quantitative negative and positive genetic interactions derived from Snc1 localization signals were used to cluster genes into functional modules identifying 20 novel members of Snc1 uptake pathways (Burston *et al.*, 2009).

Using SGA, reporter constructs can be introduced into sets of deletion mutants to monitor activation or repression of specific pathways. For example, the green fluorescent protein (GFP) reporter was introduced into the complete set of viable single deletion mutants and a select set of double mutants to detect the activation of the endoplasmic reticulum unfolded protein response (UPR) (Jonikas *et al.*, 2009). Using a similar strategy, fluorescent reporters enabled the examination of cell cycle-dependent transcription on a genome-wide scale (Costanzo *et al.*, 2004; Fillingham *et al.*, 2009).

The spectrum of phenotypic traits amenable for genome-scale mapping of genetic interactions has expanded immeasurably with technological advances in high-throughput microscopy and imaging tools (Vizeacoumar *et al.*, 2010). By combining cytological reporters with high-content screening, it is now possible to classify and measure a variety of yeast morphological and protein localization phenotypes in a large-scale manner (Huh *et al.*, 2003; Ohya *et al.*, 2005). For example, morphological data pertaining to the cell wall, actin cytoskeleton and nuclear DNA have been systematically collected and analyzed for the entire set of *S. cerevisiae* non-essential gene deletion mutants (Ohya *et al.*, 2005). Similar approaches

have been applied in mammalian cells to characterize genes regulating various cellular processes, including cell morphology and cell cycle progression (Moffat *et al.*, 2006).

Combining high content screening with SGA enabled genome-wide mapping of genetic interactions associated with mitotic spindle morphogenesis (Vizeacoumar *et al.*, 2010). A GFP-tubulin reporter was introduced into single and double mutants and, aided by automated image processing and computational scoring of phenotypes, enabled quantitative measurement of 90 cell morphological parameters. This approach identified 122 novel genetic interactions that impinge on spindle morphology and, by doing so, produced a genetic interaction network four times larger than the one derived for the fitness-based analysis of the microtubule-binding protein Bim1. Thus, integrating diverse phenotypic assays with high-throughput technologies for combining mutations, such as SGA, has the potential to uncover vast genetic interaction networks that shed light onto the organization of complex biological processes (Michaut and Bader, 2012).

1.4 Mapping genetic interactions in other model organisms

Besides *S. cerevisiae*, approaches for large-scale mapping of genetic interaction networks have been developed in other eukaryotic and prokaryotic organisms, thus enabling cross-species comparison and evolutionary analysis of genetic interaction networks (Butland *et al.*, 2008; Dixon *et al.*, 2008; Roguev *et al.*, 2008; Typas *et al.*, 2008). These techniques are directly analogous to those employed in *S. cerevisiae* in that they use genome-wide deletion collections to generate comprehensive sets of double mutants via mating procedures.

Large-scale mapping of genetic interactions in fission yeast *Schizosaccharomyces pombe* revealed numerous similarities as well as differences in the wiring of its genetic network compared to *S. cerevisiae* (Dixon *et al.*, 2008; Roguev *et al.*, 2008). Despite hundreds of millions of years of evolutionary separation, *S. cerevisiae* and *S. pombe* seem to share ~30% of their genetic interactions, while ~70% are species-specific (Dixon *et al.*, 2008; Roguev *et al.*, 2008). At least some of the differences can be explained by the unique gene functions, such as the RNA interference machinery, encoded in the fission but not budding yeast genome (Sunnerhagen, 2002). Another potential explanation is offered by a recent cross-species comparison of chemical-genetic interactions, which define deletion mutants that are sensitive to chemical compounds (Kapitzky *et al.*, 2010). This analysis showed that, while individual chemical-genetic interactions are poorly conserved, interactions involving specific functional modules are maintained at a much higher level (Kapitzky *et al.*, 2010), thus supporting the hypothesis that general network properties, rather than individual pairwise connections, are preserved through evolution. Because such network properties are likely shared by other metazoans, including humans, the comparative analysis of genetic interactions in *S. pombe* and *S. cerevisiae* will provide important insights into the functional organization of higher organisms.

Similar experimental approaches have been devised to uncover genetic interactions in the bacterium *Escherichia coli* (Butland *et al.*, 2008; Typas *et al.*, 2008). Despite the paucity of data available so far and the lack of systematic comparisons to the *S. cerevisiae* network, preliminary results suggest that prokaryotic and eukaryotic genomes, while having a two-fold difference in gene density, might bear a similar degree of redundancy. For example, system-

atic interrogation of 39 nonessential genes in *E. coli* showed that each gene has ~20 synthetic lethal partners on average (Butland *et al.*, 2008), which is comparable to the estimate of ~30 synthetic lethal interactions per nonessential gene in yeast (Tong *et al.*, 2004). Furthermore, profiles of synthetic lethal interactions in *E. coli* are highly informative about a gene's biological function and, just like in *S. cerevisiae*, enable grouping of genes into functional modules (Butland *et al.*, 2008).

In addition to unicellular organisms, large-scale genetic interaction studies have also been described for the nematode worm *Caenorhabditis elegans* (Lehner *et al.*, 2006; Byrne *et al.*, 2007; Tischler *et al.*, 2008) and the fruit fly *Drosophila melanogaster* (Horn *et al.*, 2011). These studies are made possible by the availability of organism-specific genome-wide RNAi libraries, which can mimic the effect of gene deletions by reducing the abundance of transcripts and are compatible with high-throughput combinatorial approaches.

C. elegans is particularly suitable for RNAi-mediated genetic interaction studies because a gene's expression can be inhibited in the entire organism simply by feeding the worms on a bacterial lawn expressing the dsRNA of interest. The first systematic study of genetic interactions in worms screened 37 query genes against ~1,750 individual RNAis and identified 350 synthetic lethal interactions, many of which involved human disease orthologs (Lehner *et al.*, 2006). Interestingly, genes encoding chromatin regulators appeared to be central to the worm genetic interaction network and acted as network hubs. We later showed that the high connectivity of chromatin-related genes is shared by the *S. cerevisiae* network (Chapter 3.7), suggesting a general conservation rule for genetic interaction hubs across organisms. Despite this finding, the overall degree to which individual genetic interactions are

conserved between yeast and worms is still unclear (Byrne *et al*., 2007; Tarailo *et al*., 2007; Tischler *et al*., 2008). Recent evidence suggests that conservation of genetic and other biological interactions is maintained at the functional module level, rather than individual gene level, and may be reflecting a common network evolution mechanism in various species (Zinman *et al*., 2011).

In *Drosophila*, inhibiting a gene's function in an entire organism is much more challenging, and large-scale studies of genetic interactions have instead focused on cell cultures. In a recent report, a combinatorial RNAi strategy was used to systematically map pairwise genetic interactions among 93 genes encoding signaling factors (Horn *et al.*, 2011). Negative and positive genetic interactions were identified using several phenotypic readouts, including cell number, nuclear area and fluorescence intensity of stained nuclei (Horn *et al.*, 2011). Interestingly, different phenotypes seemed to uncover different functional relationships between genes, as only ~20% of the identified genetic interactions were common to all phenotypes under consideration. Consistent with yeast and bacterial studies, genes with similar functions shared similar genetic interaction profiles, independent of the phenotypes from which the profiles were generated. This indicates that genetic interactions provide a powerful tool for annotating uncharacterized *Drosophila* genes. Indeed, comparison of genetic interaction profiles led to the discovery of a novel activator of RAS-MAPK pathway signaling, whose function is conserved from flies to humans (Horn *et al.*, 2011).

1.5 Mapping genetic interactions in mammalian cell lines

Currently, large-scale mammalian genetic interaction studies are based on RNAi-mediated genetic manipulation of cell cultures. Comparative analyses of different cell lines have the potential to uncover cell type-specific molecular vulnerabilities and provide directions for targeted therapeutic approaches. Most commonly used cancer therapies involve administering high doses of radiation or toxic chemicals to the patient, which suppress tumor growth but also cause substantial damage to normal cells. Compared to chemo- and radiotherapy, synthetic lethality emerges as a much more accurate approach, which can selectively kill cancer cells carrying a specific mutation by chemically targeting a synthetic lethal partner of the mutated gene. The enormous potential of this strategy was first noted by Hartwell and colleagues (Hartwell *et al.*, 1997) and has recently been realized by the identification of PARP inhibitors as novel treatments for breast cancer (O'Shaughnessy *et al.*, 2011). Poly(ADP-ribose) polymerase (PARP) is an enzyme involved in base excision repair, a DNA repair pathway that is not essential in normal cells with functional homologous recombination (Farmer *et al.*, 2005). In cells, however, where homologous recombination is impaired, such as BRCA-dependent cancer cells, PARP activity is required for survival, thus suggesting that drug-mediated PARP inactivation will likely result in selective arrest of cancer cell growth (Bryant *et al.*, 2005; Farmer *et al.*, 2005). Currently, several PARP inhibitors are in various stages of clinical trials and promise successful treatment of breast cancer in the nearest future (O'Shaughnessy *et al.*, 2011).

To extend this approach to other types of cancer, several groups have examined gene essentiality in human cancer cell lines using large-scale RNAi libraries (Luo *et al.*, 2008;

Silva *et al.*, 2008; Cheung *et al.*, 2011; Marcotte *et al.*, 2012). One of these studies tested over 11,000 genes in 102 cell lineages and uncovered 54 genes that are specifically essential for viability of ovarian cancer cells (Cheung *et al.*, 2011). Among these genes, *PAX8* was confirmed as a lineage-specific survival gene that is essential for proliferation, highly expressed and substantially amplified in ovarian cancers (Cheung *et al.*, 2011).

Lineage-specific essentiality presumably reflects synthetic lethal interactions between the knocked-down gene and the genetic background of the cancer cell. While in most cases the identity of the relevant endogenous mutations remains to be determined, a few studies were able to pinpoint the true effectors of the synthetic lethal phenotype. For example, tumor cells harboring oncogenic alleles of the *NRAS* and *HRAS* genes were used as queries in a screen of 5,760 short hairpin RNAs targeting ~1,000 human genes, including 571 kinases (Yang and Stockwell, 2008). This screen identified a synthetic lethal interaction between oncogenic Ras and the gene *CSNK1E*, encoding a casein kinase 1 epsilon kinase (Yang and Stockwell, 2008). It was subsequently shown that chemical inhibition of casein kinase 1 epsilon recapitulated the synthetic lethal effect observed with shRNA-mediated knockdown, validating this gene as a potential anticancer drug target.

The availability of cell lines carrying genetic mutations in loci of interest limits the number of human genes that can be tested for genetic interactions. This limitation is partially overcome by the development of combinatorial RNAi strategies, which enable simultaneous shRNA-mediated knock down of two or more genes in animal cells (Lambeth *et al.*, 2010). One of the latest combinatorial RNAi studies identified 878 negative and positive genetic interactions for 12 human lysine deacetylases (KDACs) using a pooled human short hairpin

RNA (shRNA) library (Lin *et al.*, 2012). The obtained genetic interaction profiles uncovered substrate specificity for individual KDACs and enabled characterization of the AMPK pathway in lipid metabolism.

1.6 Mapping genetic interactions in human populations

The ultimate goal of studying genetic interactions is to understand the genetic basis of human phenotypic variation, identify genetic variants responsible for disease and develop successful therapeutic strategies. This task is extremely challenging because, unlike model organisms and cultured cell lines, standard genetic methods are not applicable in human populations, and thus systematic genotype-to-phenotype mapping requires more sophisticated approaches.

Uncovering the genetic causes of human phenotypic variation is a massive epidemiological and statistical endeavor, involving design of sampling strategies, collection of health data and development of new statistical analysis tools. The field has been dominated by the success of genome-wide association studies (GWAS) in which genetic variants, identified by SNP mapping and genome-wide sequencing efforts, are tested for statistical association with a particular phenotype. GWAS experiments have quickly and dramatically increased our knowledge of human genetics, linking over 1,200 genetic variants to nearly 170 complex traits (Lander, 2011). However, only a few of these associations explain a substantial portion of trait heritability. For example, 71 genetic variants have been associated with Crohn's disease, a chronic inflammatory disorder of the gastrointestinal tract affecting ~0.1% of North American population (Zuk *et al.*, 2012). However, only ~22% of Crohn's disease heritability

is explained by the cumulative contribution of these variants, indicating that our understanding of the genetic scaffold of this disorder is far from complete (Zuk *et al.*, 2012).

A number of reasons have been proposed to explain missing heritability, including environmental contribution, influence of particularly rare and still undiscovered genetic variants or failure to detect small genetic effects due to the limited sample size of a typical GWAS study (Gibson, 2011). The penetrance of certain phenotypes may also be affected by stochastic changes in gene expression, whereby the impact of a mutation is masked by a higher abundance of an ancestral duplicate or a promiscuous buffer gene (Burga *et al.*, 2011).

Another potential cause for missing heritability is the presence of genetic interactions among the genetic variants, which, when not accounted for, lead to an underestimate of the portion of heritability explained by the variants (Zuk *et al.*, 2012). Indeed, a model that incorporates the possibility for genetic interactions between the 71 genetic variants associated with Crohn's disease showed that these variants may explain as much as ~84% of the disease heritability, compared to ~22% when only additivity between variants is considered (Zuk *et al.*, 2012). This finding suggests that we might have already uncovered most of the variants necessary to explain heritable phenotypic variation, but more work is required to identify the mathematical models that govern their interactions.

Evidence for the ubiquitous presence of genetic interactions in natural populations have been accumulating over the past several years. In yeast, more than 3% of all gene pairs tested so far exhibit a detectable positive or negative genetic interaction for fitness (Chapters 2 and 3), and this rate is substantially higher for more specific phenotypes (Chapter 1.3.3). Moreo-

ver, a comparative analysis of two closely related *S. cerevisiae* strains (S288C and Σ1278b) showed that conditional essentiality, whereby a gene is essential in one strain's genetic background but not in the other, is almost always driven by two or more genetic modifiers (Dowell *et al.*, 2010). Being two variants of the same species, S288C and Σ1278b, just like two human individuals, are >99% identical (Schacherer *et al.*, 2007), suggesting that conditional essentiality and other strain-specific phenotypes are due to genetic interactions involving their individual sequence variants (Dowell *et al.*, 2010).

In conclusion, given the importance of genetic interactions in determining the phenotype of an individual, understanding their behavior is a fundamental goal in post-genomic biology. Although this task faces many experimental and statistical challenges in multicellular organisms, particularly in humans, it is theoretically and practically feasible in yeast, which is indeed well underway to be the first organism with a complete pair-wise genetic interaction network in at least one experimental condition. In this thesis, I describe the generation and the analysis of the first 30% of the yeast genetic interaction network derived from SGA experiments conducted in our laboratory. This work provides a solid platform for similar experimental endeavors in other model systems and contributes to the elucidation of complex trait heritability in humans.

1.7 Thesis rationale

Our laboratory has undertaken the effort of mapping the complete pair-wise genetic interaction network for budding yeast *S. cerevisiae*. This ambitious task is feasible through systematic Synthetic Genetic Array (SGA) experiments, which combine arrays of single gene

mutations to generate all possible double mutants and examine their phenotypes (Chapter 1.3.2). The goal of my thesis is to extract quantitative genetic interactions from single and double mutant colony size data derived from systematic SGA experiments and analyze the properties of the resulting genetic interaction network.

Mapping genetic interactions between all genes of an organism bears extraordinary importance for understanding the functional organization of the cell: genetic interactions uncover complex functional interconnections between genes, pathways and biological processes, and ultimately underlie the relationship between genotype and phenotype of an organism. Since so far *S. cerevisiae* is the only living organism for which an extensive genetic interaction network is available, methods developed in this thesis and findings obtained through their application provide a foundation for similar analysis in other organisms, including mammalian cells and human populations.

Similarly to other genomics technologies, SGA suffers from several systematic experimental biases, which interfere with our ability to obtain accurate measures of fitness from colony sizes and introduce a substantial number of false positive and false negative interactions. The first aim of my studies, which I describe in Chapter 2, was to identify these systematic effects and develop computational methods for their removal. In doing so, I was able to extract accurate fitness information from SGA-derived single and double mutant colony sizes, which rivals the quality of smaller-scale higher-resolution published datasets. Furthermore, I generated a comprehensive catalog of yeast single mutant fitness, which serves as a reference for numerous genomic analyses (see, for example, Chapter 3.6). Finally, quantitative fitness-based genetic interaction measurements were derived using my method, provid-

ing significant improvements in data quality and gene function prediction capacity over previously available methods for measuring genetic interactions. This method, collectively referred to as the SGA score, is applicable to similar experimental contexts and was published in Nature Methods in 2010 (Baryshnikova *et al.*, 2010b).

In Chapter 3, I present the results of applying the SGA score to ~1,700 genome-wide SGA screens. Analysis of this dataset revealed quantitative genetic interactions among 5.4 million yeast double mutants (~30% of the yeast genome) and generated the first genome-wide genetic interaction map of a living organism. I developed a novel network-based method for visualizing large-scale genetic interaction data that led to an unprecedented view of the cell revealing the functional organization of a cell. Genes of similar biological processes cluster together in coherent subsets and functionally interconnected bioprocesses map next to each other. We discovered several physiological and evolutionary gene features that are characteristic of genetic interaction hubs, and explored the relationship between genetic and protein-protein interaction networks. By comparing quantitative single and double mutant phenotypes, we identified specific cases of positive genetic interactions, termed genetic suppression, and constructed a global network of suppression interactions among protein complexes. I also demonstrated that an extensive and unbiased mapping of genetic interactions provides a key for interpretation of chemical-genetic interactions and drug target identification. The genetic interaction map and the associated novel biological findings were published in Science in 2010 (Costanzo *et al.*, 2010).

Finally, in Chapter 4, I show that, in addition to genetic interactions, systematic SGA experiments produce a genome-wide genetic linkage map for *S. cerevisiae*. Since SGA gen-

erates double mutants by mating and sporulation, the frequency of double mutant meiotic progeny involving loci on the same chromosome reflects the frequency of recombination between the loci and can be used to identify recombination hotspots and coldspots across the genome. Using SGA-derived genetic linkage data, I mapped domains of high and low meiotic recombination, which recapitulated previously identified recombination patterns, including inhibition of recombination at pericentromeric chromosomal regions. Interestingly, I found that the extent of centromere-related recombination repression is directly proportional to chromosome length, whereas low recombination regions outside the centromere were, on average, constant in size. These findings corroborate the unique role played by centromeres during meiotic division and suggest a chromosome size-dependent mechanism for ensuring proper chromosome segregation.

2 Measuring quantitative fitness and genetic interactions in yeast on a genome scale

Portions of this chapter have been reprinted or adapted from:

- Baryshnikova A., Costanzo M., Dixon S., Vizeacoumar F. J., Myers C. L., Andrews B., Boone C., "Synthetic Genetic Array (SGA) Analysis in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*", Methods in Enzymology, 470:145–179, 2010 (Baryshnikova *et al.*, 2010a)
- Baryshnikova A.*, Costanzo M.*, Kim Y., Ding H., Koh J., Toufighi K., Youn J. Y., Ou J., San Luis B. J., Bandyopadhyay S., Hibbs M., Hess D., Gingras A. C., Bader G. D., Troyanskaya O. G., Brown G. W., Andrews B., Boone C., Myers C. L., "Quantitative analysis of fitness and genetic interactions in yeast on a genome scale", Nature Methods, 7(12):1017–1024, 2010 (Baryshnikova *et al.*, 2010b)
- Costanzo M.*, Baryshnikova A.*, Bellay J., Kim Y., Spear E. D., Sevier C. S., Ding H., Koh J. L., Toufighi K., Mostafavi S., Prinz J., St Onge R. P., VanderSluis B., Makhnevych T., Vizeacoumar F. J., Alizadeh S., Bahr S., Brost R. L., Chen Y., Cokol M., Deshpande R., Li Z., Lin Z. Y., Liang W., Marback M., Paw J., San Luis B. J., Shuteriqi E., Tong A. H., van Dyk N., Wallace I. M., Whitney J. A., Weirauch M. T., Zhong G., Zhu H., Houry W. A., Brudno M., Ragibizadeh S., Papp B., Pal C., Roth F. P., Giaever G., Nislow C., Troyanskaya O. G., Bussey H., Bader G. D., Gingras A. C., Morris Q. D., Kim P. M., Kaiser C. A., Myers C. L., Andrews B. J., Boone C., "The genetic landscape of a cell", Science, 327(5964):425-431, 2010 (Costanzo *et al.*, 2010)

* These authors contributed equally to the publication.

I designed and implemented the SGA score algorithm in close collaboration with Chad Myers (University of Minnesota): we jointly defined the problems and the analytical strategies for their resolution, and equally contributed to coding. In addition, I performed all technical and functional evaluations of the data, and developed the strategy for the assessment of false positive/false negative rates. I generated the single mutant fitness catalog and conducted comparisons to published datasets.

Sara Mostafavi developed and applied the GeneMANIA algorithm for evaluating the function prediction capacity of large-scale datasets.

Michael Costanzo designed and coordinated all experimental work which was performed by technicians in the Boone lab and collaborators.

2.1 Abstract

Global quantitative analysis of genetic interactions is a powerful approach for deciphering the roles of genes and mapping functional relationships among pathways. Using colony size as a proxy for fitness, I developed a method for measuring genetic interactions from high-density arrays of yeast double mutants generated by synthetic genetic array (SGA) experiments. I identified several experimental sources of systematic variation and developed normalization strategies to obtain accurate single- and double-mutant fitness measurements, which rival the accuracy of higher resolution but smaller scale studies. I showed that quantitative genetic interactions obtained from these fitness measurements are reproducible and biologically relevant, and contribute a substantial amount of unique functional information not captured by previous genetic interaction studies or other genomic datasets.

2.2 Introduction

Synthetic genetic array (SGA) analysis is an automated form of yeast genetics that uses robotic manipulations to combine arrays of mutant strains and construct double mutants in a high-throughput manner (Baryshnikova *et al.*, 2010a) (Chapter 1.3.2). SGA was developed to enable large-scale mapping of genetic interactions, which are identified as unexpected phenotypes arising from the combination of two or more mutations and can be broadly divided into "positive" and "negative" subcategories (Chapter 1.2). Negative genetic interactions refer to double mutants that exhibit a more severe fitness defect than expected, with an extreme case being synthetic lethality. Conversely, positive genetic interactions refer to double mutants with a less severe fitness defect than expected, and involve, for example, genes encoding members of the same nonessential protein complex. The expected fitness of a double mutant is typically modeled as a multiplicative combination of the single mutants phenotypes, and genetic interactions are quantitatively measured as the difference between observed and expected double mutant fitness (Chapter 1.1).

Previous studies have quantified SGA-derived genetic interactions using an interaction score, termed S-score (Collins *et al.*, 2006), which relies on colony size measurements obtained from digital images and estimates genetic interactions by comparing each double mutant screen to a set of control screens. The S-score outputs a composite value reflecting both the magnitude and reproducibility of genetic interactions, but it does not explicitly measure single or double mutant fitness, which are critical for detailed interpretation of genetic interactions (Collins *et al.*, 2006; Breslow *et al.*, 2008). In addition, quantification of genetic interactions using the S-score is not as precise as that achieved by higher-accuracy, albeit low-

er-throughput, methods (Breslow *et al.*, 2008), especially in the context of genome-wide SGA screens (Chapter 2.6).

To enable mapping accurate quantitative genetic interactions on a genome-wide scale, I developed a new method, called the SGA score. I identified several systematic effects associated with genome-scale SGA experiments and developed normalization methods for removing them. This led to single and double mutant fitness measurements of comparable precision to those obtained using other high-resolution methodologies. I applied the SGA score to a genome-scale collection of SGA screens to quantify genetic interactions among ~5.4 million gene pairs, and showed that normalization of systematic effects has a striking impact on their function prediction capacity. Moreover, genetic interactions obtained using our method contribute a substantial amount of unique functional information not captured by previous genetic interaction studies or other genomic datasets.

2.3 A quantitative fitness-based model for genetic interactions

The requirement for accurate phenotypic measurements has imposed constraints on the scale and functional scope of quantitative genetic interaction studies (Drees *et al.*, 2005; St Onge *et al.*, 2007; Breslow *et al.*, 2008). Genome-scale surveys require experimental designs that optimize throughput and, as a result, often sacrifice accuracy. I developed a strategy for deriving quantitative genetic interactions from arrays of double mutant yeast colonies generated by genome-scale SGA screens, such that all possible *Saccharomyces cerevisiae* digenic interactions could be examined in an accurate and unbiased manner.

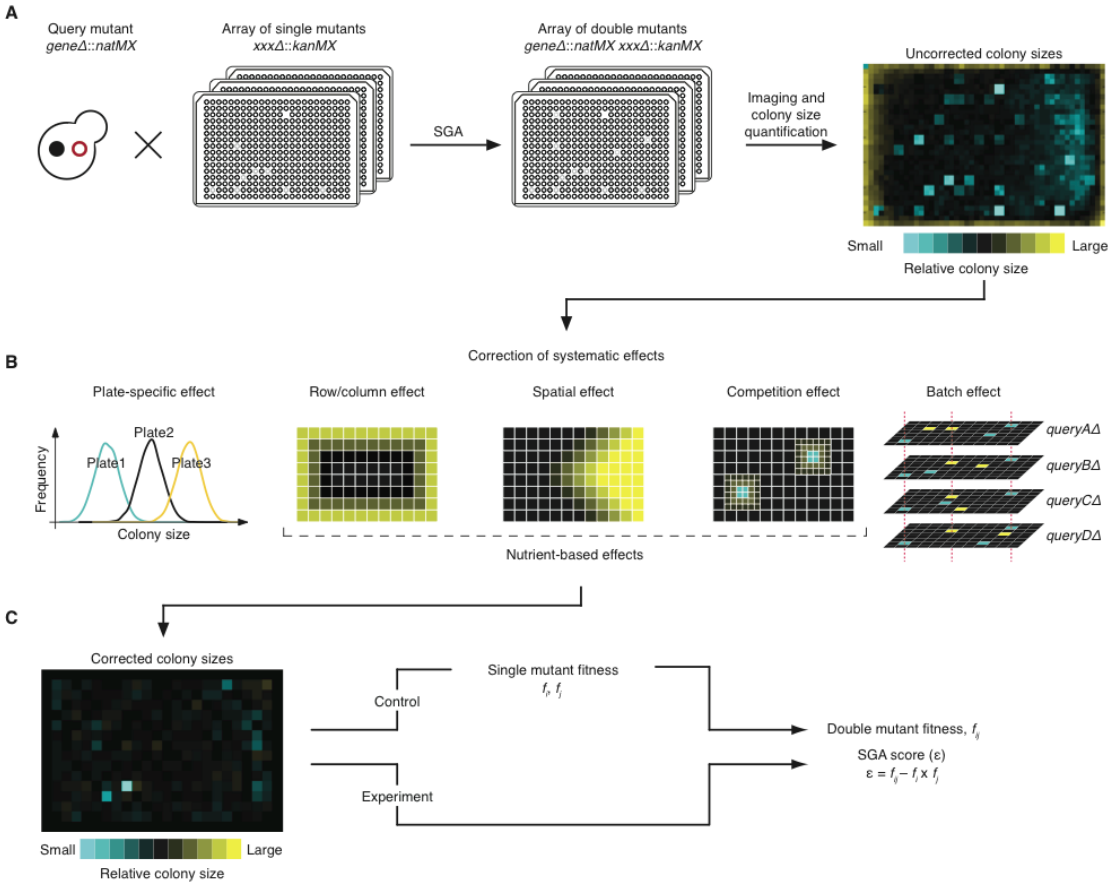


Figure 2.1. The SGA score for measuring quantitative genetic interactions.

(**A**) In a typical SGA experiment, a "query" single mutant strain (*geneΔ::natMX*) is crossed to an input array of single mutants (*xxxΔ::kanMX*), each carrying a wild-type copy of the "query" gene and a unique "array" strain mutation. After several SGA selection steps, a final output array of double mutants (*geneΔ::natMX xxxΔ::kanMX*) is generated, photographed and quantified using an image-processing software that measures colony areas in terms of pixels. Relative colony size, determined by measuring deviation of individual colonies from the median size for the same colony across 1,712 different experiments, is shown.
(**B**) Schematic depiction of the five factors that contribute to experimental variance of colony size.
(**C**) Relative colony size after normalization. Single mutant fitness (f_i, f_j) and double mutant-fitness (f_{ij}) derived from normalized colony size measurements are used to identify and measure genetic interactions (SGA score, ε).

Similar to other high-throughput technologies, genetic interaction screens using ordered mutant arrays are susceptible to systematic experimental effects that introduce noise in colony size measurements (Figure 2.1, Figure 2.2). Contributing factors include differences in growth conditions from one array plate to the next, as well as subtle differences in local nutrient availability within the same plate. For example, more nutrients are available to colonies located in less densely populated areas of the plate. As a result, colonies located in the outermost rows and columns of a plate are on average 40% larger than centrally located colonies (Figure 2.2B). Similarly, strains located next to less fit neighbors, including those showing negative genetic interactions, are also significantly larger and may be mistakenly associated with positive genetic interactions (Figure 2.2D).

The most problematic source of systematic variability is due to the so-called "batch effect", where a batch is defined as a set of screens conducted on the same day by the same person using the same robotic instrument (Figure 2.1, Figure 2.2E). Because screens conducted in a batch are influenced by common experimental factors, they often exhibit similar trends in colony size variation irrespective of the query strain identity. Such colony size variation is comparable in magnitude to the genetic interactions we aim to measure, substantially increasing our false positive rate. Indeed, without accounting for the batch effect, genetic interaction profiles of functionally unrelated genes appear to be as similar as profiles of genes encoding members of the same protein complex (Figure 2.2E). Given the importance of correlation-based approaches for genetic interaction analyses (Chapter 1.2; Chapter 3.3), strategies for normalizing batch signatures are critical for harnessing the full potential of large-scale genetic interaction data.

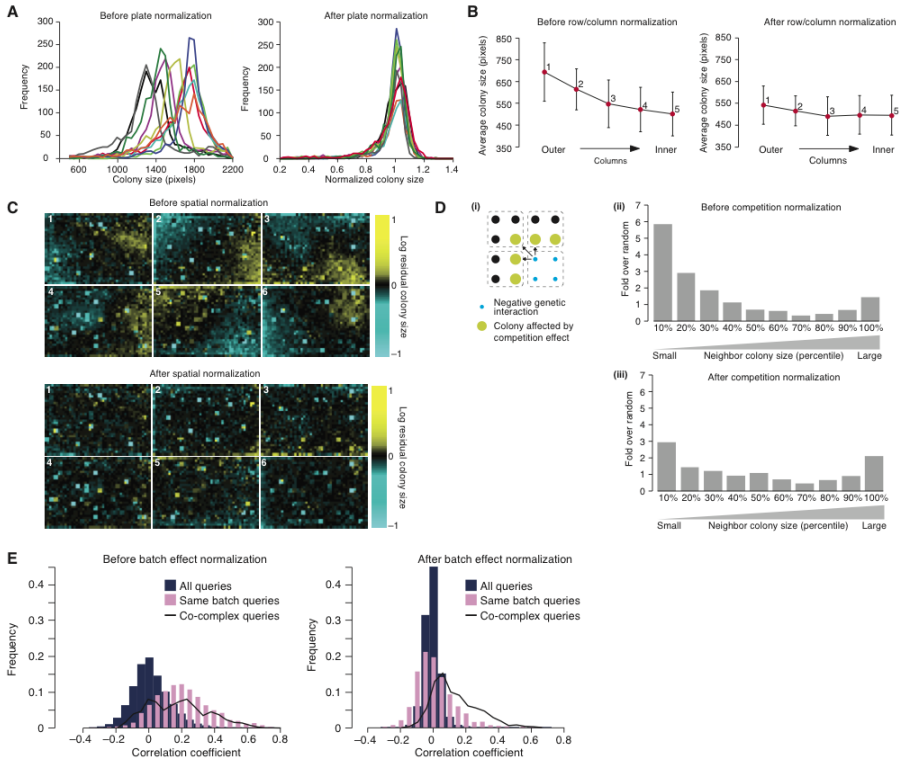


Figure 2.2. Correction of systematic experimental effects.

(**A**) Plate normalization. Colony size distributions for 10 plates randomly selected from control SGA screens are shown before and after plate normalization. Each color corresponds to a different SGA plate.
(**B**) Row/column normalization. Colony sizes located in columns 1 to 5 were averaged across 80 SGA control screens. The mean and standard deviation of colony sizes in each column are shown before and after row/column normalization.
(**C**) Spatial effect normalization. Residual colony sizes (Eq. 2.6) for 6 typical SGA plates are shown before and after spatial effect normalization. Colonies are displayed according to plate position and highlighted in yellow, blue or black depending on whether they appear larger, smaller or equal to the average colony size in that position.
(**D**) Competition normalization. (**i**) Schematic of colonies neighboring a synthetic sick/lethal interaction (blue). Replicate colonies corresponding to the same double mutant are enclosed by a dotted line. Arrows point to the colonies (yellow) that are most affected by the sick neighbor. (**ii-iii**) Colonies neighboring the top 1,000 largest colonies in our dataset were binned into 10 deciles and normalized to the background distribution of all colonies. The bar graphs show fold enrichment over background for neighbor colonies in each decile before (**ii**) and after (**iii**) competition normalization. This analysis shows that large colonies tend to have small neighbors. This trend is significantly less pronounced after competition normalization.
(**E**) Batch effect normalization. The distribution of Pearson correlation coefficients among genetic interaction profiles of all queries (blue) and queries screened within the same batch (pink) are shown before and after batch effect normalization. For reference, the distribution of correlations among genes encoding members of the same protein complex is shown as a black line. Before batch correction, the average correlation among queries screened in the same batch is as high as the correlation of co-complex genes.

I developed normalization procedures that estimate and remove systematic biases in colony size arising from experimental factors (Chapters 2.4.1–2.4.5). I applied statistical techniques including spatial smoothing, quantile normalization, and linear discriminant analysis to substantially reduce colony size variability and improve correlation between independent replicates (Chapters 2.4.1–2.4.5). After removal of experimental artifacts, I use normalized colony sizes to fit a model estimating fitness and genetic interactions for each double mutant (Chapter 2.4.9). The resulting measure, termed the SGA score, captures single and double mutant fitness measurements and provides a quantitative genome-wide assessment of genetic interactions (Chapters 2.5–2.8).

2.4 SGA genetic interaction score

I model colony size as a multiplicative combination of double mutant fitness, time and systematic experimental factors. Specifically, for a double mutant carrying mutations in genes i and j, colony area C_{ij} can be expressed as:

$$C_{ij} = \alpha f_{ij} t s_{ij} e \qquad \textbf{Eq. 2.1}$$

where f_{ij} is the fitness of the double mutant, t is time, s_{ij} is the combination of all systematic factors, α is a constant scale factor, and e is a log-normally distributed error. In addition, double mutant fitness f_{ij} can be expressed as $f_{ij} = f_i f_j + \varepsilon_{ij}$, where f_i and f_j and the single mutant fitness measures, and ε_{ij} is the genetic interaction between genes i and j.

This model is motivated by two empirical observations. First, colony sizes of replicate mutants grown different amounts of time can be normalized by a single multiplicative factor, suggesting that colony area scales linearly with time. Second, colony sizes of double mutants having the same query mutation can be normalized by a single multiplicative factor, suggesting that colony area also scales linearly with the fitness of each mutation. I also find that most experimental factors (Chapters 2.4.1–2.4.5) tend to scale colony sizes, which motivates a multiplicative model.

Our goal is to fit this model to colony area obtained from systematic genome-wide SGA experiments, eliminate the contribution of systematic experimental effects and extract the biological factors of interest: the single mutant fitnesses f_i and f_j, and the genetic interaction score ε_{ij}. A summary of all sources of systematic variation and the corresponding normalization schemes are presented below.

2.4.1 Plate effect

Plate-to-plate variance in colony size can be substantial due to the variability in each plate's incubation times as well as to the contribution of the query mutant fitness (Figure 2.2A). From Eq. 2.1 above, we assume that observed colony area is a function of single mutant fitness and time:

$$C_{ij} = \alpha f_i f_j t s_{ij} e \qquad \textbf{Eq. 2.2}$$

as genetic interactions are rare and in most cases $\varepsilon_{ij} \approx 0$.

In a typical SGA screen, a plate consists of a single query mutant *j* crossed to an array of 384 array mutants pinned in quadruplicate. Each plate is grown until most colonies reach a minimum size, which is assessed visually, and plates are only imaged once. Thus, each plate's incubation time is directly linked to the fitness of the query mutant, and their product is relatively constant, i.e. $f_j\, t \approx c$, for some constant c. Given the SGA experimental design, optimized for cost and throughput, estimating f_j and t separately is an extremely challenging task. As a result, I opted to scale each plate by a composite normalization factor, which incorporates the product of f_j and t, and is computed as the plate middle mean (*PMM*, mean of the middle 60% of colonies on the plate). The corrected colony areas are computed as:

$$C'_{ij} = C_{ij} \frac{PMM_{global}}{PMM_k} \qquad \textbf{Eq. 2.3}$$

where PMM_{global} is derived from the PMM values across all plates.

Note that $PMM_k \approx f_j\, t_k$, where t_k is the incubation time for plate k. This normalization approach eliminates any dependence of the normalized colony size on the query mutation in the absence of a genetic interaction ($\varepsilon_{ij} = 0$). The new normalized colony size can now be expressed as:

$$C'_{ij} = \alpha f_i s_{ij} e \qquad \textbf{Eq. 2.4}$$

When an interaction is present, the normalized colony size depends on the query fitness as follows:

$$C'_{ij} = \alpha\left(f_i + \frac{\varepsilon_{ij}}{f_j}\right)s_{ij}e \qquad \textbf{Eq. 2.5}$$

As observed previously (Collins *et al.*, 2006), detecting interactions using these normalized colony sizes is relatively convenient. To detect an interaction, the double mutant of interest can simply be compared to all other double mutants sharing the same array mutation, given that, after plate normalization, a double mutant's colony size should reflect only the array mutant fitness, the associated systematic effects and the genetic interaction, if present.

Following this logic, I define a colony residual R_{ij} as

$$R_{ij} = C'_{ij} - \widetilde{C'}_{ij} \approx \alpha\frac{\varepsilon_{ij}}{f_j}s_{ij}e \qquad \textbf{Eq. 2.6}$$

which is simply the difference between the normalized colony size for double mutant *ij* (C'_{ij}) and the median normalized colony size for all double mutants sharing the same array mutation *i* ($\widetilde{C'_{ij}}$). As a result, R_{ij} should reflect the genetic interactions ε_{ij} we aim to measure, in addition to lingering systematic effects, s_{ij}.

There are several advantages for correcting lingering systematic effects in colony residual space as opposed to the original colony space. Colony size variation in the original colony space is the result of variation in single mutant fitness of the array strain (f_i), the systematic effects (s_{ij}) and the genetic interactions (ε_{ij}). Variation in colony residual measures, however, is free from the array fitness component. In addition, a substantial fraction of sys-

tematic effects is linked to colony position on the plate, which is constant for any given array mutant. Thus, colony residuals indirectly normalize for the colony positional effects as well. As a result, observed variation in colony residual space is more closely related to the quantity of interest, i.e. the genetic interaction between two genes, which facilitates normalization for the remaining systematic effects.

2.4.2 Row/column effect

Similar to other array-based genomic technologies, positional effects are a major contributor to systematic variation in colony sizes. For example, colonies located in the outermost rows and columns of a plate are visibly larger than internal colonies (Figure 2.2B), mostly because of a reduced colony density near the plate edges and the resulting availability of extra nutrients. Such row/column effect, when left uncorrected, introduces an abundance of spurious positive genetic interactions and undermines our ability to map connections within and between molecular pathways.

I estimate the extent of row/column effects on a plate-by-plate basis using a linear LOWESS smoothing (Cleveland, 1979) on normalized colony sizes. While the shape of the estimated trends appears to be consistent across plates, the severity of the effect is variable and the correction must be plate-specific. Row and column correction is applied as follows:

$$C'_{ij} = C_{ij}\left(\frac{\bar{r}}{r_m}\right)\left(\frac{\bar{c}}{c_n}\right) \qquad \textbf{Eq. 2.7}$$

where r_m and c_n are estimates derived from LOWESS smoothing for row m and column n, respectively, and $\bar{r}$ and $\bar{c}$ denote average row and average column factors.

2.4.3 Spatial effect

Plates often exhibit spatial gradients in colony size, such that, for example, colonies located on the left-hand side of the plate grow larger than colonies located on the right-hand side (Figure 2.2C). Such spatial gradients are potentially due to an uneven distribution of media across the plate or to a non-uniform temperature within the incubator, and result in the appearance of false positive and negative genetic interactions.

I identify and correct spatial gradients using a series of two-dimensional smoothing filters applied to colony residuals. First, a median spatial filter is applied on a 7×7 grid of colonies surrounding the position of interest. Colony residuals used for the smoothing are calculated on a log scale as follows:

$$R_{ij} = \log\left(\frac{C_{ij}}{\widetilde{C_{ij}}}\right) \qquad \textbf{Eq. 2.8}$$

Median filter estimates are then further smoothed with a simple average filter on a 10×10 grid to derive the final estimates of the surface gradient. The spatial normalization is applied as follows:

$$C'_{ij} = C_{ij} \frac{1}{\exp(z_{ij})} \qquad \textbf{Eq. 2.9}$$

where z_{ij} is the final spatial effect estimated by the spatial filters.

2.4.4 Competition effect

In order to afford the throughput necessary to map all pair-wise genetic interactions in yeast, SGA screens are performed using high density mutant arrays, each containing 384 mutant strains pinned in quadruplicate for a total of 1,536 colonies per plate. In such a crowded environment, colonies compete for nutrients, and small positional advantages, such as, for example, proximity to a low fitness neighbor, translate into substantial growth benefits. As a result, the largest double mutant colonies on a plate are typically associated with small neighbors and may be erroneously identified as strong positive genetic interactions (Figure 2.2D).

To normalize for the nutrient competition effect, I take a step-wise approach. First, colony residuals are binned into ten deciles based on the size of their smallest neighbor. Within each decile, competition effect is normalized by scaling the colonies by a multiplicative factor estimated from the linear LOWESS fit of colony residuals against the size of their neighbors. Finally, competition effect is corrected across deciles using quantile normalization, which matches the distribution of residual colony sizes within each decile to that of colonies with relatively healthy neighbors (60–80 percentile).

2.4.5 Batch effect

We found that SGA screens performed in batch, i.e. by the same person on the same day using the same robotic instrument, tend to share a common pattern of colony size variability (Figure 2.2E). This batch effect is likely caused by the fact that concurrent screens are influenced by common experimental factors, such as media stock, incubation time and the single mutant source array. The combination of these and other factors, that we might be unaware of, generates unusually small or large colonies that are shared by all screens in the batch and can be easily misinterpreted as real negative or positive genetic interactions. The batch-related non-biological signature is often stronger than the real biological signature and represents a severe impediment for predicting biological function using profiles of genetic interactions (Figure 2.2E).

Eliminating the batch effect is statistically challenging since batches are relatively small (5–10 screens), while their quantitative readouts are extremely large (100,000–200,000 colonies). Linear discriminant analysis (LDA) is often used for finding linear combinations of features that are highly discriminative across sets of classes and thus can be applied for estimating batch signatures. Unlike other similar approaches, such as principal component analysis (PCA) and singular value decomposition (SVD), LDA is a supervised method that is able to leverage pre-defined class distinctions, which correspond to batches in our experimental context.

Specifically, we can construct a matrix X of double mutant colony data, such that rows consist of screens grouped into different batches. Subsequently, the objective of a multi-class LDA is to find projections of the data that maximize the following criterion:

$$J(v) = \frac{\det\left(v^T S_b v\right)}{\det\left(v^T S_w v\right)} \qquad \textbf{Eq. 2.10}$$

where

$$S_b = \sum_{i=1}^{C} N_i (\mu_i - \mu)(\mu_i - \mu) \qquad \text{and} \qquad S_w = \sum_{i=1}^{C} \sum_{x_k \in C_i} (x_k - \mu)(x_k - \mu)$$

for C different batches where μ_i is the centroid of batch i and μ is the global centroid.

Intuitively, this criterion measures the ratio of between-batch differences to within-batch differences and the optimal vector v represents linear combinations of the features that maximize this batch signal-to-noise ratio. Optimal projections v can be readily computed by finding the principal eigenvectors of the matrix $S_w^{-1} S_b$. Principal eigenvectors identified through LDA are effectively focused towards cross-batch variation, which should avoid removing potentially important biological signatures from the data.

To normalize the batch effect, we construct a matrix V_n consisting of the top n principal eigenvectors and normalize the original data matrix as follows:

$$X' = X - X * V_n^T V_n \qquad \textbf{Eq. 2.11}$$

This computes the projection of the original colony size data onto the space of "batch signals" and removes these signatures from the data. The number of components removed is chosen as the principal eigenvectors satisfying $\lambda_i/\lambda_{max} > 0.1$ for eigenvalues λ_i. Based on our

empirical evaluation, removing a higher or a lower number of principal eigenvectors results in over- or under-correction of batch effects, as measured by the distribution of within-batch similarities.

2.4.6 Linkage filter

Most physically linked genomic loci, i.e. loci positioned close to one another on the same chromosome, are also genetically linked, i.e. they tend to be inherited together due to the scarcity of recombination events occurring between them (Chapter 4). As a result, crossing an SGA query mutation to a genetically linked array mutation produces fewer double mutant meiotic spores, compared to a pair of independently segregating loci, and results in a visibly smaller double mutant colony, which can be confused with a negative genetic interaction (Figure 4.2).

To eliminate these spurious negative interactions from our dataset, I identify and filter out data involving the chromosomal region affected by linkage for each individual query. Array genes are sorted by chromosomal position and the resulting genetic interaction profile is smoothed using a moving average window of 40 kb. The boundaries of a linkage region are defined as chromosomal positions closest to the query such that the smoothed genetic interaction score is no longer negative. Using this approach, I found that the boundaries of linkage regions are relatively constant across large chromosomal regions, but are different in other regions, suggesting that linkage patterns reflect local patterns of meiotic recombination (Chapter 4). On average, linkage affects 410 ± 150 kb around the query locus.

2.4.7 Large colony filter

Due to unexpected diploidization of certain strains during SGA selection steps, I filtered any double mutant presenting abnormally large colonies (more than 150% the size of a median colony) from our analysis.

2.4.8 Jackknife filter

To eliminate exceptionally small or large outliers from each group of replicates for a given double mutant, I applied a jackknife procedure to assess the impact of each colony on the overall variance of the replicates. Colony i is filtered if $\sigma_i < 0.1\sigma$, where σ is the variance of all replicates and σ_i the jackknife variance derived from holding out colony i.

2.4.9 Estimation of fitness, genetic interactions and confidence measures

As mentioned above, in the absence of genetic interactions, the normalized colony sizes directly reflect the single mutant fitness of the array strain. Unlike genetic interactions, which are most sensitive to variance on colony residuals, precise estimation of single mutant fitness depends critically on proper normalization at the colony size level. For example, positional biases consistent across plates do not present a problem for detecting interactions (as these biases equally affect all SGA screens and are indirectly normalized for in colony residual space), but could lead to highly inaccurate estimates of fitness.

To avoid such inaccuracy, we constructed several array configurations, where subsets of non-essential single mutant deletion strains were placed in randomized positions. We performed control SGA screens by crossing a neutral query mutant to the various array configu-

rations, and derived single mutant fitness estimates from each array by computing the median normalized colony size across all screens. Variance in single mutant fitness estimates was obtained from bootstrap sampling of the median, and final fitness measures were derived by pooling across all array configurations.

Quantitative estimates of genetic interactions are derived from normalized colony residuals. Specifically, for each replicate colony k of a double mutant ij, colony residual is computed as:

$$R_{ij_k} = C'_{ij_k} - \widetilde{C'}_{ij_k} \qquad \textbf{Eq. 2.12}$$

where C'_{ij_k} represents the normalized colony size and $\widetilde{C'}_{ij_k}$ the median across all queries.

An interaction estimate is derived by averaging across all N_{ij} replicates of a double mutant:

$$I_{ij} = \frac{1}{N_{ij}} \sum_k R_{ij_k} \qquad \textbf{Eq. 2.13}$$

To estimate our confidence in a given interaction estimate, I calculate two distinct variance measures: (1) a measure of local variability of the replicate colonies; (2) a measure of the expected double mutant variance, given the estimated variances of the array and query single mutants.

The first variance measure is derived from the unbiased estimate of variance on the interaction mean:

$$\sigma_{ij} = \sqrt{\frac{\sum_k \left(R_{ij_k} - I_{ij}\right)}{N_{ij} - 1}} \qquad \textbf{Eq. 2.14}$$

As reported in previously (Collins *et al.*, 2006), variance estimates derived from local double mutant colony reproducibility can dramatically underestimate the true variance. This is likely due to the fact that replicate colonies are not truly independent, as they are situated next to each other on the plate and experience much of the same systematic variation. I therefore derive a second more accurate measure of variance by combining error estimates across all occurrences of the array and query strains for the double mutant of interest.

As stated in Eq. 2.2, we assume that, in the absence of genetic interactions ($\varepsilon_{ij} \approx 0$), a double mutant colony size is proportional to the product of the two single mutant fitnesses. Thus, we can assume that this colony size is log-normally distributed with variance contributions from both the query and the array mutations:

$$C_{ij} \sim \alpha f_i f_j e^X \quad \text{for} \quad X \sim N\left(0, \sigma_i^2 + \sigma_j^2\right) \qquad \textbf{Eq. 2.15}$$

where σ_i^2 and σ_j^2 are array- and query-specific variance contributions.

Array-specific variance is calculated using normalized colony sizes derived from control SGA screens. Query-specific variance is calculated by combining variances of all double mutants sharing the corresponding query mutation.

For each measured interaction, I calculate both the measured standard deviation of the interaction estimate itself (σ_{ij}; Eq. 2.14) and the expected combined variance of the array and the query mutations, computed as a geometric standard deviation:

$$\sigma_{ij_{\text{expected}}} = e^{\sqrt{\sigma_i^2+\sigma_j^2}} \qquad \textbf{Eq. 2.16}$$

To integrate these two standard deviations into a single confidence measure, I compute a geometric mean of the two normal cumulative distribution functions (*CDF*):

$$p = \sqrt{CDF\left(-\left|\frac{I_{ij}}{\sigma_{ij}}\right|\right)\cdot CDF\left(-\left|\frac{\log\left(C_{ij}\right)}{\log\left(\sigma_{ij_{\text{expected}}}\right)}\right|\right)} \qquad \textbf{Eq. 2.17}$$

The resulting p-value p measures the probability of observing a genetic interaction score equal or more extreme than I_{ij}, given the variances of the query mutant i, array mutant j and double mutant ij.

2.4.10 Post-interaction filters

In our experimental setup, array mutant strains are distributed across 14 plates in chromosomal order, such that on average each individual plate contains mutant strains from one to three chromosomes. As a result, plates harboring array mutant strains that are genetically linked to a given query mutations contain an exceptionally large number of inviable or slow growing colonies (Chapter 2.4.6). These colonies generate extreme competition effects that are not completely removed by our normalization procedures. Thus, to avoid introducing

numerous false positives, I filter any positive genetic interaction that is adjacent to a strain affected by genetic linkage.

2.5 A high-resolution catalog of yeast single mutant fitness

To validate our approach for correcting systematic experimental effects, I used a panel of control SGA screens to measure colony areas for 4,635 viable deletion mutants and 1,388 temperature-sensitive or hypomorphic alleles of essential genes. Control SGA screens are performed using a query strain deleted for the neutral *URA3* locus, which is unlikely to interact with other genes and thus produces an output array of reference single mutant colonies. I processed the colony areas using the SGA score (Chapter 2.4) and generated the first genome-wide colony size-based single mutant fitness catalog (Materials & methods 2.10.5).

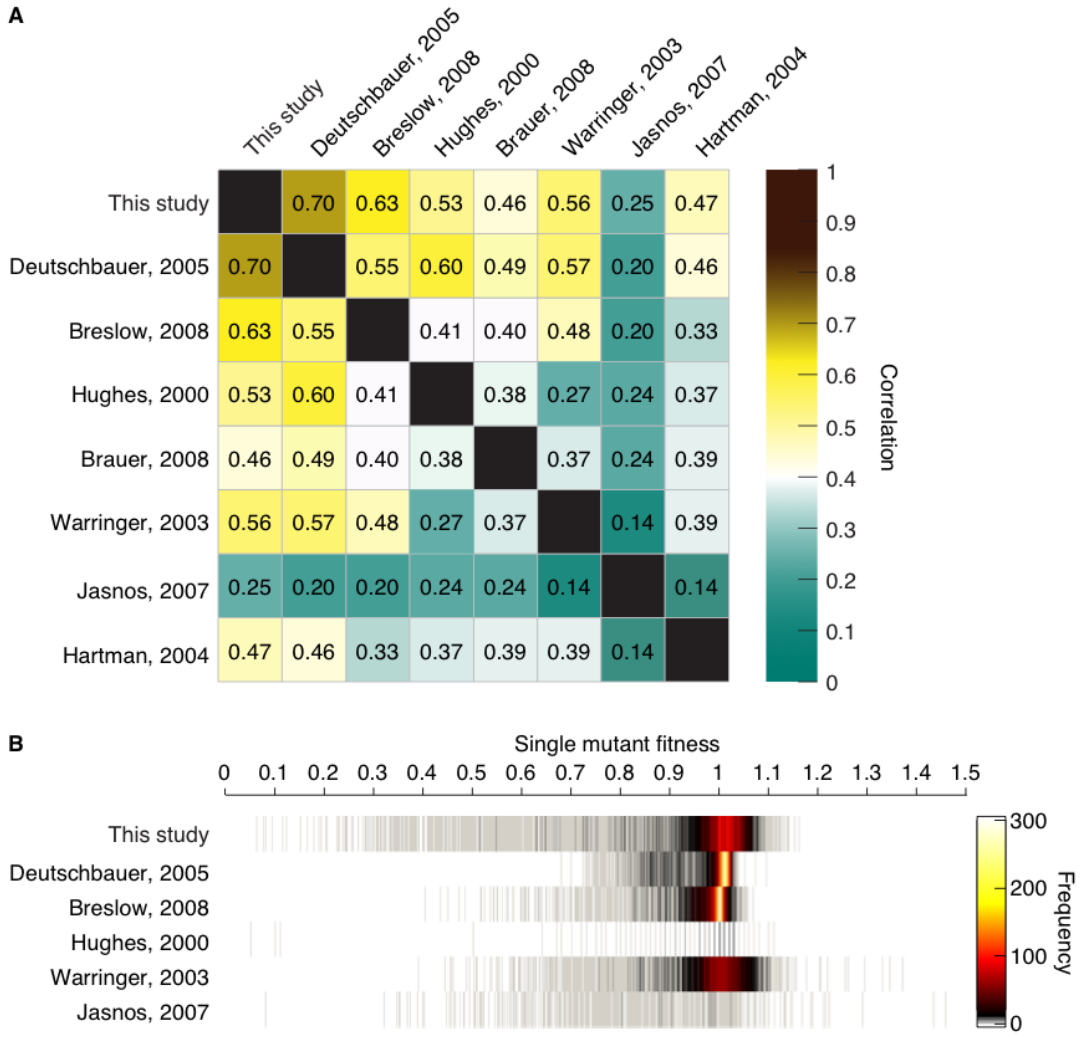


Figure 2.3. Comparison of single mutant fitness measures.

(**A**) Comparison of fitness or relative growth measurements for nonessential gene deletions mutants derived from eight independent approaches: colony size measurements (this study); competitive growth analyzed by barcode hybridization (Deutschbauer *et al.*, 2005), flow cytometry (Breslow *et al.*, 2008) or gene expression profiles (Hughes *et al.*, 2000); gene expression microarrays (Brauer *et al.*, 2008); liquid growth profiling (Warringer *et al.*, 2003; Jasnos and Korona, 2007); and a spot assay on solid growth medium (phenotypic assay) (Hartman and Tippery, 2004).
(**B**) Distribution of single mutant fitness measures reported by the studies described in (A) reporting fitness or relative growth rate.

Comparison to previously published fitness or relative growth rate studies showed that SGA captures fitness as well or better than other available methodologies. Indeed, together

with barcode-based measurements from competitive growth assays, our colony size-based fitness catalog had the highest average cross-study correlation (Figure 2.3A). In addition, the range of growth defects detectable by our method substantially exceeded that of most other studies (Figure 2.3B). For example, competitive growth assays, while testing a similar number of strains (Deutschbauer *et al.*, 2005; Breslow *et al.*, 2008), do not report fitness for mutants with relative growth defect greater than 50% (Figure 2.3B).

In order to obtain high-quality quantitative genetic interactions, both single and double mutant fitness must be accurately measured. I validated the quality of our double mutant fitness measurements by evaluating the reproducibility of 211 genome-wide SGA screens performed in duplicate. Replicate double mutant fitness measurements were highly correlated (average Pearson correlation of 211 pairs of replicate screens, $R = 0.89$; Figure 2.4), indicating that colony size-based fitness estimates should provide a robust basis for accurate assessment of genetic interactions.

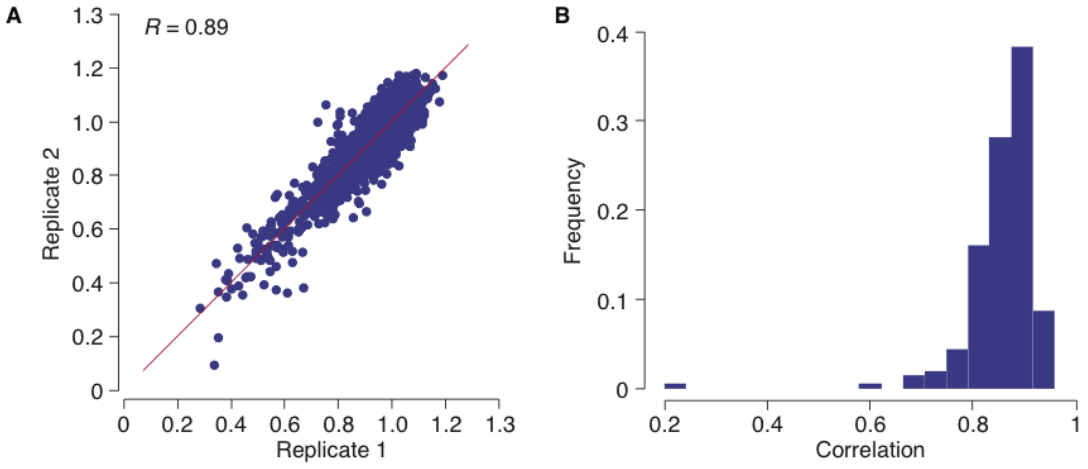


Figure 2.4. Reproducibility of double mutant fitness measurements.

(**A**) Correlation of double mutant fitness measures obtained from two independent replicates of a representative genome-wide SGA screen. Red line, $y = x$.
(**B**) Distribution of correlations between double mutant fitness measures obtained from 211 genome-wide SGA

screens conducted in duplicate. Average Pearson correlation, $R = 0.89$.

2.6 Evaluation of the reproducibility of genetic interaction measurements

We screened 1,712 *S. cerevisiae* query genes, including 334 conditional or hypomorphic alleles of essential genes, for a total of ~5.4 million gene pairs spanning all biological processes (Materials & methods 2.10.1). These queries were selected randomly with respect to function (Costanzo *et al.*, 2010); however, preference was given to mutants exhibiting fitness defects. I scored the SGA screens using the SGA score and used multiple approaches to evaluate the technical reproducibility and the biological value of the identified genetic interactions.

First, I compared genetic interactions measured from independent replicate experiments and found them to be highly reproducible ($R = 0.87$; Figure 2.5A).

Second, I assessed the reproducibility of interactions among reciprocal gene pairs (query mutant A crossed to array mutant B versus query mutant B crossed to array mutant A). In an SGA experiment, query and array deletion mutants for the same gene are subject to slightly different experimental conditions. In particular, query mutants with severe fitness defects are allowed to grow longer than their corresponding array mutants, resulting in greater resolution for detecting genetic interactions. Despite these differences, I observed good agreement among reciprocal genetic interaction pairs ($R = 0.67$; Figure 2.5B), especially when I restricted the analysis to more confident interactions ($R = 0.75$). Notably, agreement between

reciprocal interactions decreases by ~30–40% when experimental noise factors are left uncorrected (Baryshnikova *et al.*, 2010b).

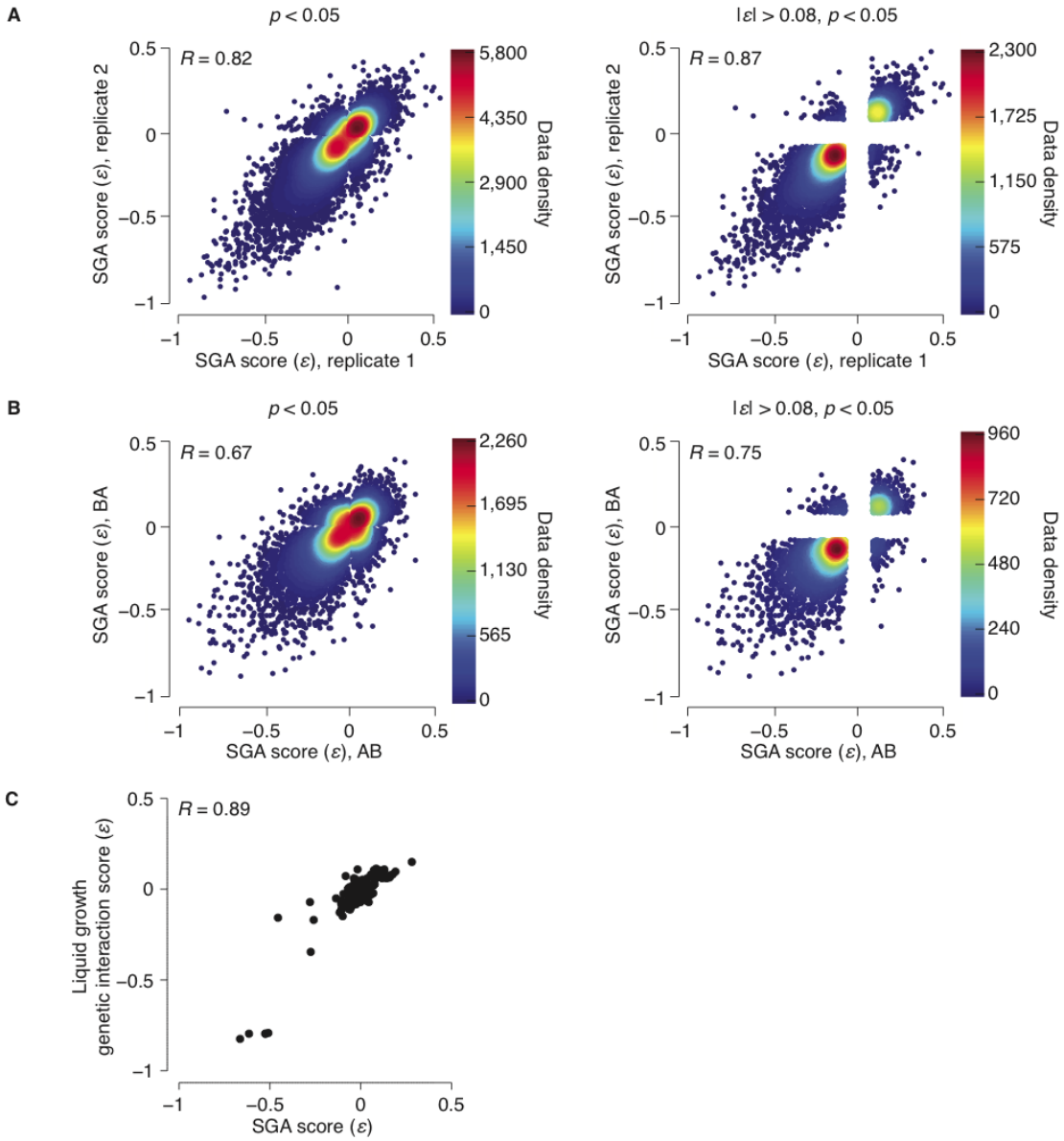


Figure 2.5. Reproducibility of quantitative genetic interactions.

(**A**) Scatter plots of genetic interactions derived from 211 genome-wide SGA screens conducted in duplicate. Pearson correlation coefficients were computed after applying a lenient ($p < 0.05$) or intermediate ($|\varepsilon| > 0.08$, $p < 0.05$) confidence threshold.
(**B**) Scatter plots of genetic interaction measures between reciprocal gene pairs. Pearson correlation coefficients were computed after applying a lenient ($p < 0.05$) or intermediate ($|\varepsilon| > 0.08$, $p < 0.05$) confidence threshold.
(**C**) A scatter plot illustrating the overlap between genetic interaction scores for 239 unique gene pairs extracted from the large-scale SGA dataset and a small-scale high-resolution liquid growth profiling study (St Onge *et al.*,

2007).

Finally, I compared our SGA scores to a smaller-scale high-resolution liquid growth study (St Onge *et al.*, 2007), which measured quantitative genetic interactions among 26 genes conferring resistance to the DNA damaging agent methyl methanesulfonate (MMS). I observed a highly significant correlation between 239 gene pairs tested in both studies ($R = 0.89$; Figure 2.5C). Such correlation was not simply attributable to extreme genetic interaction scores because even modest interactions were highly correlated ($R = 0.73$ for SGA score absolute value $|\varepsilon| < 0.4$; Figure 2.5C).

2.7 Functional impact of systematic effects

I evaluated the functional utility of positive and negative genetic interactions identified by the SGA score by comparing them to commonly used functional benchmarks: annotation to the same Gene Ontology (GO) biological process term and overlap with protein-protein interactions (Figure 2.6, Figure 2.7).

Early experimental evidence showed that genetic interactions are enriched among genes acting in the same biological process (Tong *et al.*, 2001; Tong *et al.*, 2004). As a result, the frequency of genetically interacting gene pairs annotated to the same GO term serves as an objective metric for quantifying the functional utility of a given dataset (Myers *et al.*, 2006). I used precision/recall analysis to assess the frequency of co-annotated gene pairs along a list of gene pairs sorted by their genetic interaction score (Materials & methods

2.10.2). Precision was calculated as the fraction of gene pairs above a certain genetic interaction score threshold that are co-annotated to the same GO term, while recall corresponded to the total number of co-annotated gene pairs at that threshold (Materials & methods 2.10.2). Both negative and positive genetic interactions overlapped with co-annotated gene pairs more frequently than expected by chance (Figure 2.6A). In addition, interactions derived from a version of the SGA score that omits systematic effect normalization were significantly less enriched for co-annotated gene pairs (Figure 2.6A): at 30% precision, negative SGA scores recovered over four-fold more co-annotated gene pairs than unnormalized negative SGA scores, indicating that experimental variability contributes substantially to the false positive rate of large-scale SGA screens. The SGA score had similar improvement over genome-wide screens processed using the S-score (Collins *et al.*, 2006), which does not account for many of the systematic effects we identified (Figure 2.6A). Moreover, similar trends were observed when genetic interactions were evaluated against a protein-protein interaction standard (Figure 2.6B). A shared physical interaction, similarly to a shared GO annotation, is indicative of a functional relationship between genes and their encoded proteins.

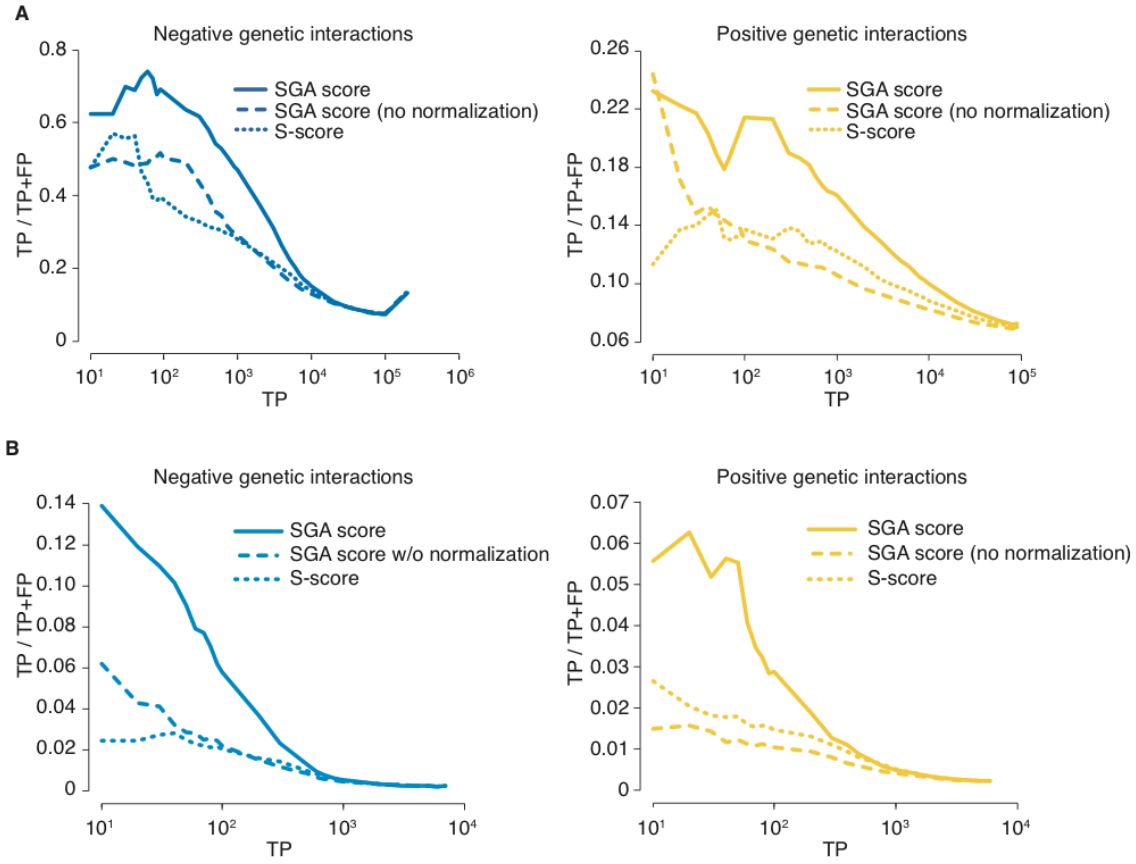


Figure 2.6. Functional evaluation of negative and positive genetic interactions.

Plots of precision (TP/TP+FP) versus recall (TP) for negative and positive genetic interactions, as determined by the SGA score or the S-score. A version of the SGA score that omits systematic effect normalization is also plotted for comparison [SGA score (no normalization)]. Precision and recall were calculated as described in Materials & methods 2.10.2.

(**A**) Evaluation against the Gene Ontology (GO) co-annotation standard. A set of gold standard GO terms was chosen using previously described criteria (Myers *et al.*, 2006). True positive interactions (TP) were defined as those involving both genes annotated to the same GO term.

(**B**) Evaluation against the protein-protein interaction standard. True positive interactions (TP) were defined as those involving genes encoding physically interacting proteins reported in the BioGRID database (Stark *et al.*, 2011).

TP, true positive; FP, false positive.

In addition to its effect on direct genetic interactions, accounting for systematic noise had a striking impact on the functional utility of genetic interaction profiles (Figure 2.7). For example, at 30% precision genetic interaction profiles computed using the SGA score recover ~3,000 co-annotated gene pairs, compared to less than ten gene pairs obtained using unnormalized scores (Figure 2.7A). The main contributor to such improvement for both individual genetic interactions and genetic interaction profiles was batch effect normalization (Figure 2.8).

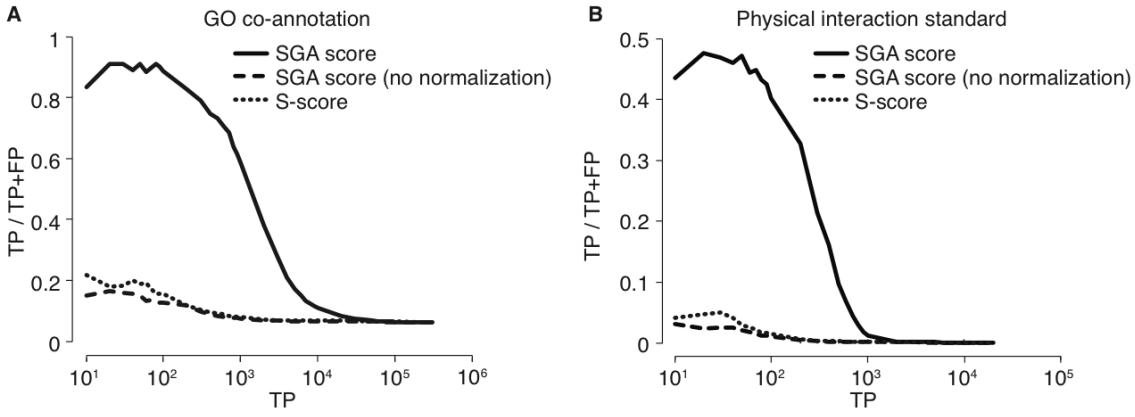


Figure 2.7. Functional evaluation of genetic interaction profile similarities.

Plots of precision (TP/TP+FP) versus recall (TP) for genetic interaction profile similarities computed using the SGA score or the S-score. A version of the SGA score that omits systematic effect normalization is also plotted for comparison [SGA score (no normalization)]. Precision and recall were calculated as described in Materials & methods 2.10.2. Pearson correlation was used to compute profile similarity for every pair of array mutant strains across profiles consisting of interactions with the 1,712 query mutant strains.
(**A**) Evaluation against the Gene Ontology (GO) co-annotation standard. A set of gold standard GO terms was chosen using previously described criteria (Myers *et al.*, 2006). True positive interactions (TP) were defined as those involving gene pairs annotated to the same GO term.
(**B**) Evaluation against the protein-protein interaction standard. True positive interactions (TP) were defined as those involving genes encoding physically interacting proteins reported in the BioGRID database (Stark *et al.*, 2011).
TP, true positive; FP, false positive.

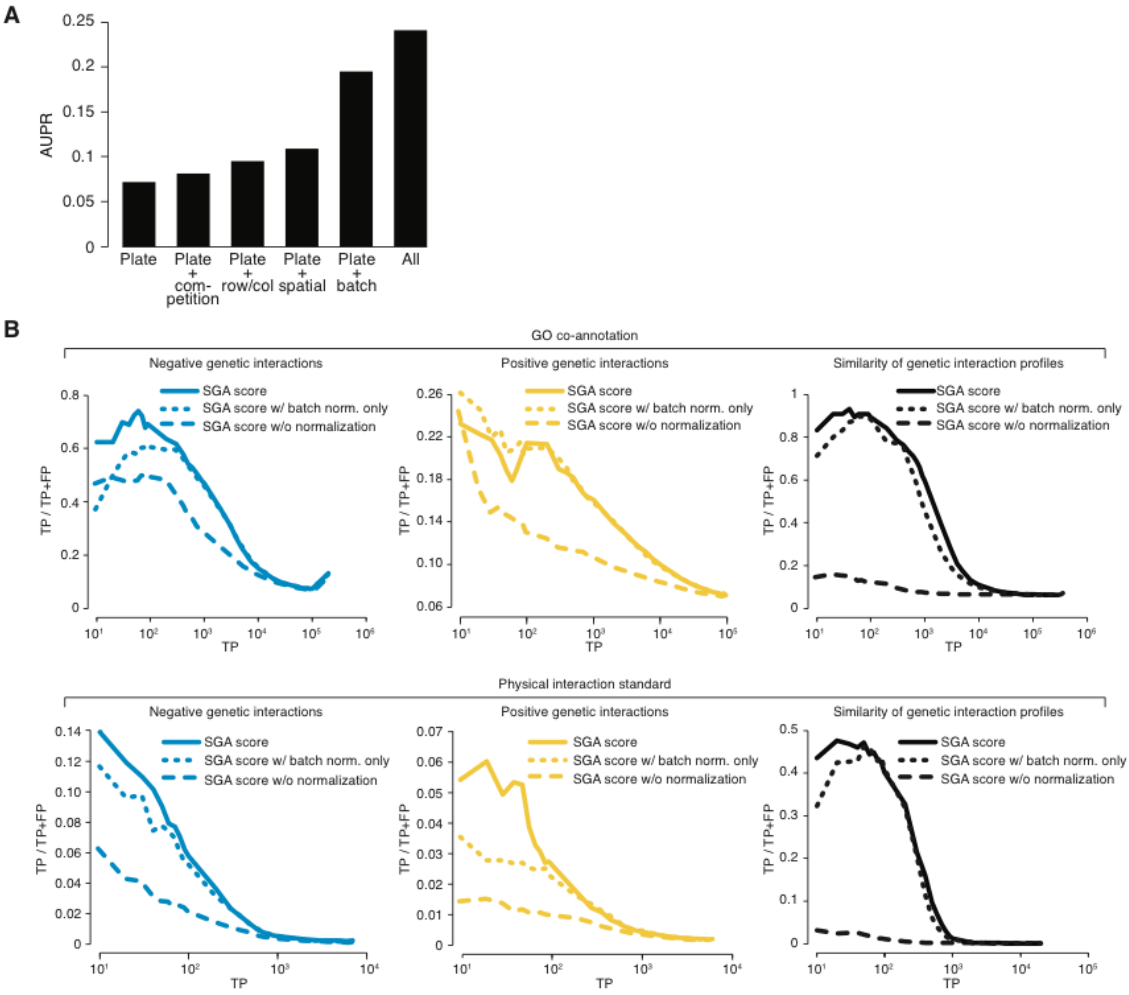


Figure 2.8. Relative contributions of each systematic effect

(**A**) The contribution of each normalization procedure to the functional performance of the genetic interaction dataset was assessed by evaluating SGA scores against co-annotated gene pairs and measuring the area under the precision-recall (AUPR) curve. The genetic interaction dataset was scored five times, each time omitting all correction steps except for plate normalization and the correction procedure indicated on the x-axis. Consistent with our other evaluations, each correction alone contributes to improve data quality. The most significant contribution is made by the batch effect correction.
(**B**) Comparative functional evaluation of negative interactions, positive interactions and genetic interaction profiles to assess the impact of the batch effect. SGA scores and profiles were evaluated using normalized data (SGA score), data without normalization (SGA score w/o normalization) or data in which only the batch effect normalization was applied (SGA score w/ batch norm. only). Pearson correlation was used to compute profile similarity for every pair of array mutant strains across 1,712 query mutant strains. True positive interactions were defined as those involving co-annotated gene pairs (Myers *et al.*, 2006) or protein-protein interactions (Stark *et al.*, 2011). Precision and recall were calculated as described in Materials & methods 2.10.2.
TP, true positive; FP, false positive.

To further characterize our dataset, I estimated false negative and false positive rates for a range of SGA score and p-value thresholds using a strategy based on the reproducibility of reciprocal gene pairs and a set of reference genetic interactions identified by small scale studies (Materials & methods 2.10.3; Figure 2.9A). I chose an intermediate stringency threshold ($|\varepsilon| > 0.08$, $p < 0.05$) and used this filtered data set for all subsequent analyses.

At the chosen threshold, I uncovered ~170,000 unique genetic interactions, a threefold increase over all existing data (Figure 2.9B). Our dataset captures ~35% and ~20% of previously reported negative and positive genetic interactions, respectively, with an estimated false discovery rate of ~40% in both cases (Table 2.1; Figure 2.9C). Notably, false discovery rate is substantially lower for SGA experiments conducted in duplicate (~21% and ~28% for negative and positive interactions, respectively), suggesting that data quality will be further improved by future experiments (Figure 2.9C).

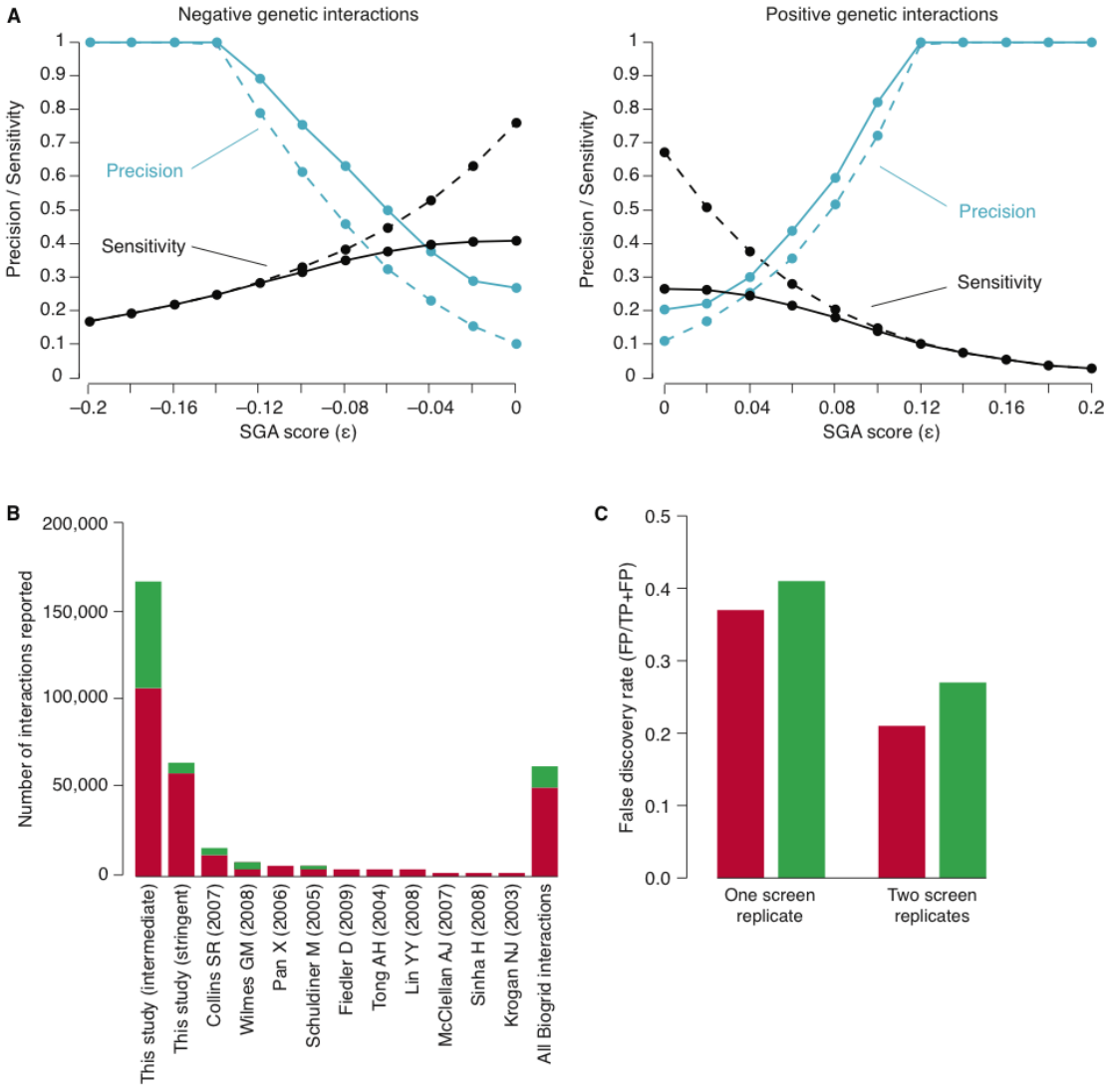


Figure 2.9. Characterization of SGA genetic interaction score thresholds.

(**A**) Sensitivity (black line) and precision (blue line) were systematically measured (as described in Materials & methods 2.10.3) for positive and negative genetic interactions at various SGA score (ε) cutoffs (x-axis) and two p-value cutoffs ($p < 0.5$, dashed line, and $p < 0.05$, solid line).
(**B**) Dataset size comparison. I compared the total number of negative (red) and positive (green) genetic interactions at our intermediate and stringent cutoffs to those reported by the 10 largest genetic interaction datasets available in BioGRID (Stark *et al.*, 2011). The intermediate confidence SGA dataset represents a 3-fold increase over the total number of genetic interactions reported so far.
(**C**) The same analysis as in (A) was repeated on 211 screens performed in duplicate. False discovery rate was computed as 1 – precision.
TP, true positive; FP, false positive.

Overall, I found about twice as many negative as positive genetic interactions (Table 2.1). Also, negative interactions tended to be more functionally informative than positive interactions, both at the direct SGA score level (Figure 2.6) and at the level of genetic interaction profile similarities (Figure 2.10). This presumably reflects a higher incidence of noise or less functionally informative interaction profiles in the positive interaction dataset. However, optimal predictive power is obtained using complete genetic interaction profiles, which include both negative and positive genetic interactions (Figure 2.10).

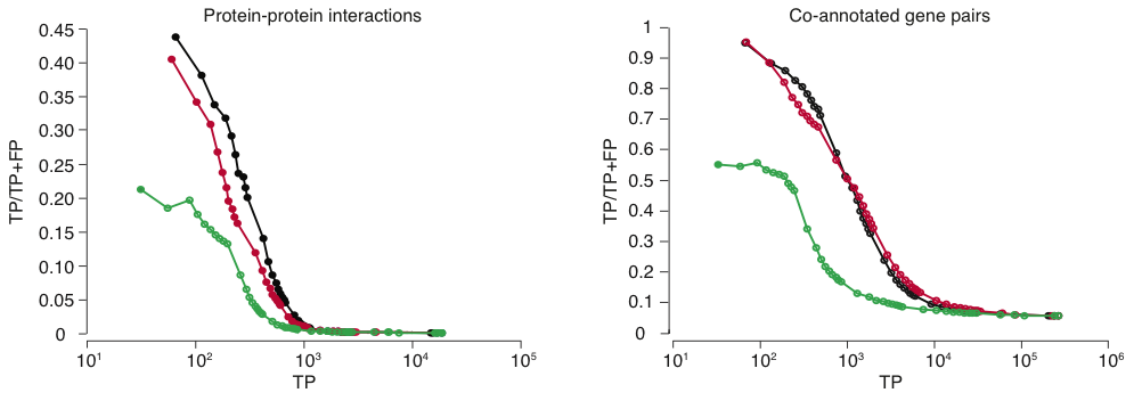


Figure 2.10. Functional evaluation of negative and positive genetic interaction profile similarities.

Plots of precision (TP/TP+FP) versus recall (TP) for genetic interaction profile similarities computed using only negative genetic interaction scores (red line), only positive genetic interaction scores (green line) or both (black line). True positive interactions were defined as those involving co-annotated gene pairs (Myers *et al.*, 2006) or protein-protein interactions (Stark *et al.*, 2011). Precision and recall were calculated as described in Materials & methods 2.10.2.
TP, true positive; FP, false positive.

2.8 Comparative analysis of function prediction capacity

We tested the ability of our genetic interactions to predict novel functional associations in comparison to other genomic datasets (Figure 2.11). We chose a subset of 1,885 GO biological process categories associated with at least 3 and at most 3,000 genes, and used the

GeneMANIA algorithm (Mostafavi *et al.*, 2008) to re-assign genes to GO categories based on specific data sources only (Materials & methods 2.10.4). For any given category, prediction performance was evaluated by measuring the number of correctly assigned genes relative to all predictions (precision), as well as the number of correctly assigned genes relative to all known annotations for the category (recall).

Despite covering only 30% of the genome, our SGA dataset was able to predict gene function as well as the combination of two exhaustive AP/MS studies and better than any other major data type available in the literature (Figure 2.11A). Compared to individual datasets, our data exhibited the second best predictive power by showing an average ~11% precision across all tested categories (Figure 2.11B). Furthermore, excluding our SGA dataset from the compendium of all available genetic interactions results in the greatest loss in average precision (Figure 2.11C), suggesting that our data contributes a substantial amount of unique functional information not captured by other genetic interaction studies.

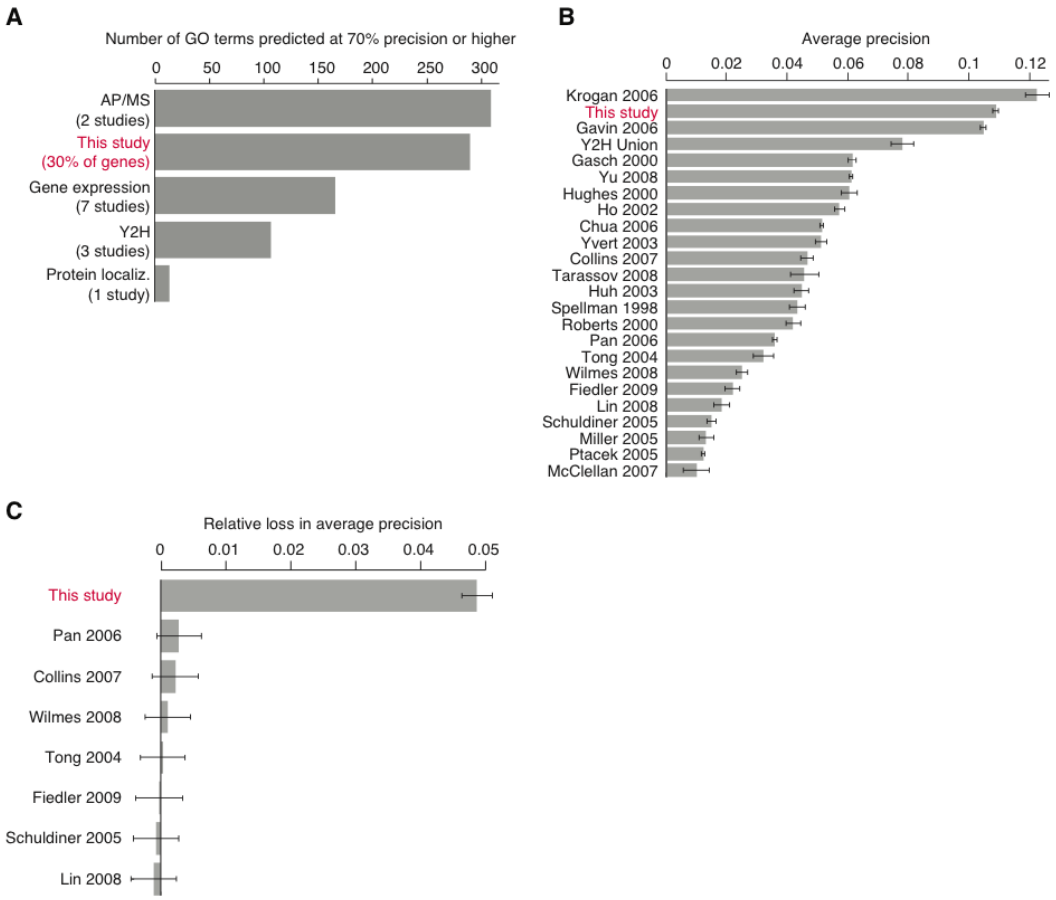


Figure 2.11. Comparative function prediction capacity of the genetic interaction dataset.

(**A**) The performance of the four major data types (genetic interactions, protein-protein interactions, gene expression and protein localization) in predicting gene function was determined by the GeneMANIA algorithm (Mostafavi *et al.*, 2008) as described in Materials & methods 2.10.4, and measured as the total number of GO terms predicted at 70% precision or higher at 10% recall.
(**B**) Comparison of individual high-throughput datasets in predicting gene function.
(**C**) Comparison of high-throughput genetic interaction studies using the leave-one-out approach.

2.9 SGA score application to other functional genomics datasets

The development of the SGA score established a standard computational framework for processing colony size data in various experimental contexts, and made an important contribution to our global understanding of yeast biology.

In addition to the collection of 1,700 genome-wide SGA experiments, described in this chapter, I used the SGA score to map quantitative genetic interactions among ~185,000 gene pairs acting in yeast metabolism (Szappanos *et al.*, 2011). Metabolism is one of the most extensively characterized cellular systems and an ideal test bed for investigating the molecular mechanisms underlying genetic interactions. Hundreds of chemical reactions connecting enzymes and their products have been modeled through high quality network reconstructions, as well as computational algorithms that predict the phenotypic outcomes of network perturbations. Systematic SGA experiments, together with rigorous SGA score processing, provided experimental evidence to refine these theoretical models and correct previous misannotations (Szappanos *et al.*, 2011). The theoretical models, in turn, offered a mechanistic explanation for the distribution of negative and positive genetic interactions in the empirical data and showed that the genes with the highest genetic interaction degree are the ones that impact the abundance of the largest number of key metabolites in the cell and therefore contribute to the largest number of biological processes (Szappanos *et al.*, 2011).

Experimental strategies similar to SGA have been developed in other model organisms, which, just like *S. cerevisiae*, form single colonies that can be quantitatively measured as a proxy for fitness and genetic interactions (Chapter 1.4). For example, I applied the SGA score methodology to map genetic interactions between 222 genes involved in DNA replication and repair, chromatin remodeling and intracellular transport in the fission yeast *Schizosaccharomyces pombe* (Dixon *et al.*, 2008). The SGA score identified 669 high confidence negative genetic interactions, which were highly reproducible and enabled the investigation of cross-species connectivity of ortholog genes (Chapter 1.4).

In addition to loss-of-function alleles, such as deletion mutants, SGA has also been used to examine gain-of-function perturbations, such as increased gene dosage (Sopko *et al.*, 2006; Sharifpoor *et al.*, 2012). I adapted the SGA score algorithm to enable quantitative analysis of synthetic dosage lethality (SDL) experiments, where a genome-wide overexpression library was combined with a set of deletion mutants to uncover mutant-specific sensitivity to gene dosage. This approach exploits the idea that high levels of a protein are often innocuous in a wild-type cell but may be deleterious in a deletion mutant defective for the protein's negative regulator (Kroll *et al.*, 1996; Sopko *et al.*, 2006). Consistent with this idea, SDL analysis of 92 protein kinases, coupled with quantitative SGA score processing, identified ~1,300 interactors, which were highly enriched for known kinase targets, physical interactors of kinases and phosphoproteins (Sharifpoor *et al.*, 2012). The integration of SDL with positive and negative genetic interactions revealed numerous recurring network motifs, which were predictive of regulatory relationships between kinases, their substrates and other proteins (Sharifpoor *et al.*, 2012).

Knowledge gained through application of the SGA score to colony-based high-throughput genetic surveys spans multiple biological processes, model organisms and perturbation strategies (Dixon *et al.*, 2008; Costanzo *et al.*, 2010; Szappanos *et al.*, 2011; Sharifpoor *et al.*, 2012). More importantly, the development of the SGA score has improved our understanding of the experimental conditions affecting colony growth in high density mutant arrays and helped us to refine the experimental design of our future high-throughput endeavors. The ideas behind the SGA score can be extended beyond genetic interactions to other array-based interaction mapping technologies, such as yeast two-hybrid (Yu *et al.*,

2008) or protein-fragment complementation assay (Tarassov *et al.*, 2008), which rely on colony growth to identify physically interacting proteins. Protein-DNA binding can also be assessed using colony-based methods, such as for example yeast one-hybrid, which maps transcription factors interacting with genomic regions of interest (Reece-Hoyes *et al.*, 2011). A version of the SGA scoring method is currently being adapted to yeast-one hybrid experiments (C. Myers, personal communication).

2.10 Materials & methods

2.10.1 SGA query construction and screening

SGA query strain construction and screening were conducted as described in (Baryshnikova *et al.*, 2010a). A total of 1,712 query mutants were screened against an array of 4,293 non-essential deletion mutants. 408 array mutants were removed by quality control procedures or due to incompatibility with the SGA technology, resulting in a final dataset of 1,712 queries × 3,885 arrays. All screens were conducted a single time with 4 replicate colonies per double mutant, except for 211 screens that were performed twice.

2.10.2 Precision/recall analysis against functional standards

Gene pairs are sorted based on their genetic interaction scores or Pearson correlation coefficient, such that the strongest values are sorted to the top of the list (i.e., in descending order for positive values and ascending order for negative values). Each pair is associated with functional association score $S = \{1,-1\}$, based on the chosen functional standard (GO co-annotation or protein-protein interactions), such that 1 indicates that the pair is a positive in

the standard and –1 indicates that it is a negative. Pairs not reported in the functional standard are assigned a score of 0 and do not contribute to the calculation.

At each position n in the list, recall $R(n)$ is calculated as:

$$R(n) = \sum_{n_i \le n} \left[\frac{S(n_i) + 1}{2} \right]$$

where $S(n_i)$ is the functional score at position n_i.

At each position n in the list, precision $P(n)$ is calculated as:

$$P(n) = \frac{R(n)}{\sum_{n_i \le n} |S(n_i)|}$$

where $R(n)$ is recall at position n and $S(n_i)$ is the functional score at position n_i

2.10.3 Estimation of false negative and false discovery rates

Sensitivity is defined as the fraction of known interactions identified by an assay:

$$sensitivity = \frac{TP}{TP + FN}$$

where TP is the number of true positives and FN is the number of false negatives.

I estimated SGA sensitivity by evaluating previously published genetic interactions stored in the BioGRID database (Stark *et al.*, 2011) as of May 4th, 2009. By systematically varying SGA genetic interaction score (ε) and p-value cutoffs, I calculated the number of true

positives (*TP*) as the number of BioGRID interactions with SGA scores and p-values more extreme than the chosen cutoff (Figure 2.9, sensitivity). For negative genetic interactions, I considered BioGRID interactions annotated as phenotypic enhancement, synthetic growth defect and synthetic lethality. For positive genetic interactions, I considered BioGRID interactions annotated as phenotypic suppression and synthetic rescue.

Precision is the fraction of true interactions in the set of all identified hits:

$$precision = \frac{TP}{TP + FP}$$

where *TP* is the number of true positives and *FP* is the number of false positives.

I estimated precision of our positive and negative genetic interactions using the following strategy. Given a square matrix of SGA genetic interaction scores where every gene A is present both as query (AB) and as array (BA), the total number of interactions at a certain cutoff is the sum of interactions called by AB pairs only, BA pairs only and those called by both. The interactions called by both AB and BA ($N_{overlap}$) are expected to be a mixture of true positives and false positives. The true positive portion of $N_{overlap}$ depends on the total number of interactions called by AB (N_{AB}), the probability of them being true positives (precision, p) and the probability of them not being missed by BA (sensitivity = 1 – *FNR*). The false positive portion of $N_{overlap}$ depends on the total number of interactions called by AB (N_{AB}), the probability of them being false ($1 - p$) and the probability of them being incorrectly called positive by BA (specificity = 1 – *TNR*).

Overall, the number of interactions called by both AB and BA is given by:

$$N_{overlap} = N_{overlap\_true} + N_{overlap\_false} = pN_{AB}(1-FNR) + (1-p)N_{AB}(1-TNR)$$

Given that $N_{overlap}$, N_{AB} and FNR are known, and $1-TNR$ can be at most equal to the total density of interactions in the matrix (adopting the conservative assumption that all the reported interactions are false), precision can be calculated as follows (Figure 2.9, precision):

$$p = \frac{N_{overlap} - N_{AB}(1-TNR)}{N_{AB}(1-FNR) - N_{AB}(1-TNR)}$$

Table 2.1. Sensitivity and precision of SGA genetic interaction scores.

		Lenient cutoff $p < 0.05$	Intermediate cutoff $\|\varepsilon\| > 0.08$, $p < 0.05$	Stringent cutoff $\varepsilon < -0.12$, $p < 0.05$ $\varepsilon > 0.16$, $p < 0.05$
Negative interactions	Number of interactions	366,085	108,417	58,508
	Sensitivity	0.41	0.35	0.28
	Precision	0.27	0.63	0.89
Positive interactions	Number of interactions	323,935	59,887	6,185
	Sensitivity	0.26	0.18	0.05
	Precision	0.20	0.59	~1

2.10.4 Comparative analysis of function prediction capacity

The input to the GeneMANIA algorithm is one or more functional association networks, i.e. networks where nodes represent genes or proteins, and edges are weighted according to evidence of shared function between the connected nodes (Mostafavi *et al.*, 2008). We obtained 23 large-scale datasets from BioGRID (Stark *et al.*, 2011) or the Supplementary da-

ta of the corresponding publications (Table 2.2, below), and used them to construct dataset-specific and data type-specific functional association networks (Figure 2.11). For protein-protein and genetic interaction data, we represented shared function using direct binary interactions, as well as similarity of interaction profiles as defined by Pearson correlation coefficients (PCC) or other similarity measures. For gene expression and localization data, only profile similarity was used. All correlation-based networks were sparsified by setting to zero all but the top 100 highest connections for each gene.

Table 2.2. List of datasets used in comparative function prediction analysis.

Data type	Dataset name	Reference	Source
AP/MS	Gavin 2006	(Gavin *et al.*, 2006)	BioGRID
	Krogan 2006	(Krogan *et al.*, 2006)	BioGRID
	Ho 2002	(Ho *et al.*, 2002)	BioGRID
Y2H	Yu 2008	(Yu *et al.*, 2008)	BioGRID
	Y2H Union	(Yu *et al.*, 2008)	BioGRID
PCA	Tarassov 2008	(Tarassov *et al.*, 2008)	BioGRID
Other PPI	Miller 2005	(Miller *et al.*, 2005)	BioGRID
Gene expression	Chua 2006	(Chua *et al.*, 2006)	Supplementary data
	Spellman 1998	(Spellman *et al.*, 1998)	Supplementary data
	Gasch 2000	(Gasch *et al.*, 2000)	Supplementary data
	Hughes 2000	(Hughes *et al.*, 2000)	Supplementary data
	Roberts 2000	(Roberts *et al.*, 2000)	Supplementary data
	Yvert 2003	(Yvert *et al.*, 2003)	Supplementary data
Protein localization	Huh 2003	(Huh *et al.*, 2003)	Supplementary data
Genetic interactions	Tong 2004	(Tong *et al.*, 2004)	BioGRID
	Schuldiner 2005	(Schuldiner *et al.*, 2005)	BioGRID
	Pan 2006	(Pan *et al.*, 2006)	BioGRID
	Collins 2007	(Collins *et al.*, 2007)	BioGRID
	Lin 2008	(Lin *et al.*, 2008)	BioGRID
	Wilmes 2008	(Wilmes *et al.*, 2008)	Supplementary data
	Fiedler 2009	(Fiedler *et al.*, 2009)	Supplementary data
Phosphorylation	Ptacek 2005	(Ptacek *et al.*, 2005)	Supplementary data
Chemical genomics	McClellan 2007	(McClellan *et al.*, 2007)	Supplementary data

2.10.5 Data access

Matlab source code for the SGA score algorithm

https://research.cs.umn.edu/csbio/SGAScore/Supplementary_software_1_Matlab_source_code.zip

Supplementary software 1, (Baryshnikova *et al.*, 2010b)

Raw colony size data

https://research.cs.umn.edu/csbio/SGAScore/Supplementary_data_4_Raw_sgadata_fullgenome_101209.txt.gz

Supplementary data file 4, (Baryshnikova *et al.*, 2010b)

Single mutant fitness standard

https://research.cs.umn.edu/csbio/SGAScore/Supplementary_data_1_SMF_standard_100209.xls

Supplementary data file 1, (Baryshnikova *et al.*, 2010b)

Genetic interaction scores

http://drygin.ccbr.utoronto.ca/~costanzo2009/sgadata_costanzo2009_rawdata_101120.txt.gz

Supplementary data file 1, (Costanzo *et al.*, 2010)

3 The genetic landscape of a cell

Portions of this chapter have been reprinted or adapted from:

- Baryshnikova A.*, Costanzo M.*, Kim Y., Ding H., Koh J., Toufighi K., Youn J. Y., Ou J., San Luis B. J., Bandyopadhyay S., Hibbs M., Hess D., Gingras A. C., Bader G. D., Troyanskaya O. G., Brown G. W., Andrews B., Boone C., Myers C. L., "Quantitative analysis of fitness and genetic interactions in yeast on a genome scale", Nature Methods, 7(12):1017–1024, 2010 (Baryshnikova *et al.*, 2010b)
- Costanzo M.*, Baryshnikova A.*, Bellay J., Kim Y., Spear E. D., Sevier C. S., Ding H., Koh J. L., Toufighi K., Mostafavi S., Prinz J., St Onge R. P., VanderSluis B., Makhnevych T., Vizeacoumar F. J., Alizadeh S., Bahr S., Brost R. L., Chen Y., Cokol M., Deshpande R., Li Z., Lin Z. Y., Liang W., Marback M., Paw J., San Luis B. J., Shuteriqi E., Tong A. H., van Dyk N., Wallace I. M., Whitney J. A., Weirauch M. T., Zhong G., Zhu H., Houry W. A., Brudno M., Ragibizadeh S., Papp B., Pal C., Roth F. P., Giaever G., Nislow C., Troyanskaya O. G., Bussey H., Bader G. D., Gingras A. C., Morris Q. D., Kim P. M., Kaiser C. A., Myers C. L., Andrews B. J., Boone C., "The genetic landscape of a cell", Science, 327(5964):425-431, 2010 (Costanzo *et al.*, 2010)

* These authors contributed equally to the publication.

I generated the SGA dataset by performing quality control procedures on raw colony size data and applying the SGA score to measure genetic interactions. I produced the correlation-based genetic interaction network and performed the downstream analysis of functional clusters. I integrated genetic and chemical-genetic interaction networks and identified the novel drug target.

Jeremy Bellay performed the comparison of genetic interaction degree against physiological and evolutionary properties, and computed the frequency of interactions between biological processes.

Judice Koh developed and ran the GINECA clustering algorithm.

Yungil Kim analyzed the genetic suppression network between protein complexes.

Michael Costanzo designed and coordinated all experimental work which was performed by technicians in the Boone lab and collaborators. Follow-up experiments were conducted in the laboratories of Chris Kaiser, Anne-Claude Gingras, Guri Giaever and Corey Nislow.

3.1 Abstract

A genome-scale genetic interaction map was constructed by examining ~5.4 million gene pairs in *Saccharomyces cerevisiae* using the Synthetic Genetic Array (SGA) technology. A network based on quantitative genetic interaction profiles for ~75% of yeast genes revealed a functional map of the cell, in which genes of similar biological processes clustered together in coherent subsets. Highly similar genetic interaction profiles delineated specific pathways and defined new functional roles for uncharacterized genes. The global network identified high level functional cross-connections between bioprocesses and mapped the wiring diagram of pleiotropy. Genetic interaction degree correlated with a number of different gene attributes, which may be informative about genetic network hubs in other organisms. Comparison of physical and genetic interaction networks showed that, in contrast to previous expectation, positive genetic interactions often connect functionally distinct protein complexes and form a network of genetic suppression among loss-of-function alleles. Finally, extensive and unbiased mapping of genetic interactions provided a key for interpretation of chemical-genetic interactions and enabled identification of novel drug targets.

3.2 Introduction

The relationship between genotype and phenotype of an organism is governed by a myriad of genetic interactions (Chapter 1). Although a complex genetic landscape has long been anticipated (Waddington, 1957), exploration of genetic interactions on a genome-wide level has been limited. Systematic deletion analysis in the budding yeast *Saccharomyces cerevisiae* demonstrates that the majority of its ~6,000 genes are individually dispensable, with only a relatively small subset (~20%) required for viability (Giaever *et al.*, 2002). This finding suggests that cells have evolved extensive buffering against genetic perturbations (Hartwell, 2004), which can be uncovered by genome-scale screens for genetic interactions. In yeast, synthetic genetic array (SGA) methodology enables the systematic mapping of genetic interactions through an automated form of genetic analysis that produces high-density arrays of double mutants (Chapter 1.3.2). In this chapter, I report the systematic application of the SGA methodology to ~30% of the yeast genome and the construction of the first genome-scale functionally unbiased genetic interaction map for a eukaryotic cell.

In our definition, a genetic interaction occurs when the fitness a double mutant shows a significant deviation from the multiplicative combination of the two single mutant fitnesses (Chapter 1.1). Negative interactions refer to a more severe fitness defect than expected, with the extreme case being synthetic lethality. Positive genetic interactions refer to double mutants with a less severe fitness defect than expected. To quantitatively score genetic interactions in large-scale SGA screens, I developed a model to estimate fitness defects directly from double mutant colony sizes (Chapter 2.4). We screened 1,712 *S. cerevisiae* query genes, including 334 conditional or hypomorphic alleles of essential genes, for a total of ~5.4 mil-

lion gene pairs spanning all biological processes. Data evaluation, using several different approaches, indicated that the identified genetic interactions were reproducible and functionally informative (Chapters 2.6–2.7).

3.3 A functional map of the cell

Genes acting within the same pathway or biological process tend to share common sets of genetic interactions, i.e. they have similar genetic interaction profiles (Chapter 1.2). I exploited this property to construct a global network, in which nodes represent genes and edges connect gene pairs with similar genetic interaction profiles (Figure 3.1). This network highlights the genetic relationships between diverse biological processes and reveals the inherent functional organization of the cell. Genes displaying tightly correlated genetic interaction profiles form discernible clusters corresponding to distinct bioprocesses, and the relative distance between them appears to reflect shared functionality (Figure 3.1). For example, the role of the microtubule cytoskeleton in bridging nuclear chromosomal- and actin cytoskeleton-based functions is illustrated by close proximity and relative positioning of gene clusters involved in cell polarity and morphogenesis, mitosis and chromosome segregation, and DNA replication and repair (Figure 3.1).

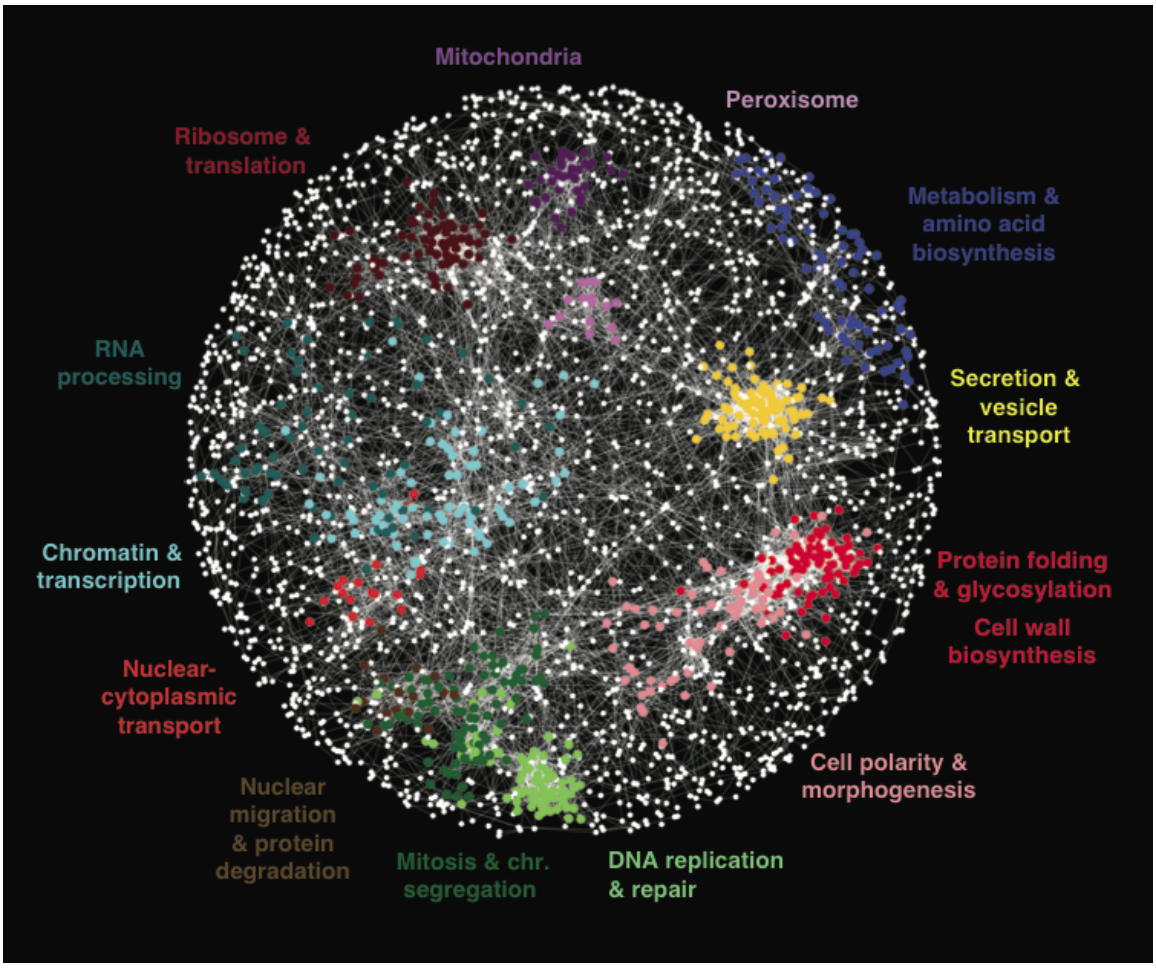


Figure 3.1. A correlation-based network connecting genes with similar genetic interaction profiles.

Genetic profile similarities were measured for all gene pairs by computing Pearson correlation coefficients (PCCs) from the complete genetic interaction matrix. Gene pairs whose profile similarity exceeded a PCC > 0.2 threshold were connected in the network and laid out using an edge-weighted spring-embedded network layout algorithm implemented in Cytoscape (Smoot *et al.*, 2011). Genes sharing similar patterns of genetic interactions are proximal to each other, while less similar genes are positioned farther apart. Colored regions indicate sets of genes identified as described in Materials & methods 3.13.1, and enriched for GO biological processes summarized by the indicated terms.

Despite screening only ~30% of all yeast genes as queries, we recovered genetic interactions for ~75% of the genome because nearly all nonessential genes are present on the array and obtained partial genetic interaction profiles across 1,712 queries.

3.4 Predicting function and relations

The genetic interaction network contains functional information at multiple levels of resolution. At the general level, it reveals functional connections between broad biological processes (Figure 3.1). At a higher level of resolution, the network dissects biological processes into more specific genetic cohorts, such as individual molecular pathways and protein complexes (Figure 3.2A). Even further, direct positive and negative genetic interactions reveal the connections that underlie a gene's functional clustering within a given cohort, and relate the gene to other members of the same biological pathway or protein complex (Figure 3.2B–D). As predicted previously (Segre *et al.*, 2005), connections within and between different pathways and protein complexes were often monochromatic, i.e. composed almost exclusively by a single type of genetic interaction, either all negative or all positive.

I used similarity of genetic interaction profiles to predict function for several uncharacterized genes (Figure 3.3). *PAR32*, *ECM30* and *UBP15* had interaction profiles similar to those of members of the Gap1-sorting module (Figure 3.2B), and consistent with a role in this process, deletion mutants for all three genes led to defects in Gap1 sorting (Figure 3.3A) and transport (Figure 3.3B). Additional experimental results (Huber *et al.*, 2009; Costanzo *et al.*, 2010) suggest that Par32 may function in target of rapamycin (TOR)-dependent regulation of the Gln3, Gat1, Rtg1, and Rtg3 transcription factors (Chen and Kaiser, 2003), whereas Ecm30 forms a stoichiometric complex with the Ubp15 ubiquitin protease that may modulate Gap1 localization, perhaps by controlling its ubiquitination state.

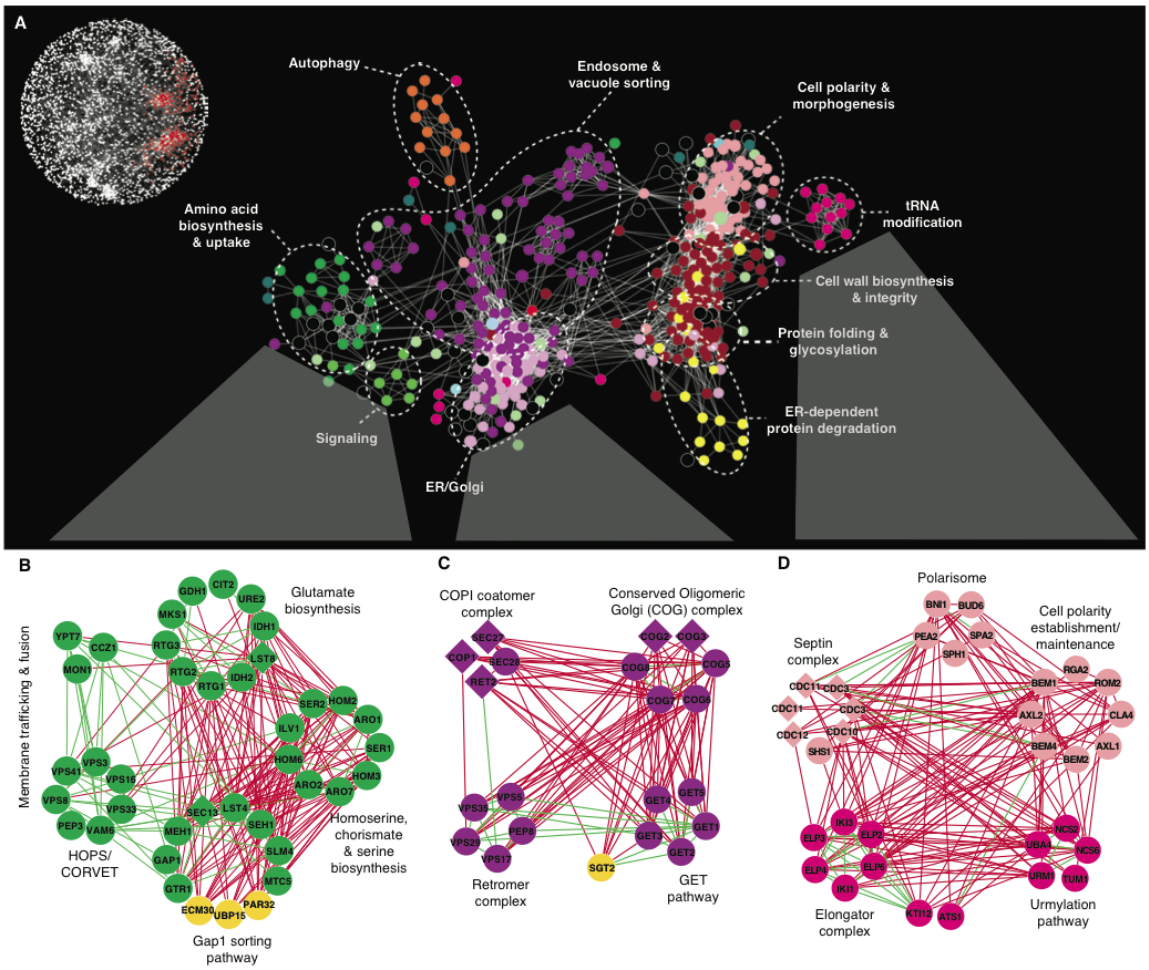


Figure 3.2. Magnification of the functional map better resolves cellular processes.

(**A**) A subnetwork corresponding to a region of the global map described in Figure 3.1 is indicated in red (inset). Node color corresponds to a specific biological process: dark green, amino acid biosynthesis and uptake; light green, signaling; light purple, ER-Golgi; dark purple, endosome and vacuole sorting; yellow, ER-dependent protein degradation; red, protein folding and glycosylation, cell wall biosynthesis and integrity; fuchsia, tRNA modification; pink, cell polarity and morphogenesis; orange, autophagy; black, uncharacterized. Individual genetic interactions contributing to genetic profiles revealed by (**A**) are illustrated for three specific subnetworks in (**B**) to (**D**).

(**B** to **D**) Subsets of genes belonging to amino acid biosynthesis and uptake, ER-Golgi, and tRNA modification regions of the network were selected, and, in some cases, additional genes were included from the complete network shown in Figure 3.1. Nodes are grouped according to profile similarity, and edges represent negative (red) and positive (green) genetic interactions ($|\varepsilon| > 0.08$, $p < 0.05$). Nonessential (circles) and essential (diamonds) genes are colored according to the biological process indicated in (**A**), and uncharacterized genes are depicted in yellow.

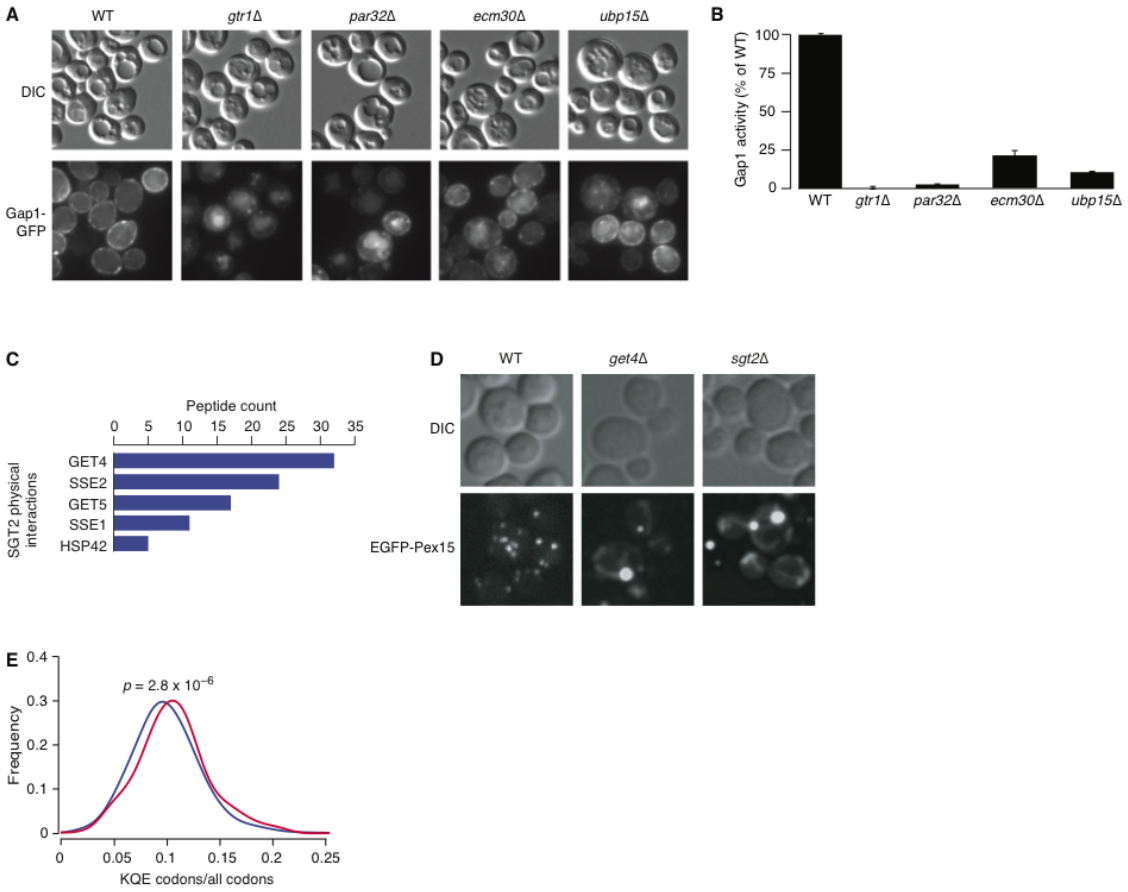


Figure 3.3. Genetic interaction profiles reflect diverse types of functional relationship.

(**A**–**B**) *PAR32*, *ECM30*, and *UBP15* are required for plasma membrane localization (**A**) and activity (**B**) of the Gap1 amino acid permease (Materials & methods 3.13.3). DIC, differential interference contrast; GFP, green fluorescent protein.
(**C**) Stg2 physically interacts with components of the GET pathway and members of the Hsp70 chaperone family. Proteins identified with high confidence as specific interactors for tandem affinity purification (TAP)-tagged Sgt2 (Sgt2-TAP) are shown in decreasing order of spectral counts (Materials & methods 3.13.4).
(**D**) Similar to GET pathway mutants, *SGT2* is required for proper Pex15 localization (Materials & methods 3.13.5).
(**E**) Distribution of Elp and Urm modified codon usage among synthetic sick or lethal interaction partners (Materials & methods 3.13.6). The fraction of Elp and Urm modified codons (lysine, glutamine, and glutamic acid) relative to all codons was measured for all negative interactors of the Elp or Urm complex members (red) relative to the background usage of all genes (blue).

In another example, similarity of genetic interaction profiles suggested a strong functional relation between the GET pathway and the poorly characterized gene, *SGT2* (Figure 3.2B). Consistent with a role in endoplasmic reticulum (ER)-dependent membrane targeting

(Jonikas *et al.*, 2009) or protein folding (Metzger and Michaelis, 2009), we found that Sgt2 physically interacts with Get4, Get5 and heat shock 70 (Hsp70) protein family members (Figure 3.3C). In addition, similarly to GET pathway mutants (Jonikas *et al.*, 2009), deletion of *SGT2* results in mislocalization of the tail-anchored protein, Pex15 (Figure 3.3D).

3.5 Deciphering complex regulatory relations from the global genetic network

Because the global genetic interaction map represents a broad functional survey, it should provide insights into the regulatory wiring diagram of the cell. For example, synthetic lethal interactions between members of the elongator (Elp) complex and those of the urmylation (Urm) pathway suggested that the Urm pathway collaborates with the Elp complex in the modification of specific transfer RNAs (tRNAs) (Figure 3.2D) (Leidel *et al.*, 2009). In addition to their synthetic lethal relationship, Elp and Urm pathway genes shared highly similar genetic interaction profiles. Notably, these interactions were enriched for cell polarity and secretion genes ($p < 10^{-3}$), which reflects a specific cell polarity defect associated with Elp pathway mutants (Rahl *et al.*, 2005).

The elongator tRNA modification machinery has been postulated either to broadly affect the translation of a suite of mRNAs whose genes have cell polarity roles or to selectively influence the activity of a key polarity regulatory gene (Esberg *et al.*, 2006). We were intrigued by the finding that the subset of Elp-Urm negative interactors, as well as cell polarity and secretion genes in general, encode proteins that are significantly enriched for the amino acids that charge Elp- and Urm-modified tRNAs (Figure 3.3F). These findings suggest that

Elp and Urm pathways may be biased toward the regulation of a functionally specific subset of cellular proteins. *ELP1* is a highly conserved gene whose human ortholog, inhibitor of kappa light polypeptide gene enhancer in B cells, kinase complex-associated protein (IKBKAP), is associated with a neurological disorder, familial dysautonomia, which leads to disruption of cytoskeletal organization when mutated (Johansen *et al.*, 2008; Naumanen *et al.*, 2008). Thus it is possible that disease manifestation may involve impaired IKBKAP-dependent translation of a set of human genes belonging to a specific functional group.

3.6 Genetic network connectivity

Consistent with the degree distribution of other biological networks, the majority of genes had few genetic interactions, whereas a small number were highly connected and served as network hubs (Figure 3.4A). Biases in interaction number were supplemented by biases in interaction type. About 2% of array genes exhibited preferentially negative genetic interactions, presenting a positive to negative degree ratio lower than 1:8. Conversely, ~1% of array genes were biased towards positive genetic interactions and showed a positive to negative degree ratio greater than 4:1 (Figure 3.4B). Genes displaying this distinct behavior were also functionally distinct. The set of negatively biased genes included preferentially those required for normal cell cycle progression ($p < 10^{-8}$), reflecting the central role of checkpoints in maintaining viability in dividing cells. Predominantly positive interactions were indicative of genes involved in translation, ribosomal RNA processing, and mRNA decay ($p < 10^{-5}$), which may suggest that defects in the translation machinery somehow mask phenotypes that would otherwise be expressed in normal cells.

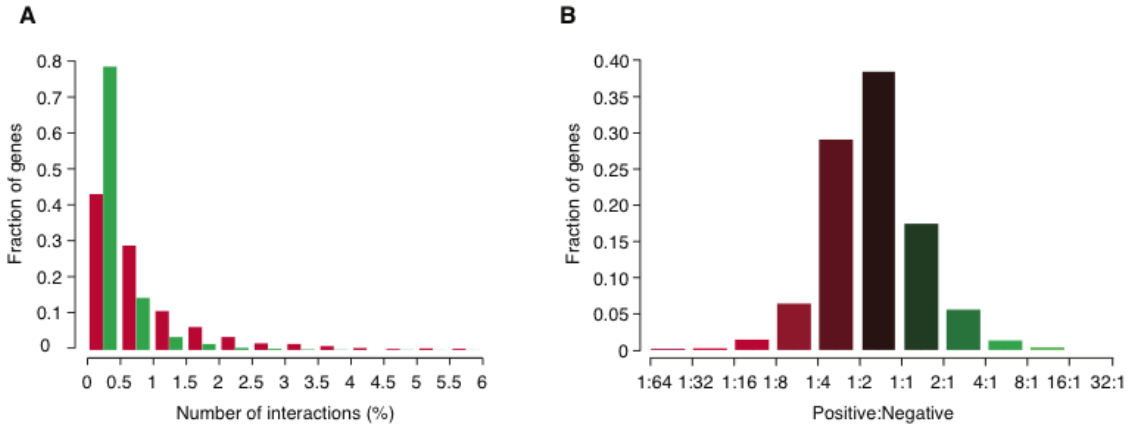


Figure 3.4. Genetic network connectivity.

(**A**) The distribution of genetic interaction network degree for negative (red) and positive (green) interactions involving query genes.
(**B**) The distribution of the ratio of positive to negative interactions for query genes.

Genetic interaction hubs showed a clear association with several fundamental physiological and evolutionary properties (Figure 3.5A), which may be predictive of genetic interactions in other organisms. In particular, we uncovered a strong correlation between genetic interaction degree and single mutant fitness ($R = 0.73$). Single mutants with increasingly severe fitness defects tended to exhibit an increased number of both negative and positive interactions (Figure 3.5B–C). This relation was also observed for essential genes where the average number of interactions involving a temperature-sensitive mutant allele was inversely proportional to allele fitness at a given semipermissive temperature (Figure 3.5D). The increased connectivity of genes with fitness defects was not due to nonspecific interactions derived from a generally compromised cell or experimental noise: interactions involving these genes overlap with known functional relationships just as frequently as other interactions (Costanzo *et al.*, 2010).

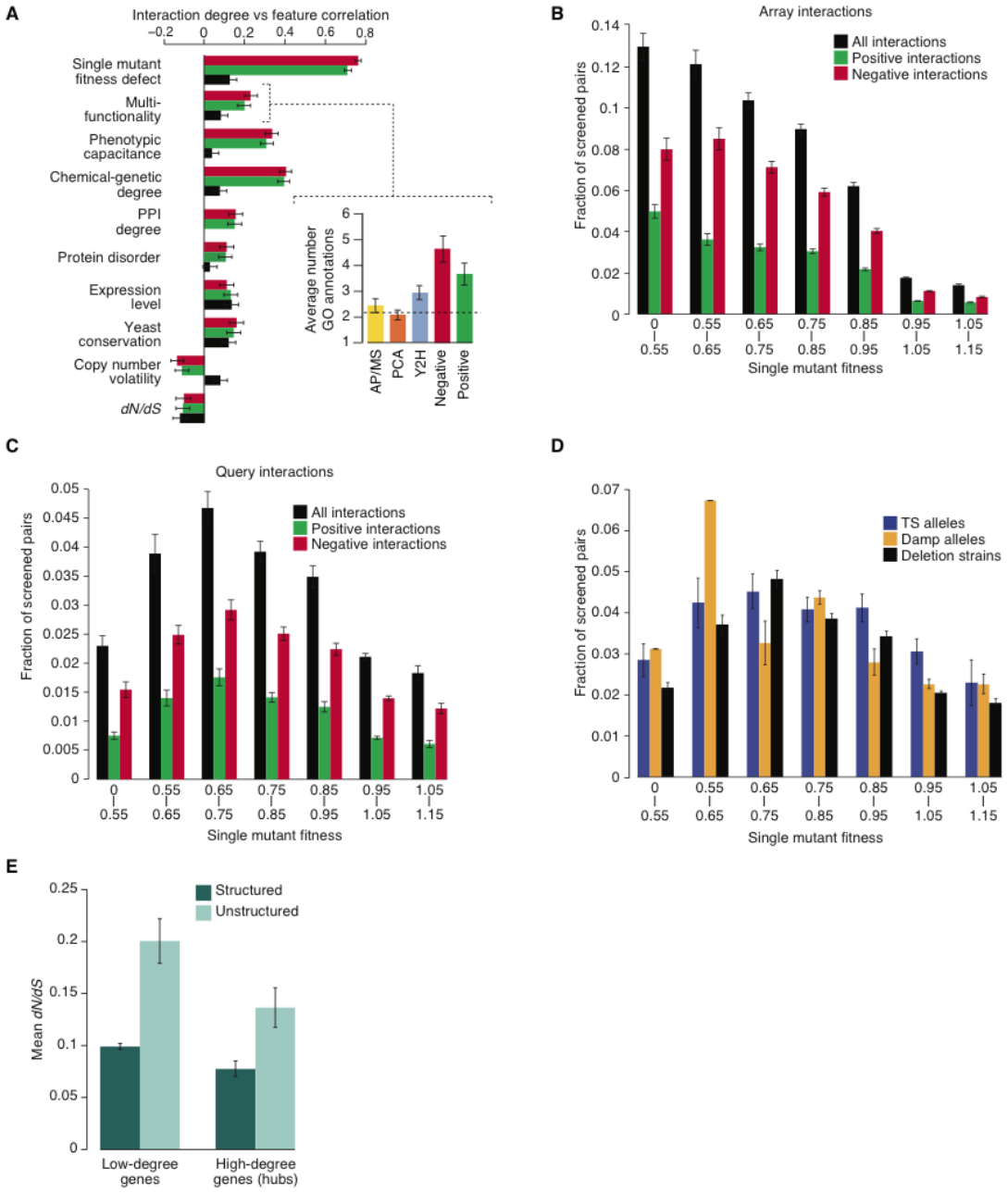


Figure 3.5. Genetic interaction degree, fitness, multifunctionality, and pleiotropy.

(**A**) Pearson correlation between interaction degree and a number of physiological and evolutionary features was measured for positive (green), negative (red) and protein-protein (black) interactions. Chemical-genetic degree refers to the number of chemical perturbations to which a gene exhibits hypersensitivity. (**Inset**) The relation to gene multifunctionality for each of the interaction datasets is illustrated by measuring the average number of annotations to specific biological process GO terms for the top 1% highest degree genes for each interaction type.

(**B–C**) Relationship between genetic interaction degree and single mutant fitness. 3,885 array genes (**B**) and 1,712 query genes (**C**) were binned into seven groups on the basis of their single mutant fitness. For each bin, the fraction of tested gene pairs showing a genetic interaction at the intermediate confidence threshold

($|\varepsilon| > 0.08$, $p < 0.05$) was measured for positive (green), negative (red) and all (black) genetic interactions. (**D**) The relationship between genetic interaction degree and single mutant fitness was examined for different types of genetic perturbations. 1,712 query genes were separated by the type of genetic perturbation (deletion, temperature-sensitive allele (TS), or DAmP allele) and genetic interaction degree was measured for each single mutant fitness bin. Genetic interaction degree of essential genes correlated with allele fitness measured at semi-permissive temperature (26ºC).

In addition to the correlation with single mutant fitness defects, the number of genetic interactions for a particular hub was significantly correlated with the number of biological functions annotated to that gene (Figure 3.5A). The connection between hubness and multi-functionality was further supported by an increased phenotypic capacitance, which measures the number of different morphological phenotypes linked to a specific gene (Figure 3.5A) (Levy and Siegal, 2008). This relation suggests that genetic network hubs play key roles in the integration and execution of multiple morphogenic programs.

It is important to note that these correlations persisted after we controlled for fitness defects of single mutants (Costanzo *et al.*, 2010). Furthermore, they reveal characteristics that distinguish genetic network hubs from hubs on the physical interaction network (Figure 3.5A). Notably, the correlation to both fitness and multifunctionality was several fold stronger for genetic interaction degree than for physical interaction degree (Figure 3.5A). This likely reflects the ability of genetic perturbation analysis to identify broad phenotypic connections that cannot be captured in networks subject to physical constraints and suggests that large-scale genetic interaction networks will be of wide utility for defining the functional wiring diagrams of cells and organisms.

Although there are several distinguishing characteristics of genetic interaction hubs, we measured a significant correlation ($R \sim 0.2$) between the genetic and physical interaction de-

gree for any given gene (Figure 3.5A). Similar to protein-protein interaction hubs (Pal *et al.*, 2001; Fraser *et al.*, 2003; Kim *et al.*, 2006), we found that genetic hubs tend to be expressed at higher mRNA levels (Figure 3.5A). Based on a comparison of whole-genome sequences of 23 different Ascomycota fungi species, genetic interaction degree correlated positively with gene conservation and negatively with copy number volatility, indicating that hubs tend to be lost or duplicated less frequently. Genetic interaction hubs also evolved more slowly (as measured by the *dN*/*dS* ratio) than genes with fewer interactions (Figure 3.5A), which suggests that hubs generally tend to be evolutionary constrained. Only a subset of genetic interaction hubs seemed to behave differently: proteins exhibiting a higher level of native disorder appear to evolve faster than their ordered counterparts (Figure 3.5E), suggesting that they may represent a distinct class of genetic interaction hubs (Bellay *et al.*, 2011).

3.7 Distribution of genetic interactions by bioprocess

We assessed the distribution of both negative and positive genetic interactions across different cellular processes (Figure 3.6A–B). The analysis identified functions which are enriched or depleted for genetic interactions relative to the expected frequency of a random gene set. As observed previously (Tong *et al.*, 2001), genes involved in the same biological process were enriched for negative interactions (Figure 3.6A). However, we also observed numerous genetic interactions connecting across bioprocesses. Specifically, genes involved in chromatin, transcription, ER-Golgi transport and Golgi-endosome transport showed a significant number of interactions that bridge diverse functions, suggesting that many of these genes are highly interconnected or multifunctional. Indeed, these bioprocess-level findings

concur with individual gene analyses, which indicate that genes involved in processes related to chromatin structure and transcription, as well as secretion and vesicle transport, were among the most highly connected genes in our network ($p < 10^{-14}$ and $p < 10^{-9}$, respectively). The central role for chromatin- and transcription-related processes identified in the yeast genetic network is consistent with large-scale genetic network mapping in *Caenorhabditis elegans* (Lehner *et al.*, 2006), and the bridging function for secretory pathway genes emphasizes their role as communication conduits for the cell.

Particular gene sets, such as meiosis, drug or ion transport, and metabolism or mitochondrial genes, showed a reduced number of genetic interactions (Figure 3.6A–B). This may arise because some processes are more buffered than others and require more complex genetic analysis to uncover their interactions, whereas other biological functions may only be needed under certain environmental conditions.

In contrast to genetic interactions, protein-protein interactions connect relatively fewer bioprocesses (Figure 3.6C), thus suggesting that, although highly informative of local pathway architecture, they fail to provide a complete picture of multifunctionality or interconnections between cellular processes.

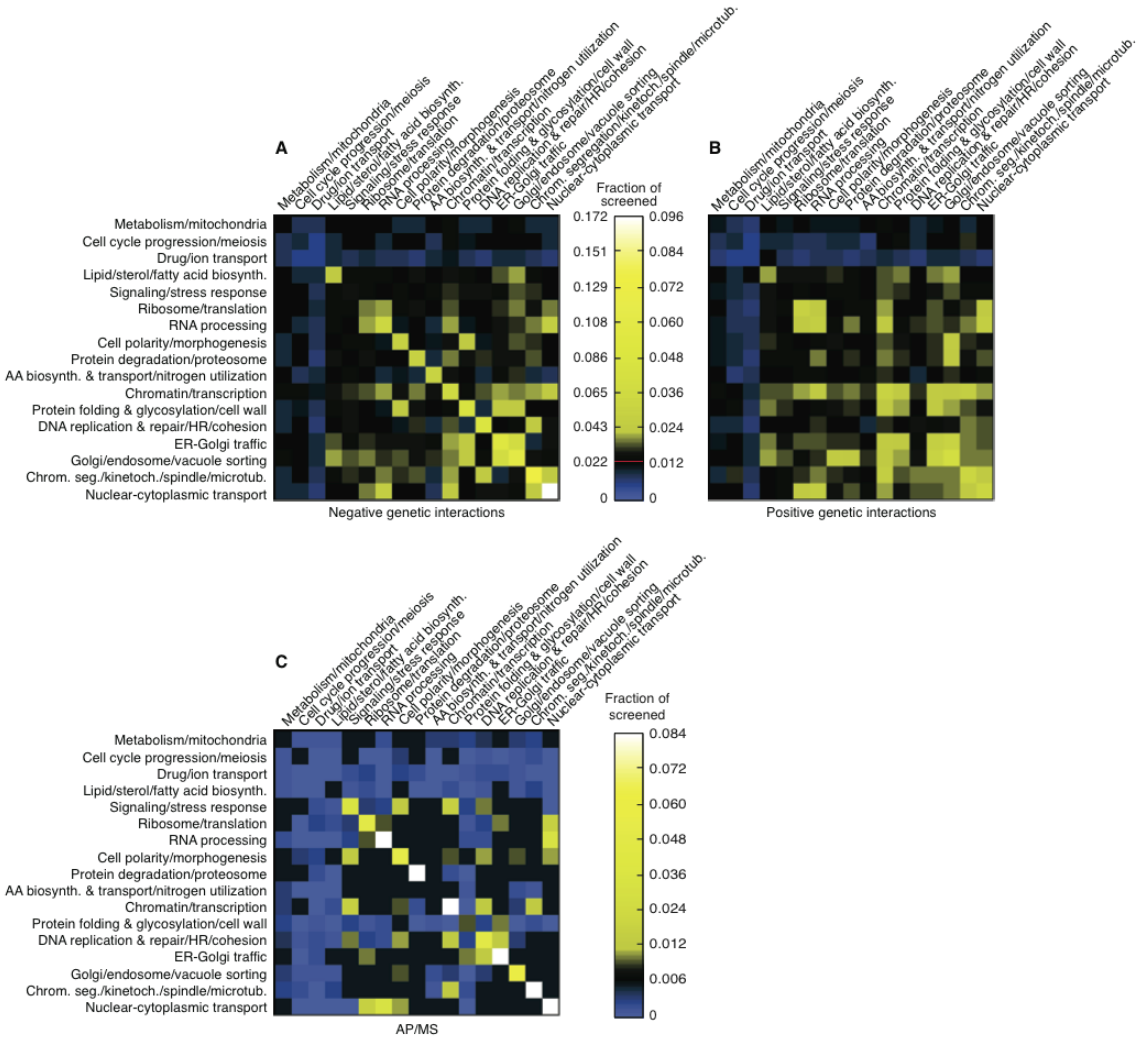


Figure 3.6. Distribution of genetic and physical interactions by bioprocess.

Frequency of negative (**A**) and positive (**B**) genetic interactions, as well as protein-protein interactions (**C**) was calculated within and between 17 broadly defined functional categories (Materials & methods 3.13.9). The color of each process-process element reflects interaction frequency relative to random expectation (blue, below random; yellow, above random; black, statistically indistinguishable from random). The diagonal represents within-process interactions. The red line in the color scale bar indicates random background.

Because variation was observed in the average number of genetic interactions across different bioprocesses, we tested whether gene-specific properties (Figure 3.5A) were predictive of this variation. For example, we found that gene duplicates exhibited fewer interactions than singleton genes (Costanzo *et al.*, 2010; VanderSluis *et al.*, 2010), and we asked if bio-

processes with relatively few genetic interactions could be explained by a high percentage of duplicated genes. An analysis of covariance (ANCOVA; Materials & methods 3.13.11) showed than a linear model including the gene-specific properties predictive of genetic interaction hubs was sufficient to explain the number of genetic interactions for the majority of bioprocesses (12 out of 17 for negative interactions, and 13 out of 17 for positive interactions; Figure 3.7). For example, the relatively few genetic interactions observed for genes involved in drug and ion transport are explained by a combination of high gene duplication rate (~50–60%) and copy number volatility. This is consistent with the tendency of genes encoding protein pumps to undergo numerous duplication events (Dunham *et al.*, 2002) and confirms that extensive redundancy associated with large gene families complicates the identification of digenic interactions.

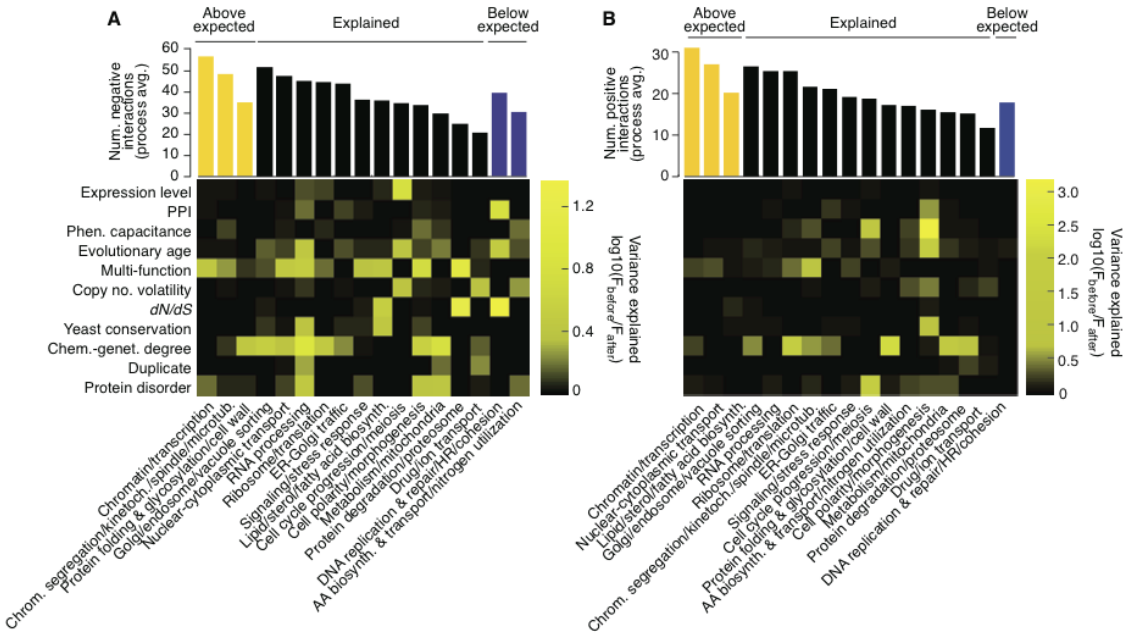


Figure 3.7. Gene-specific factors explaining the variation in genetic interaction number across biological processes.

ANCOVA analysis was used to assess the significance of gene-specific factors (Figure 3.5A) at predicting the number of negative (**A**) and positive (**B**) genetic interactions observed for a specific biological process (Materi-

als & methods 3.13.9). The top panel shows the average number of interactions within each process. The color of the bars indicates whether the factors listed on the y-axis explain the interaction degree of the process (black) or whether the interaction degree is significantly above (yellow, $p < 0.05$) or below (blue, $p < 0.05$) expectation. For each process, the proportion of variance explained by each of the gene-specific factors is indicated by the corresponding column in the heatmap.

Three bioprocesses had significantly more negative interactions than predicted ($p < 0.05$; Figure 3.7A), including those that show functional enrichment for genetic interaction hubs (Figure 3.5A). Conversely, DNA replication and repair and amino acid biosynthesis showed significantly fewer negative interactions than predicted ($p < 0.05$; Figure 3.7A), suggesting that these genes are more buffered and inherently less connected on the digenic network or that more genetic interactions will surface when these genes are tested under different environmental conditions.

3.8 Overlap between genetic and protein-protein interaction network

We examined the relationship between quantitative genetic interactions and protein-protein interactions as defined by affinity purification-mass spectrometry (Gavin *et al.*, 2006; Krogan *et al.*, 2006), yeast two-hybrid protocol (Yu *et al.*, 2008), and protein-fragment complementation assay (PCA) (Tarassov *et al.*, 2008). We found that genetic interactions overlap 10–20% of protein-protein interaction pairs, depending on the physical interaction mapping technology (Figure 3.8). This overlap is significantly higher than expected randomly (~3%) and includes roughly an equivalent number of negative (~7%) and positive (~5%) genetic interactions.

In contrast, only a small fraction of gene pairs that show a negative or a positive genetic interaction are also physically linked (0.4% and 0.5%, respectively). This suggests that the vast majority of both positive and negative interactions occur between, rather than within, complexes and pathways, connecting those that presumably work together or buffer one another, respectively.

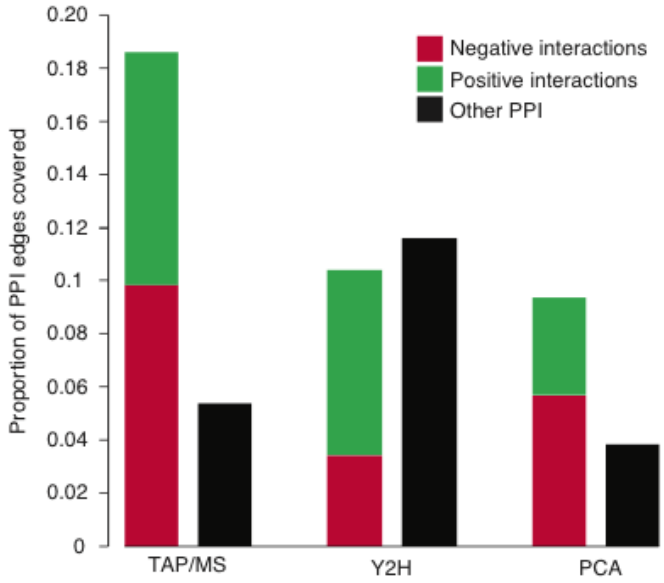


Figure 3.8. Overlap of genetic and physical interaction networks

For each of the three major mapping technologies (TAP/MS, yeast two-hybrid and PCA), we measured the fraction of physically interacting pairs that also showed a positive or a negative genetic interaction at the intermediate confidence threshold ($|\varepsilon| > 0.08$, $p < 0.05$). For comparison, the fraction of pairs identified by other protein-protein interaction technologies (excluding the dataset of interest) was also measured.

3.9 Analysis of genetic interactions within and between protein complexes

We measured the frequency of genetic interactions within and between 161 protein complexes (Materials & methods 3.13.12) for which more than one protein pair had been screened in our study. Consistent with a smaller-scale analysis (Bandyopadhyay *et al.*, 2008),

a large portion of these complexes (92 out of 161) were significantly enriched for genetic interactions among their members (hypergeometric test, $p < 0.05$). Interestingly, most complexes showed a strong preference for a single type of genetic interactions: 46% of the enriched complexes were connected by only positive interactions, while 37% by only negative interactions (Figure 3.9A–B). This finding confirms previous theoretical observations (Segre *et al.*, 2005), where similar groups of genes were described as "monochromatic".

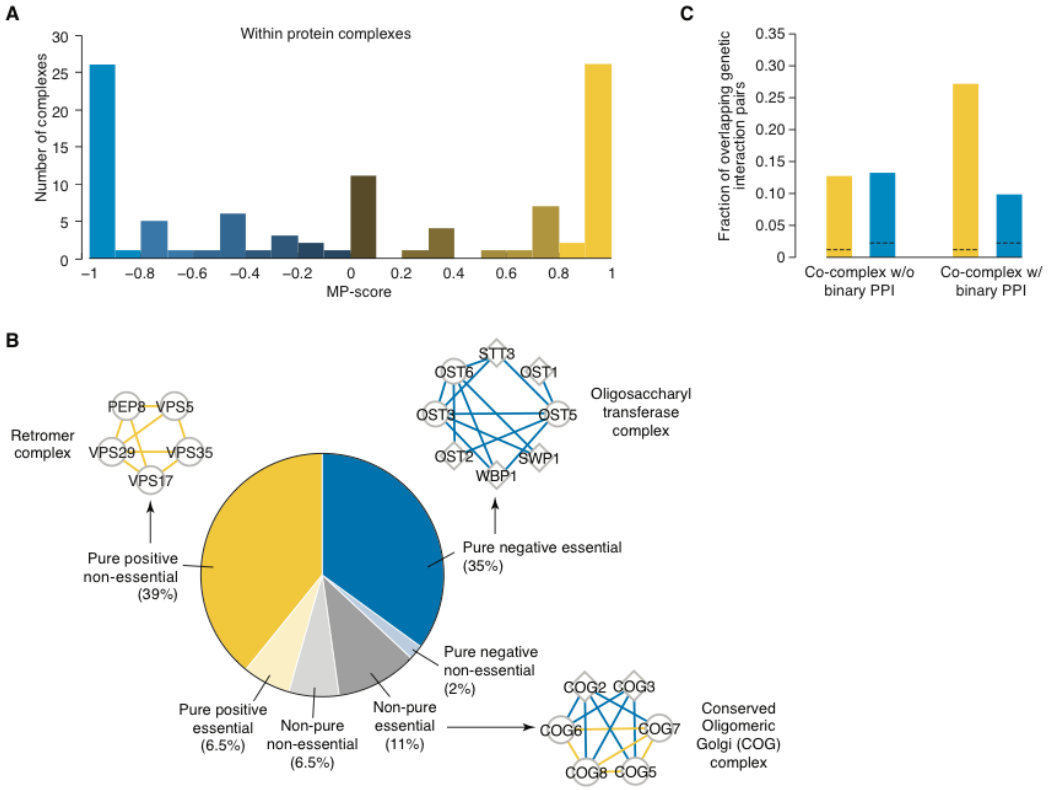


Figure 3.9. Analysis of genetic interactions within protein complexes.

(**A**) Distribution of protein complexes with respect to within-complex monochromatic purity scores (Materials & methods 3.13.12.3).
(**B**) Ninety-two protein complexes enriched for negative and/or positive genetic interactions were subdivided into monochromatic positive (yellow), monochromatic negative (blue) and mixed (gray) categories. Within each group, the fraction of complexes containing an essential gene is indicated with a lighter shade of the corresponding color. Three protein complexes representative of the major categories are shown on the sides. Nonessential genes are represented as circles and essential genes as diamonds.
(**C**) Genetic interaction frequency among co-complex members with and without a direct protein-protein interaction. The background rate of positive and negative interactions among random gene pairs is indicated by the black lines.

Virtually all (94%) monochromatic negative protein complexes contained at least one essential gene (Figure 3.9B), which is much higher than appreciated previously (Bandyopadhyay *et al.*, 2008). This observation suggests that essential complexes may contain partial internal redundancy, allowing the complex to tolerate the loss of a single nonessential component, whereas any additional perturbation disrupts the complex function and impairs cell growth. In contrast, only a relatively small fraction of monochromatic positive protein complexes (14%) contained an essential gene (Figure 3.9B).

Complexes with mixed interactions (both positive and negative; 17.5%) tended to be essential (Figure 3.9B) and, in some cases, had a well-defined structural organization. For example, the conserved oligomeric Golgi (COG) complex consists of an essential and a nonessential domains (Ungar *et al.*, 2006) which exhibit negative and positive genetic interactions, respectively (Figure 3.9B). By differentiating the two domains of the COG complex, the SGA score demonstrates its capacity to capture fine-grained genetic interactions, a task that had been so far accomplished only using high-resolution competitive-growth assays (Breslow *et al.*, 2008).

In addition to essentiality, the type of genetic interaction observed within a protein complex was also influenced by the presence of a direct physical interaction, as determined in the yeast two-hybrid assay (Figure 3.9C). Complex members with a direct physical link were nearly threefold more likely to share a positive genetic interaction. Conversely, complex members with no evidence of a direct physical interaction showed a modest preference for negative genetic interactions (Figure 3.9C).

We analyzed the genetic interactions occurring between members of different protein complexes (Figure 3.10A) and observed that monochromaticity of a given complex was predictive of its connectivity to other complexes (Figure 3.10B–C).

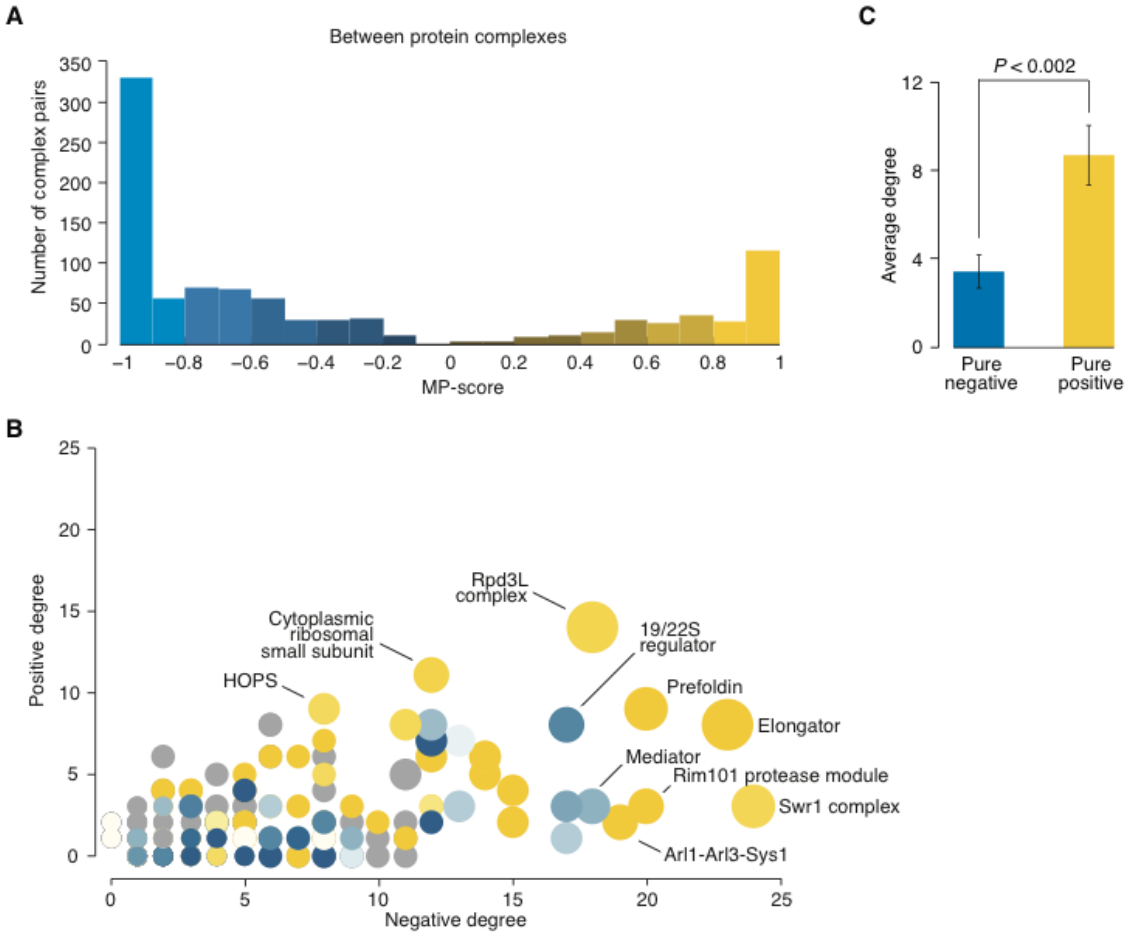


Figure 3.10. Analysis of genetic interactions between protein complexes.

(**A**) Distribution of protein complexes with respect to between-complex monochromatic purity scores (Materials & methods 3.13.12.3).
(**B**) Negative and positive degree was calculated as the number of negative and positive genetic interactions for each protein complex. Node color reflects the prevalence of positive (yellow) or negative (blue) genetic interactions within the complex. Gray nodes denote complexes with insufficient data to assess within-complex interactions. Node size indicates the number of proteins associated with the complex.
(**C**) Number of between-complex interactions was measured for monochromatic negative and monochromatic positive protein complexes. Error bars, s.e.m. ($n = 37$, positive, and $n = 25$, negative). Significance was calculated using a rank-sum test.

Nonessential members of monochromatic positive protein complexes had on average twice as many genetic interactions as nonessential genes within monochromatic negative complexes (Figure 3.10B–C). This behavior may reflect the internal redundancy of essential protein complexes: a single perturbation is less likely to compromise complex activity, and thus, these nonessential genes exhibit fewer genetic interactions. Conversely, deletion of a nonessential gene in a monochromatic positive complex may have a more severe impact on complex activity and, consequently, exhibited more genetic interactions with the rest of the genome.

3.10 Defining subclasses of positive genetic interactions

Even though both positive and negative genetic interactions are enriched among members of protein complexes, the vast majority of genetic interactions do not overlap physical interactions (Chapter 3.8). This observation strongly suggests that genetic interactions reflect functional rather than physical relationships between genes, a concept that is well established for negative genetic interactions (Tong *et al.*, 2004). However, the role of positive interactions outside individual pathways or protein complexes is much less understood.

We found 1,182 pairs of protein complexes that were significantly enriched for positive interactions connecting them (false discovery rate of 5%; Materials & methods 3.13.12.2). Among these complex pairs, we observed a surprisingly high incidence of genetic suppression, which is a particular type of positive genetic interaction indicative of mutations that rescue growth defects caused by other mutations (Hartman and Roth, 1973; Prelich, 1999). We identified genetic suppression among protein complexes by comparing single- and double-

mutant fitness of the corresponding mutants and by selecting cases where members of a particular complex consistently rescued growth defects associated with another complex (Materials & methods 3.13.13). By doing so, we constructed the first genome-wide network of suppression interactions between protein complexes (Figure 3.11).

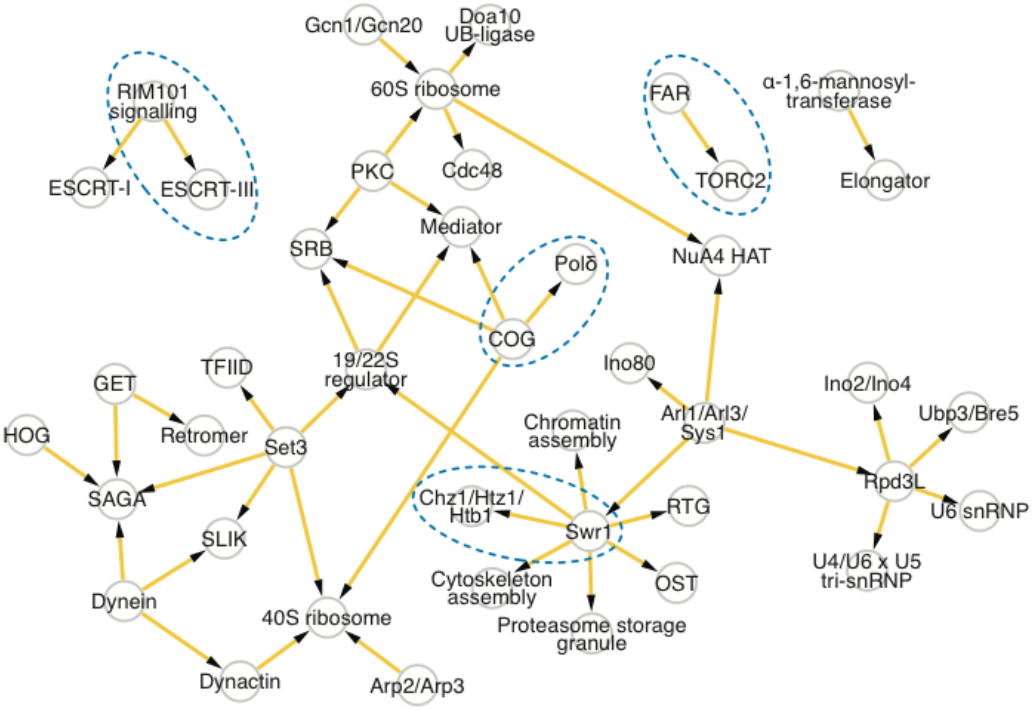


Figure 3.11. Cross-complex genetic suppression network.

A network illustrating suppression interactions between protein complexes. Arrows point to the complex whose fitness defect is suppressed.

We identified several cross-complex suppression interactions that reflect known functional relationships between protein complexes and are supported by independent experimental validation (Figure 3.12–Figure 3.15).

HTZ1 encodes the yeast histone variant H2A.Z which is deposited into chromatin by the Swr-C protein complex (Zlatanova and Thakar, 2008). Our SGA analysis confirmed the

results of a previous high-resolution competitive growth analysis (Breslow *et al.*, 2008) by showing that loss-of-function mutations involving members of the Swr-C complex suppress the growth defects associated with the loss of *HTZ1* (Figure 3.12). While the nature of this suppressive interaction is still unclear, it provides further evidence that the relationship between histones and their exchange factors is complex and still unresolved.

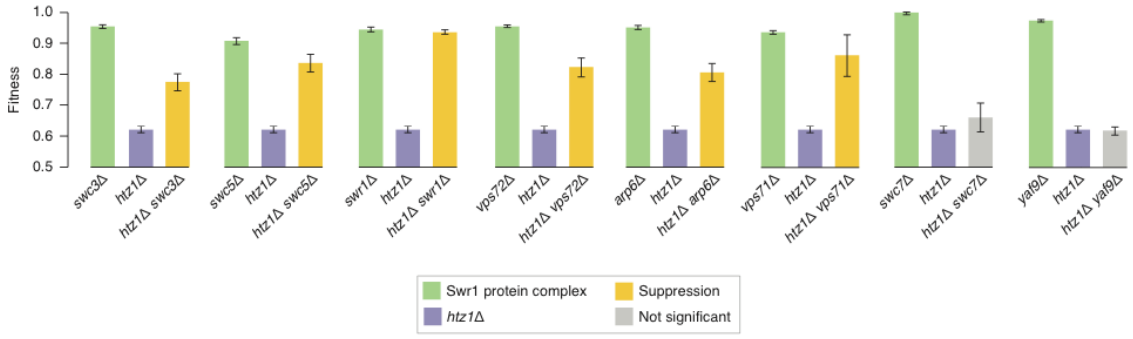


Figure 3.12. Genetic suppression between Swr1 protein complex and *htz1*Δ

Comparison of colony-size derived single and double mutant fitness measures suggests that mutations in genes encoding members of the Swr1 complex (*swc3*Δ, *swc5*Δ, *swr1*Δ, *vps72*Δ, *arp6*Δ and *vps71*Δ) suppress growth defects associated with deletion of *HTZ1*.

We also confirmed loss-of-function suppression interactions involving genes acting in the Rim101 signaling pathway (*RIM8*, *RIM9* or *DFG16*), multivesicular body sorting pathway (*DID4* and *VPS24*, which comprise an ESCRT-III subcomplex), and the AAA-type ATPase gene, *VPS4* (Figure 3.13).

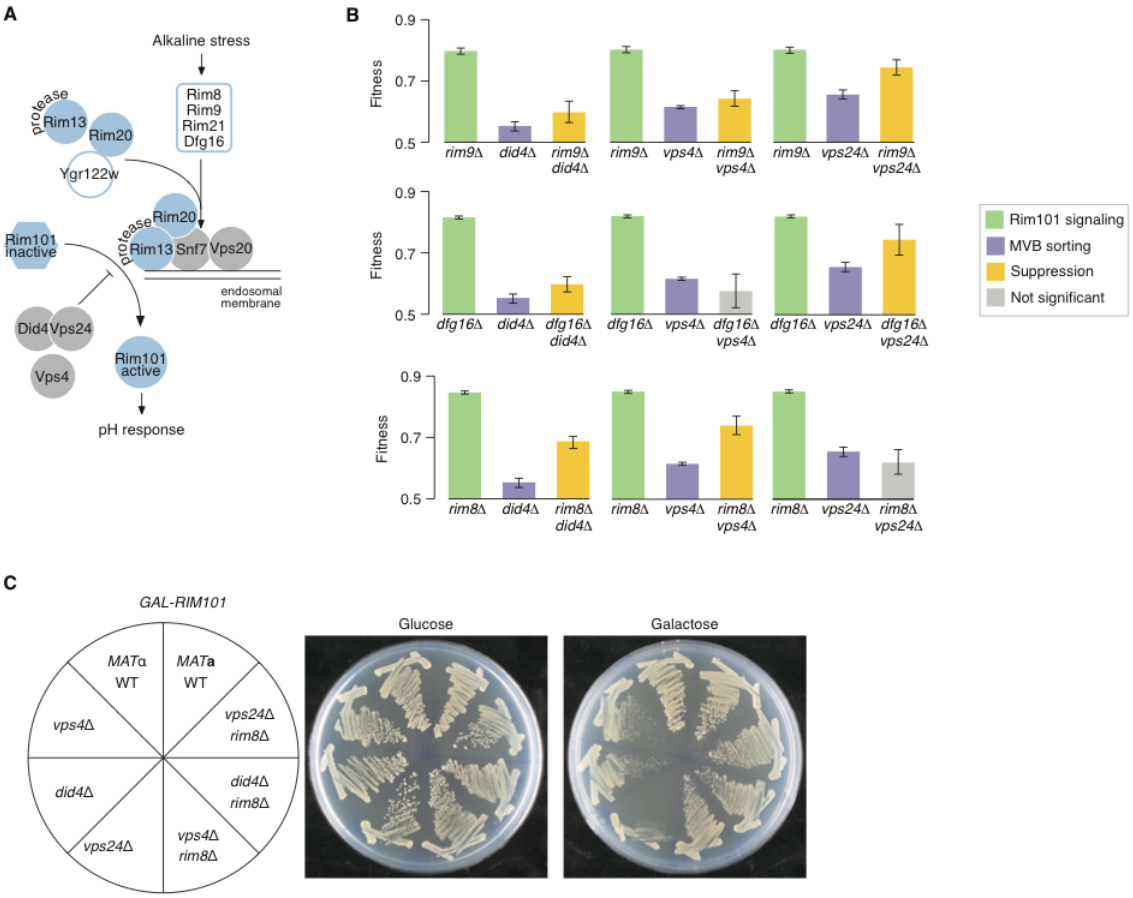


Figure 3.13. Genetic suppression between Rim101 signaling and MVB sorting pathways.

(**A**) Schematic showing activation of the Rim101 pathway in response to alkaline stress.
(**B**) Comparison of colony-size derived single and double mutant fitness measures suggests that *rim8*Δ, *rim9*Δ and *dfg16*Δ suppress growth defects associated with deletion of *DID4*, *VPS4* or *VPS24*. Error bars for single mutants: s.e.m. derived from bootstrapping (n = 800). Error bars for double mutants: s.d. (n = 4).
(**C**) Analysis of *RIM101* overexpression in the indicated single and double mutant backgrounds.

A functional relationship between Rim101 signaling and multivesicular body sorting has been established previously (Hayashi *et al.*, 2005; Mitchell, 2008). *RIM101* encodes a transcription factor activated in response to alkaline growth conditions via proteolytic cleavage at endosomal membranes (Figure 3.13A). Our suppression network indicated that deletion of genes encoding upstream signaling components of the Rim101 pathway suppressed

fitness defects associated with deletion of *DID4*, *VPS24* or *VPS4* (Figure 3.13B). This may occur because a defect in upstream signaling prevents constitutive activation caused by loss of function of downstream negative regulators. We confirmed that *did4*Δ, *vps24*Δ and *vps4*Δ mutants were sensitive to *RIM101* overexpression, and this sensitivity was rescued by deletion of *RIM8* (Figure 3.13C), a putative upstream component of the Rim101 pathway (Hayashi *et al.*, 2005) (Figure 3.13A). Interestingly, in the presence of lithium chloride, a condition in which the Rim101 pathway is activated and required for viability, suppression is observed in the opposite direction (Hayashi *et al.*, 2005), thus emphasizing the importance of considering condition-specificity when inferring pathway architecture based on genetic interactions (Avery and Wasserman, 1992). In addition to characterized pathway components, our genetic interaction analysis placed *DFG16*, an uncharacterized gene associated with Rim101 signaling (Barwell *et al.*, 2005), upstream in the pathway along with *RIM8* and *RIM9*. Another gene previously implicated in the Rim101 signaling, *YFR122W* (Rothfels *et al.*, 2005), likely functions further downstream and closer to the Rim13 protease (Costanzo *et al.*, 2010).

In another example derived from our cross-complex suppression network, disruption of the FAR complex, originally implicated in cell-cycle control (Kemp and Sprague, 2003), rescued growth defects associated with TORC2 kinase complex mutant alleles, *tor2-29* and *tsc11-1* (Figure 3.14A–B). Moreover, *FAR11* deletion suppressed actin polarization defects of a *tsc11-1* mutant in a nonpermissive condition (37ºC, Figure 3.14C). These results suggest that the FAR complex may function downstream to negatively regulate TORC2 function in actin organization. Similar to TORC2, FAR complex members are conserved from yeast to humans, and mammalian Far protein orthologs belong to a multiprotein complex that con-

tains the PP2A phosphatase (Goudreault *et al.*, 2009). Suppression of TORC2 growth and actin polarity defects was also achieved by loss of *PPG1* (Figure 3.14A–C), which encodes a PP2A-related phosphatase (Posas *et al.*, 1993). Thus it is possible that the FAR complex mediates its function by working with Ppg1 to dephosphorylate and inactivate proteins that normally control actin-based cell polarity.

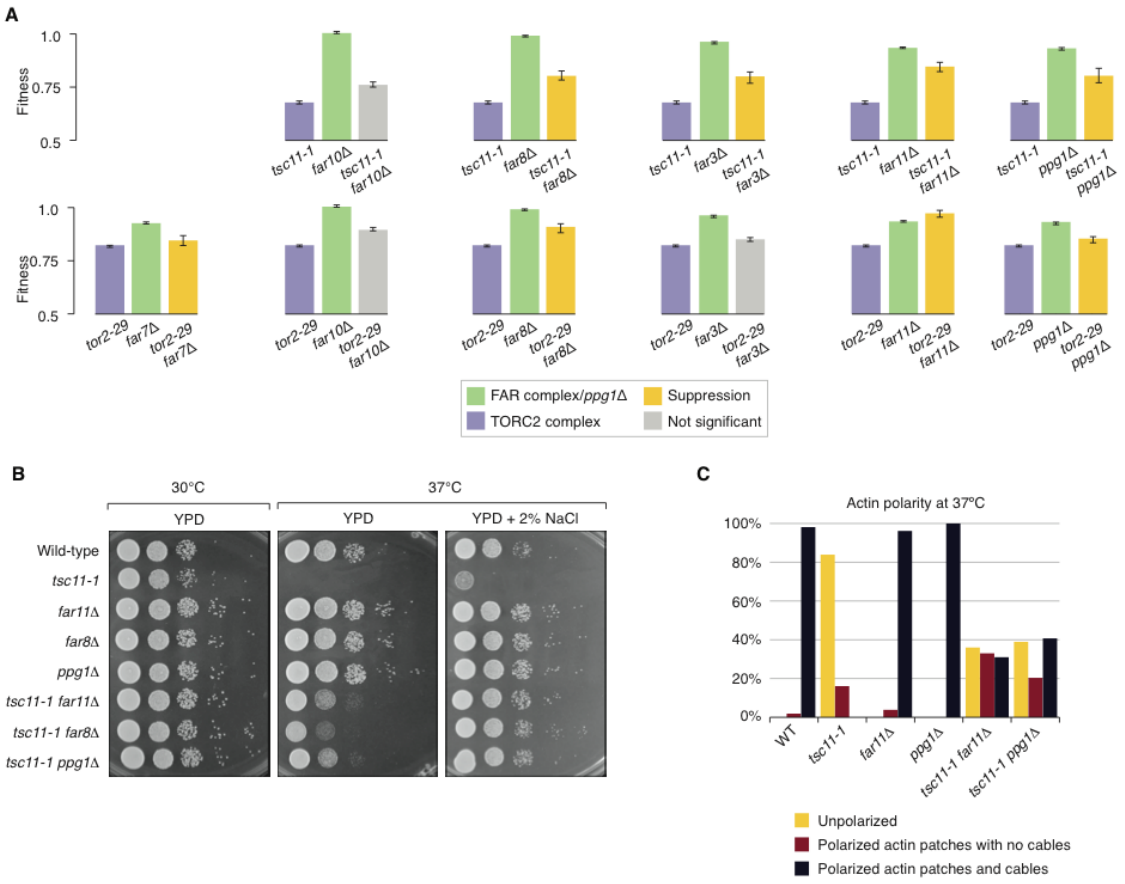


Figure 3.14. Genetic suppression between FAR and TORC2 protein complexes.

(**A**) Comparison of colony size-derived single and double mutant fitness measures suggests that mutations in genes encoding members of the Far3–11 protein complex suppress growth defects associated with mutant alleles of TORC2 kinase complex gene members, *tor2-29* and *tsc11-1*. Similarly, *tor2-29* and *tsc11-1* mutants are also suppressed by deletion of *PPG1*, a PP2A-related serine-threonine phosphatase.
(**B**) Serial dilution growth assays confirm that *tsc11-1* growth defects are suppressed by deletion of *FAR11*, *FAR8* and *PPG1* under semipermissive (30ºC) and nonpermissive (37ºC) temperatures (Materials & methods 3.13.14.1). Suppression interactions are enhanced in the presence of 2% (0.4 M) NaCl.
(**C**) A *tsc11-1* temperature-sensitive mutant exhibits an abnormal actin morphology that is suppressed by loss of *FAR11* and *PPG1*. The extent of actin polarization was quantified in wild type and the indicated single and double mutants.

Genes involved in chromatin and secretory functions act as hubs in the global genetic interaction map (Chapter 3.6) and our suppression network exhibited a similar topology (Figure 3.11). Specifically, protein complexes involved in chromatin modification or secretion suppressed growth defects associated with disruption of several different pathways and/or complexes. One such interaction involved suppression of DNA polymerase delta (Polδ) mutants by the disruption of the COG complex (Figure 3.11, Figure 3.15A). Unlike Polδ, which functions in the nucleus, the COG complex is important for the structure and function of the Golgi apparatus (Ungar *et al.*, 2006), indicating genetic suppression analysis can uncover very broad connections between functionally diverse genes. To explore this genetic relationship, we assessed whether *cog*Δ deletion mutants also rescued the UV light-sensitivity of a strain lacking a nonessential Polδ gene, *POL32* (Figure 3.15B). Upon UV-light exposure, *cog7*Δ *pol32*Δ and *cog8*Δ *pol32*Δ double mutants grew as well as *cog7*Δ and *cog8*Δ strains, confirming the COG-Polδ suppression interactions.

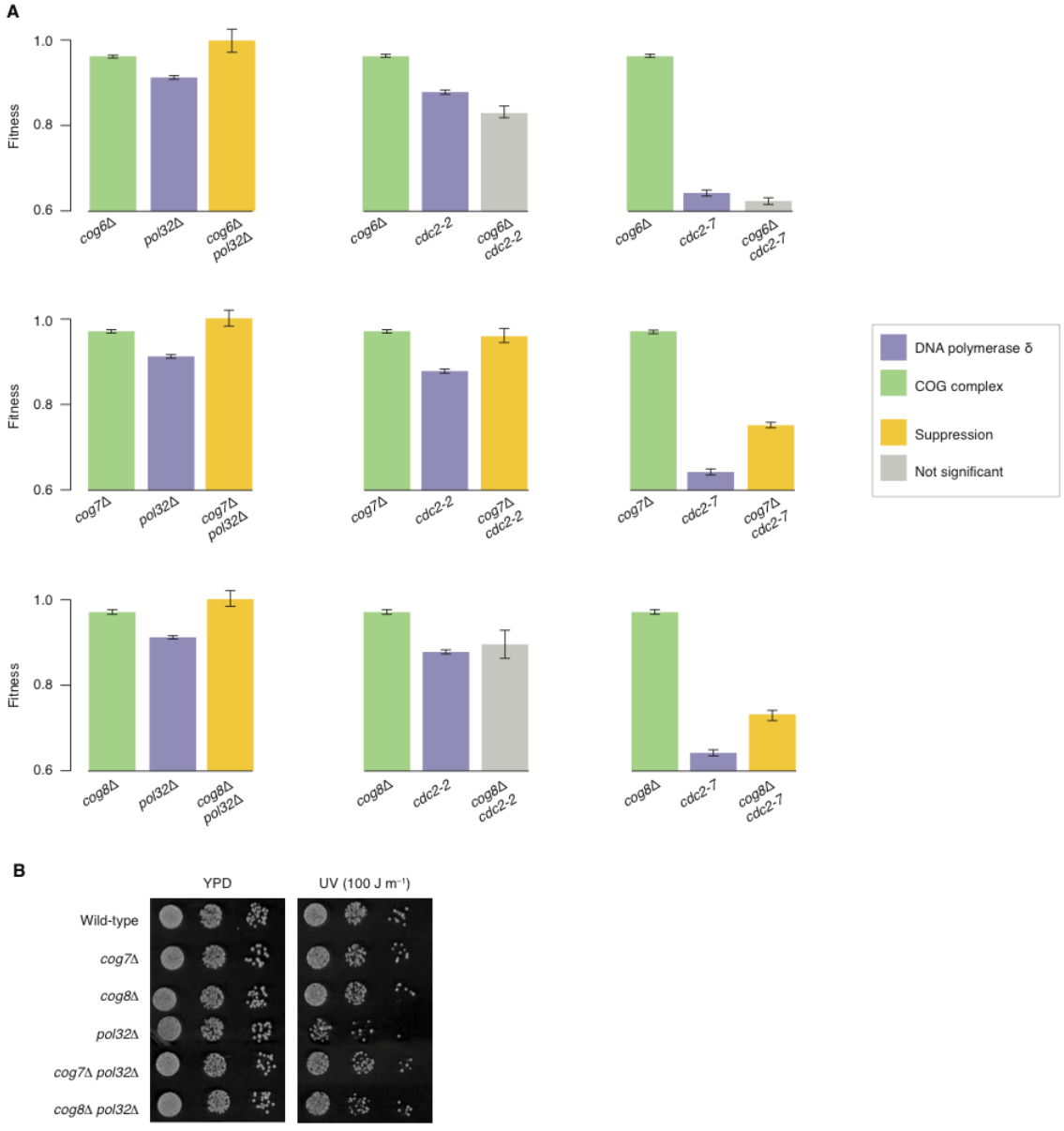


Figure 3.15. Genetic suppression between Polδ and COG protein complex.

(**A**) Comparison of colony size-derived single and double mutant fitness measures suggests that mutations in genes encoding members of the COG protein complex suppress growth defects associated with mutant alleles of DNA polymerase δ genes, *pol32*Δ, *cdc2-2* and *cdc2-7*.
(**B**) Serial dilution growth assays show that sensitivity to UV light of the *pol32*Δ mutant is suppressed by deletion of *COG7* and *COG8*.

3.11 Integrating genetic and chemical-genetic interaction networks

Findings obtained in various model organisms suggest that cell buffering strategies against genetic perturbations are similar to the strategies employed to resist chemical and environmental insults (Ho *et al.*, 2011). This observation is particularly relevant for the development of successful therapeutic approaches where the administration of chemical compounds seeks to selectively arrest the growth of specific cell types, such as tumor cells, which often carry mutations that distinguish them from the normal cells of an organism.

To explore the chemical resistance mechanisms in yeast, the set of ~4,700 viable deletion mutants has been exposed to hundreds of different chemical compounds (Hillenmeyer *et al.*, 2008). Numerous chemical-genetic interactions were identified as specific drug-mutant combinations causing a severe growth phenotype. We compared the number of chemical or nutritional environments causing growth defects in a particular deletion mutant to the number of genetic interactions associated with the corresponding gene, and found them to be highly correlated ($R = 0.4$, $p < 10^{-5}$; Figure 3.5). This finding suggests that hubs on a chemical-genetic network are predictive of hubs on the genetic interaction network and can be used to link environmental capacitance to genetic robustness. Furthermore, our data suggest that the same genes buffer the cell against both environmental and genetic insults. It is not known whether natural selection favors genetic robustness (de Visser *et al.*, 2003), but the positive correlation between genetic interaction degree and environmental capacitance suggests that genetic and environmental robustness may coevolve (Meiklejohn and Hartl, 2002).

Because chemical perturbations can mimic genetic perturbations, the genetic network should be useful for predicting the cellular targets of bioactive molecules (Parsons *et al.*, 2004). I identified genetic interaction profiles that are significantly correlated to a chemical-genetic profile of a particular compound (Materials & methods 3.13.15) and showed that compounds often clustered to dense regions of the genetic network indicative of specific bio-processes (Figure 3.16A). For example, hydroxyurea, a compound that inhibits ribonucleotide reductase and blocks DNA synthesis, clusters with genes involved in DNA replication and repair (Figure 3.16). These results demonstrate that clustering of chemical-genetic and genetic interaction profiles complements haploinsufficiency profiling, which has the potential to identify drug targets directly (Hillenmeyer *et al.*, 2008).

I used this network approach to examine a previously uncharacterized compound, 0428-0027. This compound clustered with genes associated with protein folding, glycosylation, and cell wall biosynthesis functions (Figure 3.16A) because its chemical-genetic profile most closely resembled the genetic interaction profile of *ERO1*, an essential member of the oxidative protein folding pathway (Figure 3.16B,D). I hypothesized that Ero1 might indeed be the target of 0428-0027 and named the compound erodoxin.

Several additional lines of evidence supported our hypothesis. *ERO1* encodes an essential enzyme participating in disulfide bond formation by recharging protein disulfide isomerase (PDI) into its oxidized state (Sevier *et al.*, 2007) (Figure 3.16E). Heterozygous deletion mutants *ero1*Δ/+ and *fad1*Δ/+, which encodes a cofactor that functions in the Ero1 reaction cycle (Wu *et al.*, 1995), were the most hypersensitive mutants identified from haploinsufficiency profiling of erodoxin (Figure 3.16C). Moreover, we found that erodoxin leads to inhi-

bition of Trx1 oxidation, an *in vitro* substrate of Ero1 (Figure 3.16F), and delays the processing of carboxy peptidase Y (CPY), a widely used reporter for protein folding and maturation (Figure 3.16G). Both these findings suggest that erodoxin inhibits Ero1 activity both *in vitro* and *in vivo*.

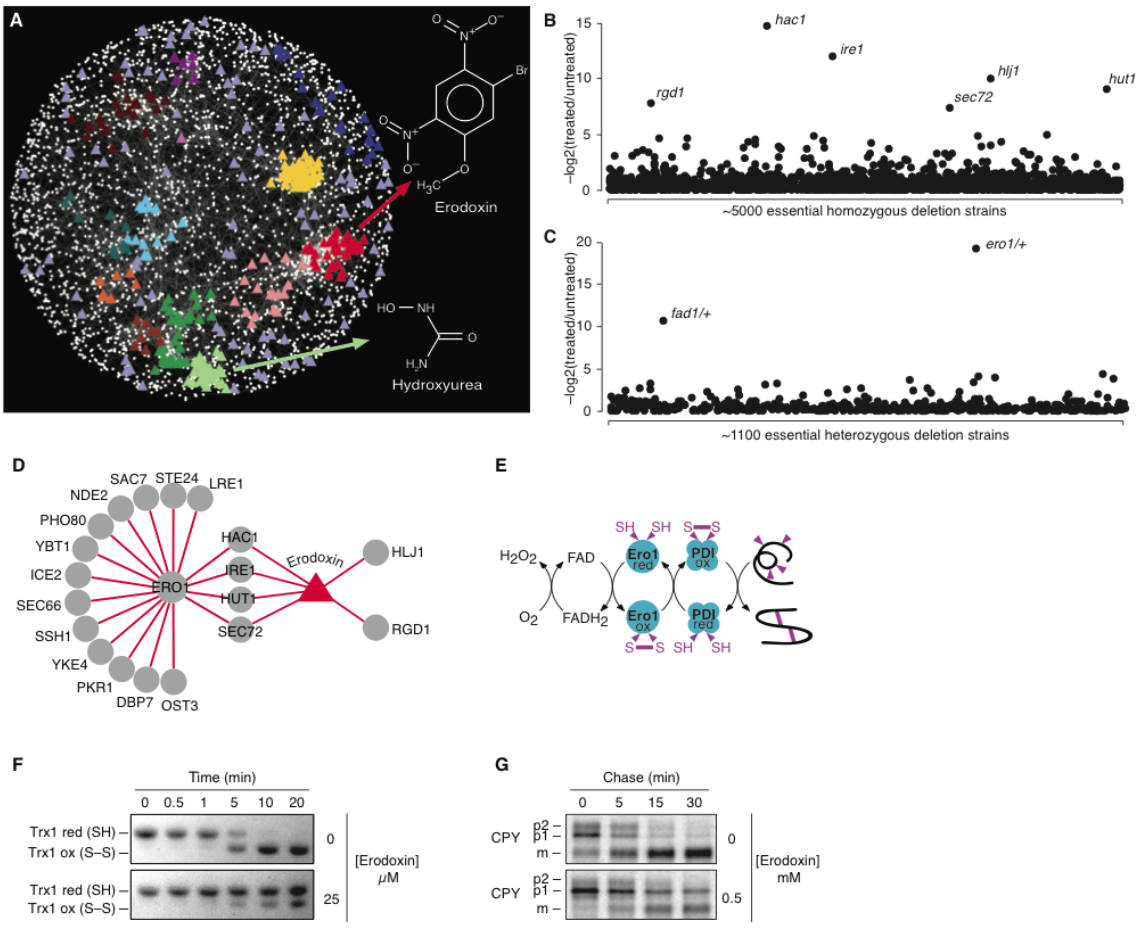


Figure 3.16. Integrating genetic and chemical-genetic interaction networks

(**A**) A chemical-genetic map was constructed on the basis of the genetic interaction map as in Figure 3.1. A compound was positioned next to the gene whose genetic interaction profile most closely resembles the chemical-genetic profile of the compound. Genes are represented as white circles. Chemical compounds are represented as triangles, which are colored according to the color of the corresponding functional cluster as in Figure 3.1. The chemical-genetic profile of hydroxyurea clustered with genes involved in DNA replication and repair, whereas that of erodoxin clustered with genes involved in protein folding, glycosylation, and cell wall biosynthesis. Compounds positioned outside functional clusters are colored light purple.
(**B–C**) Scatter plot visualizations of homozygous (**B**) and heterozygous (**C**) deletion profiling of erodoxin. The y-axis represents the sensitivity of homozygous deletion strains of nonessential genes (**B**) or heterozygous deletion strains of essential genes (**C**) when grown for 5 generations in the presence of the compound. Strains exhibiting severe sensitivity are labeled. Experiments were conducted as described previously (Hillenmeyer *et al.*,

2008).
(**D**) Network displaying overlap between *ERO1* negative genetic interactions and genes hypersensitive to erodoxin.
(**E**) *ERO1*-dependent pathway for oxidative protein-folding.
(**F**) Erodoxin inhibits Ero1-dependent oxidation of Trx1 *in vitro*.
(**G**) Erodoxin inhibits CPY processing to the vacuolar form *in vivo*. ER (p1), Golgi (p2) and vacuolar (m) forms of CPY are indicated.

3.12 Summary & conclusions

The data presented in this chapter shows that unbiased, systematic, and quantitative analysis of pairwise genetic interactions assigns a rich phenotypic profile to every gene and enables construction of a functional map of the cell, organizing genes and higher-order bio-processes according to their related roles (Figure 3.1). The functional connections uncovered by genetic interactions complement the information derived from physical interaction networks (Figure 3.8), and link previously uncharacterized genes to specific pathways and protein complexes (Figure 3.2). Moreover, genetic interactions reveal long-ranging connections between pathways and complexes, and in some cases identify complex regulatory relationships (Figure 3.2).

We show that genetic interaction hubs share numerous physiological and evolutionary properties (Figure 3.5) and tend to participate in specific biological processes (Figure 3.6). Genetic interaction hubs in other organisms, such as *C. elegans*, seem to play similar biological roles (Lehner *et al.*, 2006), thus suggesting that, while conservation of individual genetic interactions may be limited, the overall network topology and its properties may be more highly preserved. This is particularly important as global mapping of genetic networks becomes approachable in more complex cells and higher organisms (Costanzo *et al.*, 2011). Indeed, the ability to prioritize genes based on their centrality in the network, estimated from

their biological and evolutionary properties, is key for time-efficient and cost-efficient experiments.

We show that genetic and physical interaction networks, despite several common properties, are mostly orthogonal to each other and share little overlap (Figure 3.8). A small subset of negative genetic interactions connect members of essential protein complexes (Figure 3.9), potentially uncovering internal redundancy among complex members. In contrast, positive genetic interactions tend to link members of nonessential protein complexes (Figure 3.9), especially those sharing a direct physical interaction as defined by yeast-two-hybrid assays (Figure 3.9). However, a surprisingly high number of positive interactions span across protein complexes (Figure 3.10) and chart a cross-complex genetic suppression network (Figure 3.11).

The integration of genetic and chemical-genetic perturbation data shows that genetic interaction hubs also tend to be hypersensitive to numerous chemical compounds (Figure 3.5), suggesting a common cellular mechanism for protection against both genetic and chemical/environmental insults. Moreover, our global genetic interaction network helps linking bioactive compounds to their targets (Figure 3.16) and offers the potential to design synthetic lethal therapies for targeting genetically defined tumors (Fong *et al.*, 2009). In addition, it may be possible to devise drug combination therapies based on the presence of genetic interactions between drug targets (Lehar *et al.*, 2008), although a recent study suggests that this may be more challenging than anticipated (Cokol *et al.*, 2011).

In summary, genetic interaction maps provide a model for understanding the functional organization of a cell and the general organizing principles of complex biological networks. Genetic interactions underlie the intricate relationship between the genotype and the phenotype of an organism, and must play a key role in governing inherited human disease. As recognition of the importance of genetic interactions grows among human geneticists (Zuk *et al.*, 2012), yeast remains an invaluable tool for studying genetic interactions in a relatively simple and controlled setting. Discoveries made in the yeast model system will have an immense impact on our global understanding of human biology.

3.13 Materials & methods

3.13.1 Identifying subnetworks from the global map

The correlation-based network (Figure 3.1) was clustered using Markov Clustering Algorithm (MCL; inflation = 1.4) (van Dongen, 2000), as well as GINECA, a newly developed approach for identifying partially overlapping clusters (Materials & methods 3.13.2). Clusters identified in both methods were aggregated into eight subnetworks on the basis of their relatedness, measured by the degree of cluster overlap for GINECA and by average network proximity after 25 network layout iterations for MCL. Eighty functionally unrelated genes were removed from the network because their similarities may be due to experimental artifacts, such as residual systematic effects whose influence on clustering is enhanced in the absence of strong genetic interactions.

3.13.2 GINECA clustering algorithm

GINECA (Genetic Interaction Network Enumerative Clustering Algorithm) is a scalable semi-supervised graph clustering algorithm for identifying overlapping clusters of functionally related genes from dense weighted molecular interaction networks. The core of GINECA is a set of matrix operations, simulating density-based searches of local maxima in a 2-dimensional data space, applied recursively to extract submatrices from a large dense matrix representation of weighted interaction scores between the genes. This iterative approach enables some genes to have overlapping memberships in more than one functional cluster, since a gene can be involved in multiple biological processes. Through a constraint-based semi-supervised clustering approach, GINECA exploits known functional assignments of the genes, such as the Gene Ontology (GO) annotations, to guide the clustering algorithm towards more meaningful data partitioning, thus generating clusters of highly connected genes that are also functionally enriched. One or more clusters meeting statistical functional significance criteria are pruned from the network at each iteration, with the algorithm converging when all functionally enriched clusters are extracted, or when the remaining network is loosely connected.

3.13.3 Gap1 permease localization and activity

Strains deleted for *PAR32*, *ECM30* and *UBP15*, and expressing *GAP1-GFP* were grown in SD ammonium medium and Gap1 localization was visualized as described previously (Risinger *et al.*, 2006). Wild type (WT) cells and *gtr1*Δ mutant cells were included as positive and negative controls, respectively. Gap1 permease activity was measured by 14C-

citrulline uptake assays as described previously (Roberg *et al.*, 1997). Data shown are the average of four independent experiments.

3.13.4 Identification of Sgt2 and Ubp15 physical interactors

3.13.4.1 Tandem Affinity Purification

TAP-tagged strains were grown to OD_{600} ~0.7 in Synthetic Complete media lacking histidine. Cells were collected by centrifugation and lysed (vortexing with glass-beads), and the lysate cleared by centrifugation, followed by filtering (0.45 μm filters). Purification and identification of TAP-tagged proteins and their interacting partners was performed as described in (Gingras *et al.*, 2005), except that human IgG cross-linked to tosyl-activated magnetic beads were employed instead of IgG-sepharose. Following elution with 25 mM EGTA in 50 mM ammonium bicarbonate (pH 8), 500 ng of sequencing-grade trypsin was added, and the samples were incubated at 37ºC overnight. Digested samples were analyzed by nanoscale liquid chromatography coupled to tandem mass spectrometry (nLC-MS/MS), as described in (Peng *et al.*, 2003).

3.13.4.2 Protein identification and initial data analysis

LTQ data files were converted from binary .raw files to Mascot generic files (*.mgf) and searched using Mascot 2.2.1 database search tool (Perkins *et al.*, 1999) against the *S. cerevisiae* complement of RefSeq release 21 containing both forward and reverse protein entries (total 11,668 entries). Variable modifications were deamidation of asparagine and glutamine, oxidation of methionine and 1 missed trypsin cleavage. Mass accuracy requirements

for searches were set to 3 Da on the parent, and 0.6 Da on product ions. Peptide scores above 35 corresponded to a protein false discovery rate of < 4%, determined using a target-decoy strategy (Elias and Gygi, 2007).

3.13.4.3 Data filtering and interpretation

A list of likely contaminants in TAP samples was generated by compiling results from 62 unrelated TAP purification experiments. Proteins detected in 10% or more of these 62 samples were placed on a contaminant list provided in Supplementary Table 1 in Costanzo *et al.* (Costanzo *et al.*, 2010). All other proteins identified with 2 or more unique peptides were considered for subsequent analysis, and are listed in Supplementary Table 2 in Costanzo *et al.* (Costanzo *et al.*, 2010). To further validate the specificity of the detected interactions, we compared spectral counts for each protein detected in the Sgt2-TAP and Ubp15-TAP experiments with spectral counts detected across 19 unrelated experiments. In particular, we analyzed the occurrence of detection of a given hit across these 19 unrelated samples, but also the maximal spectral counts with which this hit was detected across these samples. This allowed us to calculate an enrichment ratio for each of the observed hits in the Sgt2-TAP or Ubp15-TAP samples. For example, Sse1 (detected with 11 peptides in Sgt2-TAP) was detected in 2 out of 19 unrelated samples, with a maximal spectral count of 2 in these samples, indicating high enrichment in the Sgt2-TAP sample. Only those peptides detected with a number of spectra ≥ 5 and 5-fold "enrichment" are reported in the main portion of the text.

3.13.5 Pex15 localization

A pAG-EGFP-*PEX15* plasmid was constructed by sub-cloning the *PEX15* ORF from the Yeast FLEXGene collection (Hu *et al.*, 2007) into the Gateway(R) donor vector pDONR201 and subsequent cloning into an N-terminal GFP fusion expression vector (pAG416GAL-GFP) as described previously (Alberti *et al.*, 2007). The EGFP-*PEX15* plasmid was transformed into WT, *sgt2*Δ and *get4*Δ strains. Cell cultures were grown at 30ºC in SD-HIS media to early log phase, harvested by centrifugation. A 1.5 µl suspension was spotted onto a glass slide and images were captured using Quorum WaveFX Spinning Disc Confocal System.

3.13.6 Elp/Urm amino acid usage analysis

The three codons that pair with the URM/ELP modified tRNAs are AAA, GAA and CAA. For every ORF in the yeast genome, we calculated the relative frequency of the three codons by dividing their occurrence by the total number of codons for any amino acid in the sequence. The relative frequencies of these codons in URM- and ELP-negative interacting genes were compared to a background distribution composed of all the genes tested in our study, excluding the URM/ELP interactors. Significance was evaluated using t-test and Wilcoxon rank sum statistics.

3.13.7 Identification of genetic interaction hubs and monochromatic genes

We measured the number of positive and negative genetic interactions for all 3,885 non-essential array deletion mutants at the intermediate confidence threshold ($|\varepsilon| > 0.08$,

$p < 0.05$). Genetic interaction hubs were selected as the top 10% most connected genes. Genes with a bias towards positive interactions were selected by finding genes with at least 30 total interactions and positive to negative ratio greater than 1, which is twice the background ratio. Genes with a bias towards negative interactions were selected by finding genes with at least 30 total interactions and positive to negative ratio lower than 0.25, which is one-half of the background ratio. Both of these sets consist of approximately 130 genes.

3.13.8 Functional enrichment analysis of genetic interaction hubs and monochromatic genes

Genes were evaluated for functional enrichment using the standard approach described in (Boyle *et al.*, 2004) based on the hyper-geometric distribution with the Westfall and Young step-down procedure for multiple hypothesis correction (Westfall and Young, 1993). Enrichment analysis was based on annotations downloaded from the Saccharomyces Genome Database (www.yeastgenome.org) on July 9th, 2009.

3.13.9 Biological process annotations

Biological process categories were derived from functional enrichment of network clusters identified in Figure 3.1 and further refined by examining cluster enrichment of network sub-regions in Figure 3.2. Genes were assigned to one or more of these functional categories on the basis of their Gene Ontology annotations and gene function descriptions available from the *Saccharomyces* Genome Database (SGD, www.yeastgenome.org). Specific annotations are provided in the Supplementary material in Costanzo *et al.* (Costanzo *et al.*, 2010). A subset of uncharacterized or multifunctional genes were not assigned to any functional category.

3.13.10 Correlation analysis of genetic and physical interaction degree

The number of positive and negative interactions for all 3,885 non-essential array deletion mutants was calculated using the intermediate confidence cutoff ($|\varepsilon| > 0.08$, $p < 0.05$). Pearson correlation coefficients (PCC) were measured between each of the quantitative features described below and positive genetic interaction degree, negative genetic interaction degree and protein-protein interaction degree (Figure 3.5A). Protein-protein interaction degree was computed using the union of AP/MS (Gavin *et al.*, 2006; Krogan *et al.*, 2006), two-hybrid (Yu *et al.*, 2008), and PCA (Tarassov *et al.*, 2008) datasets.

Given that single mutant fitness is a major correlate of each feature, we repeated the analysis controlling for the effect of single mutant fitness through the calculation of partial correlation (Kendall *et al.*, 1994). Correlation between genetic interaction degree and each feature remains significant, as does the distinction between genetic and protein-protein interaction hubs (Costanzo *et al.*, 2010).

3.13.10.1 Single mutant fitness defect

Single mutant fitness was derived from colony size data as described in Chapter 2.5. For correlation analysis, we used fitness defect $d_i = 1–f_i$ where f_i is the fitness of single mutant i.

3.13.10.2 Multi-functionality

For each gene, multi-functionality was measured as the total number of annotations to a set of functionally distinct GO terms described in (Myers *et al.*, 2006).

3.13.10.3 Phenotypic capacitance

Measures of phenotypic capacitance was derived directly from (Levy and Siegal, 2008).

3.13.10.4 Chemical-genetic degree

A measure of chemical-genetic degree was derived from (Hillenmeyer *et al.*, 2008), which measured the sensitivity of all non-essential deletion mutants to a library of drugs and a variety of environmental conditions. Specifically, for each deletion mutant, we summed the number of drug and environmental sensitivities in the homozygous profiling dataset that met a p-value < 0.05 cutoff.

3.13.10.5 PPI degree

The protein-protein interaction degree was measured as the total number of interactions in the union of four high-throughput physical interaction datasets (Gavin *et al.*, 2006; Krogan *et al.*, 2006; Tarassov *et al.*, 2008; Yu *et al.*, 2008). Interactions from (Gavin *et al.*, 2006; Krogan *et al.*, 2006) were obtained from BioGRID. Interactions from (Yu *et al.*, 2008) correspond to the Y2H-Union dataset reported in the publication. Interactions from (Tarassov *et al.*, 2008) were defined using the 97.7% confidence cutoff as recommended.

3.13.10.6 Protein disorder

The protein disorder measure is the percent of unstructured residues as predicted by the Dispred2 software (Ward *et al.*, 2004) and reported in (Gsponer *et al.*, 2008). For correlation

and ANCOVA analyses, this percentage was used directly as a correlate. For categorical analysis, we adopted the binning suggested in the original study: genes reporting scores lower than 0.1 were considered to be structured, scores between 0.1 and 0.3 to be moderately unstructured and scores above 0.3 to be unstructured.

3.13.10.7 Expression level

Expression level measurements were obtained from (Holstege *et al.*, 1998).

3.13.10.8 Yeast conservation

A measure of conservation across the Ascomycota phylum was computed as the number of species (out of a total of 23) that possess an ortholog for a given gene (Wapinski *et al.*, 2007). Ortholog data was downloaded from www.broadinstitute.org/regev/orthogroups/. The 23 species are an expanded set of the 17 species described in (Wapinski *et al.*, 2007), with the additions of *S. octosporus*, *S. japonicus*, *L. elongosporus*, *C. parasilosis*, *C. tropicalis* and *C. guilliermondii*.

3.13.10.9 Copy number volatility

Volatility measures the frequency with which a given gene is gained (including duplication) or lost across 23 species of Ascomycota fungi (Wapinski *et al.*, 2007).

3.13.10.10 dN/dS ratios

We computed the average *dN*/*dS* ratio for *S. cerevisiae* in comparison to the *sensu strictu* yeast species (*S. paradoxus*, *S. bayanus* and *S. mikatae*). Sequences were aligned using MUSCLE (Edgar, 2004) and *dN*/*dS* ratios were computed using PAML (Yang, 2007).

3.13.10.11 Evolutionary age

As an approximate measure of evolutionary age, we used a reconstruction of the ancestral species as described in (Kunin and Ouzounis, 2003).

3.13.10.12 Gene duplicates

The list of duplicate pairs consists of those identified as the result of the whole genome duplication events as reconciled from several sources (Byrne and Wolfe, 2005). Additionally, any pair of genes fulfilling established similarity requirements (Gu *et al.*, 2002) was reasoned to be a duplicate pair resulting from a small scale duplication event. Specifically, the gene pair must have a sufficient sequence similarity score (FASTA Blast, E = 10) and sufficient protein alignment length (> 80% of the longer protein). The pair must also have an amino acid level identity of at least 30% for proteins with aligned regions longer than 150 amino acids. For shorter proteins the identity must exceed $0.01n + 4.8L - 0.32^{(1 + \exp(-L/1000))}$, where L is the aligned length and $n = 6$ (Rost, 1999; Gu *et al.*, 2002). Pairs from the whole genome duplication event (WGD) were combined with pairs determined through sequence alone (SSD).

3.13.11 Analysis of variation in genetic interactions number across bioprocesses

Analysis of covariance (ANCOVA) was used to investigate variation in genetic interaction degree across biological processes. We measured the number of positive and negative interactions for all 3,885 non-essential array deletion mutants at the intermediate confidence threshold ($|\varepsilon| > 0.08$, $p < 0.05$), and used ANCOVA to estimate the contribution of a gene's functional category to determining its genetic interactions degree, in addition to other gene-specific properties that affect connectivity (Figure 3.5A). Specifically, we included 17 biological processes (Materials & methods 3.13.9) as independent variables in a linear model relating genetic interaction degree to various gene-specific features (Figure 3.5A). We assessed whether each biological process had a significant effect on genetic interaction degree in the presence of other factors. The complete results of the ANCOVA analysis for negative and positive genetic interactions are reported in the Supplementary material in Costanzo *et al.* (Costanzo *et al.*, 2010).

3.13.12 Analysis of genetic interactions within and between protein complexes

3.13.12.1 Construction of the protein complex standard

A literature-curated standard of 430 protein complexes was compiled by combining the SGD Macromolecular Complex GO annotations (www.yeastgenome.org), the CYC2008 protein complex catalog (Pu *et al.*, 2009) and 26 manually curated complexes/pathways. Redundant protein complex annotations were minimized by eliminating all but one complex with identical components and by excluding smaller complexes if all their members also belong to

a larger complex. A complete list of protein complexes and their members is provided in Supplementary Data 2 in (Baryshnikova *et al.*, 2010b).

3.13.12.2 Analysis of genetic interaction enrichment

We calculated enrichment of genetic interactions among members of protein complexes where at least two gene pairs were tested in our study. We ignored all dubious ORFs and restricted our analysis to interactions satisfying a lenient confidence threshold ($p < 0.05$). Moreover, interactions involving different alleles of the same gene were merged.

Significance of enrichment within a protein complex i was evaluated using the hypergeometric distribution as follows:

$$P_i = 1 - \sum_{n_i=0}^{X_i-1} \frac{\binom{K}{n_i}\binom{M-K}{N_i-n_i}}{\binom{M}{N_i}} \qquad \textbf{Eq. 3.1}$$

where M is the total number of gene pairs screened in this study, K is the total number interactions at the chosen cutoff, N_i is the number of gene pairs screened within complex i and X_i is the number of genetic interactions within complex i.

Enrichment of genetic interactions between protein complexes i and j was assessed in a similar manner but replacing N_i and X_i with N_{ij} and X_{ij}, where N_{ij} is the number of gene pairs screened between i and j and X_{ij} is the number of interactions recovered.

3.13.12.3 Monochromatic analysis of genetic interactions

To evaluate the prevalence of positive or negative genetic interactions within a protein complex *i*, we computed a monochromatic purity score (MP-score) as follows:

$$MP_i = \frac{\frac{P_i - N_i}{P_i + N_i} - \frac{P - N}{P + N}}{1 - \alpha \frac{P - N}{P + N}} \qquad \textbf{Eq. 3.2}$$

where P and N are the total number of positive and negative scores in the dataset, P_i and N_i are the number of positive and negative interactions between members of complex i, and $\alpha = sign(P_i - N_i)$.

The MP-score is designed to assume a value of +1 for a monochromatic positive protein complex, and a value of –1 for a monochromatic negative complex. A complex reflecting the background ratio of positive to negative interactions will present an $MP_i = 0$.

Monochromaticity of between-complex genetic interactions was assessed in a similar way but with P_i and N_i replaced by P_{ij} and N_{ij} referring to the number of positive and negative interactions spanning across complex i and complex j, respectively.

3.13.13 Protein complex suppression network

To construct a cross-complex genetic suppression network, we identified all pairs of protein complexes that (a) share at least 5 positive genetic interactions in common, and (b) are statistically enriched for shared positive interactions (Materials & methods 3.13.12.2).

We then categorized between-complex positive interactions into suppression and masking subclasses using the following strategy.

Assuming that (1) f_a, f_b and f_{ab} are the fitness measures for single mutants *a*, *b* and the double mutant *ab*, respectively, (2) σ_a, σ_b and σ_{ab} are the corresponding standard deviations, and (3) $f_a = \max(f_a, f_b)$, we filtered all positive interactions such that $\varepsilon > 0.08$, $p < 0.05$, $f_a < 1$, $f_b < 1$ and $|f_a - f_b| \geq \sigma_a + \sigma_b$. Then, we classified all the remaining interactions into suppression and masking categories using the following rules:

(1) if $f_{ab} > f_b + \sigma_b$, mutant *a* was determined to suppress the phenotype of mutant *b*;

(2) if rule (1) was not true, but $f_{ab} < f_a - \sigma_a$, then mutant *b* was determined to mask the phenotype of mutant *a*.

In constructing the network of cross-complex genetic suppression, we only included protein complex pairs where more than 80% of gene pairs were classified as suppression and showed consistent directionality.

3.13.14 Suppression confirmation experiments

3.13.14.1 Spot dilutions: tsc11-1 suppression

Overnight cultures were serially diluted 20-fold and spotted onto agar medium as indicated. Strains were grown for two days at 30ºC or 37ºC as indicated.

3.13.14.2 Actin staining

Cultures were grown at 30ºC to early log phase (OD600 of 0.1~0.2) and then shifted to nonpermissive temperature (37ºC) for 6 hours. Actin staining was done as described (Friesen *et al.*, 2006) without added fluorescent brightener. Stained cells were imaged using a DMI 6000B fluorescence microscope (Leica Microsystems) equipped with a spinning-disk head, an argon laser (458, 488, and 514 nm; Quorum Technologies) and ImagEM-charge-coupled device camera (Hamamatsu C9100-13, Hamamatsu Photonics). Cells were imaged in 7 Z stacks with 0.3 μm intervals and all Z stacks were collapsed into one extended focus.

3.13.14.3 UV sensitivity: pol32Δ suppression

The UV sensitivity assay was performed as described previously (Bellaoui *et al.*, 2003).

3.13.15 Integrating genetic and chemical-genetic interaction data

Chemical sensitivity profiles of all non-essential yeast deletion mutants were obtained from previously published studies (Parsons *et al.*, 2006; Hillenmeyer *et al.*, 2008). In addition, homozygous and heterozygous profiling of Erodoxin and nine other chemical compounds were conducted as described previously and released as Supplementary Data File 7 in Costanzo *et al.* (Costanzo *et al.*, 2010).

For each chemical compound, Pearson correlation coefficients were computed between its chemical-genetic profile and the genetic interaction profiles of 1,712 queries in our dataset. Each compound was then assigned a position on the genetic interaction map corre-

sponding to its top most correlated query gene. If that position was already occupied by another compound, having a higher similarity to the same query gene, the compound was assigned a different position corresponding to its second best correlated query gene, and so on.

3.13.16 Ero1 *in vitro* activity and CPY processing assays

Ero1 *in vitro* activity was monitored in the absence or presence of 25 μM Erodoxin by examining oxidation of reduced Trx1. Proteins were resolved by non-reducing SDS-PAGE and Trx1 was visualized as described previously (Sevier *et al.*, 2007). *In vivo* activity of Ero1 was assessed by carboxypeptidase Y (CPY) processing. 0.5 mM Erodoxin was added to wild-type yeast cells in minimal medium, cells were incubated for 10 min at 30ºC, pulse labeled for 7 min, and chased for 0–30 min. CPY was immunoprecipitated and resolved by SDS-PAGE as described in (Sevier and Kaiser, 2006).

4 Mapping genome-wide recombination patterns using SGA

I conceived and performed all the analyses described in this Chapter.

Experimental data were obtained from the large-scale SGA experiments, designed and coordinated by Michael Costanzo.

4.1 Abstract

Synthetic Genetic Array (SGA) is a high-throughput method for constructing yeast double mutants and mapping genetic interactions. In addition, SGA experiments produce a genome-wide genetic linkage map for *S. cerevisiae*, because construction of double mutants involving loci on the same chromosome requires meiotic recombination, whose frequency affects double mutant colony size. Genetically linked loci, i.e. those that recombine rarely, form small double mutant colonies and are usually removed from SGA analysis as they can be mistaken for negative genetic interactions. I have developed a computational approach to correct for these so-called "genetic linkage effects" and recovered 7,958 novel negative interactions, which contribute to a more comprehensive coverage of the yeast genetic interaction network. In addition, I used genetic linkage data to map regions of high and low meiotic recombination. This analysis recapitulated previously identified recombination patterns, including inhibition of recombination at pericentromeric chromosomal regions. Interestingly, I found that the extent of centromere-related recombination repression is proportional to chromosome length, whereas low recombination regions outside the centromere are, on average, constant in size. The molecular mechanisms underlying the relationship between chromosome size and genetic linkage remain a mystery; however, this novel finding supports the existence of a chromosome size-dependent mechanism that counteracts the tension generated by microtubules during meiosis I. These results highlight SGA methodology as a unique approach for systematic analysis of yeast meiotic recombination.

4.2 Introduction

In most sexually reproducing organisms, genetic diversity is promoted by meiotic recombination, a molecular mechanism in which homologous chromosomes exchange genetic information and generate new allelic combinations (Petes, 2001). In addition, recombination is necessary to physically link homologous chromosomes during the first meiotic division, preventing their premature separation and ensuring that the reshuffled genetic information is accurately transmitted to the progeny (Zickler and Kleckner, 1998; Zickler and Kleckner, 1999). As a result, meiotic recombination is central to many aspects of chromosome biology, genetic robustness and evolution, and plays a primary role in shaping the genetic structure of natural populations. Failure to recombine is incompatible with growth and successful reproduction, as it results in severe cellular abnormalities, including chromosome nondisjunction, infertility, and birth defects.

In yeast, meiotic recombination is initiated by the endonuclease Spo11, which forms directed double-strand breaks (DSBs) during prophase of the first meiotic division (Keeney *et al.*, 1997) (Figure 4.1). Spo11 is a topoisomerase II-related protein that works as a dimer to break both DNA strands via a transesterification reaction (Keeney and Neale, 2006). Following the formation of the DSB, Spo11 remains covalently attached to DSB ends and must be removed by the endonucleolytic activity of the Mre11-Rad50-Xrs2 (MRX) complex in order for the DSB to be processed and repaired (Figure 4.1). DSB ends are then resected by exonucleases to produce 3’ single-strand DNA (ssDNA) overhangs, which are substrates for strand invasion, mediated by the RecA-homologs Dmc1 and Rad51 (Figure 4.1). Resolution of the recombinant structures formed by strand invasion can lead to a crossover event, resulting in

reciprocal exchange of the chromosome arms flanking the DSB, or a non-crossover event, in which only local exchange of genetic information occurs (Figure 4.1).

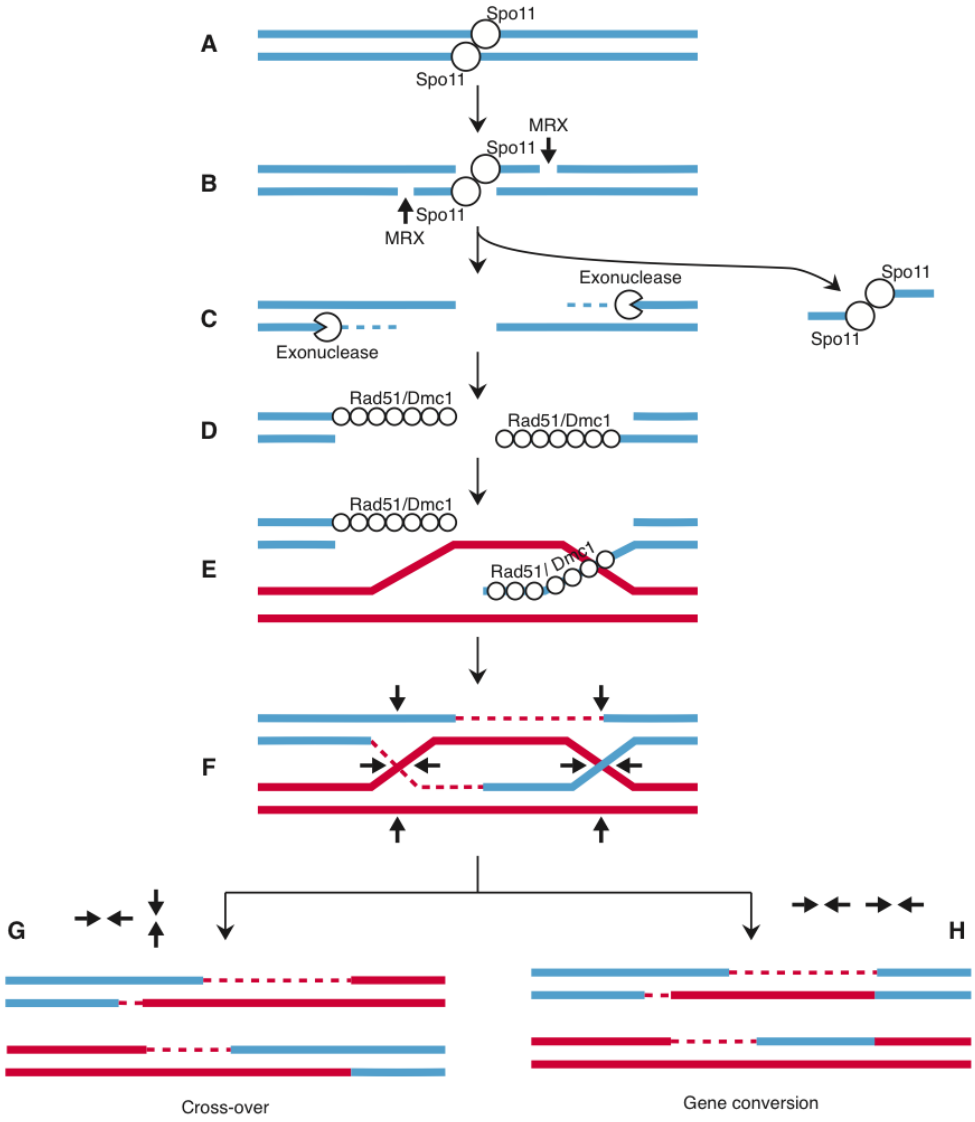


Figure 4.1. Mechanism of meiotic recombination.

A Spo11 dimer binds one of the four sister chromatids aligned at meiotic metaphase I (**A**) and initiates the formation of the DSB (**B**). Spo11 remains covalently attached to the 5' ends of the DSB and must be removed by two endonucleolytic nicks catalyzed by the MRX complex (**B**, vertical arrows). The ends of the DSB are then resected by exonucleases to form 3'-ssDNA overhangs (**C**), which are bound by strand-exchange proteins Rad51 and Dmc1 (**D**) and initiate strand invasion (**E**). The 3'-terminus of the invading strand primes DNA synthesis (**F**, dashed lines) using the homologous chromosome (solid red lines) as a template and displacing the resident strand. The displaced strand pairs with its complement from the homologous chromosome and forms a heteroduplex structure containing two Holliday junctions (**F**). Cleavage of the Holiday junctions in two alterna-

tive ways (arrows) leads to a cross-over (**G**) where chromosome arms are exchanged reciprocally, or gene conversion (**H**), where only local exchange of information occurs.

Observations in yeast and other organisms indicate that recombination events are nonrandomly distributed across the genome (Petes, 2001). Chromosomal regions with high recombination activity, known as recombination hotspots, are separated by regions where recombination is rare and genes tend to be inherited jointly (Petes, 2001). In humans, these sets of genetically linked loci, called haplotypes, are invaluable for associating genes with diseases and understanding the structure and history of populations (Lander, 2011). For example, large-scale genotypic projects, such as HapMap (The International HapMap Consortium, 2003), rely on genetic linkage to capture up to 90% of the total genetic variance in a population by genotyping only 5–10% of the sequence variants (Daly *et al.*, 2001; Lander, 2011).

In order to understand the causes of the nonrandom distribution of recombination events and to gain insight into their biochemical mechanisms, global maps of meiotic recombination have been produced in yeast using several experimental approaches. These studies focused primarily on mapping the initial phases of recombination, including Spo11 DNA binding (Gerton *et al.*, 2000; Pan *et al.*, 2011) and DSB formation (Baudat and Nicolas, 1997), as well as early recombination intermediates, such as ssDNA produced by resection (Blitzblau *et al.*, 2007; Buhler *et al.*, 2007). In addition, crossover and non-crossover recombination outcomes have been identified by genotyping single nucleotide polymorphisms (SNPs) in parents and progeny of sampled meioses (Mancera *et al.*, 2008).

Findings derived from these global studies of meiotic recombination are largely consistent with one another and uncover a number of common factors influencing the positioning

and relative activity of recombination hotspots and coldspots. First, meiosis-specific DSBs occur preferentially between, rather than within, gene coding sequences (Wu and Lichten, 1994; Baudat and Nicolas, 1997; Blitzblau *et al.*, 2007; Pan *et al.*, 2011). Specifically, they involve nucleosome-depleted, DNAse-I sensitive regions of promoters (Wu and Lichten, 1994; Pan *et al.*, 2011), indicating that recombination initiation requires open chromatin structure. Although transcription itself is not required (White *et al.*, 1992), the activity of a recombination hotspot can depend on the binding of a transcription factor (White *et al.*, 1993). Finally, recombination activity is strongly correlated to high GC content (Baudat and Nicolas, 1997; Gerton *et al.*, 2000; Pan *et al.*, 2011), possibly reflecting local DNA packaging and accessibility for the recombination machinery.

A reduction in recombination frequency has been observed in centromere- and telomere-proximal regions (Lambie and Roeder, 1988; Baudat and Nicolas, 1997; Gerton *et al.*, 2000). Centromeres play an essential role during meiotic chromosome segregation and are likely protected from interference with the recombination machinery (Brar and Amon, 2008). Indeed, crossover events near centromeres in yeast, flies and humans have been associated with premature separation of sister chromatids and severe aneuploidy (Koehler *et al.*, 1996a; Lamb *et al.*, 1996; Rockmill *et al.*, 2006), possibly due to the disruption of pericentric cohesion necessary to align homologous chromosomes during the first meiotic division (Rockmill *et al.*, 2006). Similarly, recombination near telomeres, the rDNA locus and other highly repetitive genomic sequences increases the risk of deleterious chromosomal rearrangements and, thus, appears to be actively repressed in these regions (Petes and Botstein, 1977; Su *et al.*, 2000).

Synthetic Genetic Array (SGA) technology provides a novel systematic approach to survey the global landscape of yeast meiotic recombination (Jorgensen *et al.*, 2002). SGA is a high-throughput method designed for constructing yeast double mutants via large-scale mating and sporulation (Chapter 1.3.2). An ordered array of single mutant strains, each carrying a drug-resistance cassette at a specific locus, forms a set of colinear mapping markers covering every 3 kb of the yeast genome (Giaever *et al.*, 2002). Systematic crosses of mutants located on the same chromosome allow estimation of the efficiency of double mutant production and, thus, the frequency of meiotic recombination via colony size measurement of haploid meiotic progeny (Jorgensen *et al.*, 2002; Menne *et al.*, 2007; Costanzo and Boone, 2009).

Genetic linkage defines chromosomal regions where frequency of meiotic recombination is low and loci tend to be inherited jointly. In SGA and other genetic methods based on mating and sporulation, linked gene pairs form small double mutant colonies (Figure 4.2B) and are normally removed from analysis to prevent them from being mistaken for true negative genetic interactions. As a result, the functional relationships between linked gene pairs remain largely unknown. To facilitate the investigation of linked gene pairs, I have developed a computational approach to correct for linkage effects in SGA data and recover negative genetic interactions. I show that interactions recovered from linked gene pairs are functionally informative and provide additional insights into the functional organization of the yeast genome.

In addition, I used genome-wide SGA data to map profiles of genetic linkage along all sixteen yeast chromosomes. These linkage profiles are accurate and reproducible, and reca-

pitulate previously identified recombination patterns, including inhibition of recombination at pericentric chromosomal regions. Interestingly, I found that the extent of centromere-related recombination repression is directly proportional to chromosome length, whereas low recombination regions outside the centromere are, on average, constant in size. These findings corroborate the unique role played by centromeres during meiotic division and suggest a chromosome size-dependent mechanism for ensuring proper chromosome segregation.

4.3 Mapping genetic linkage using SGA

In a typical SGA experiment (Figure 1.4), a "query" strain, carrying a deletion or a conditional mutation in a gene of interest, is crossed to an array of ~4,800 viable deletion mutants (Chapter 1.3.2). Both query and array mutations are linked to drug-resistance cassettes, which enable selection of the double mutant meiotic progeny, following mating and sporulation, and act as markers of meiotic recombination. When the query and the array mutations are located on the same chromosome, recombination is required to produce double mutants (Figure 4.2A). Physical distance between the loci and the local recombination rate determine the efficiency of double mutant production, which, together with double mutant fitness, defines the size of the double mutant colony formed at the final stage of SGA selection (Figure 4.2A). As a result, recombination frequency can be estimated from double mutant colony size, which is measured quantitatively using the SGA data acquisition pipeline described in Chapter 2 (Materials & methods 2.10.1).

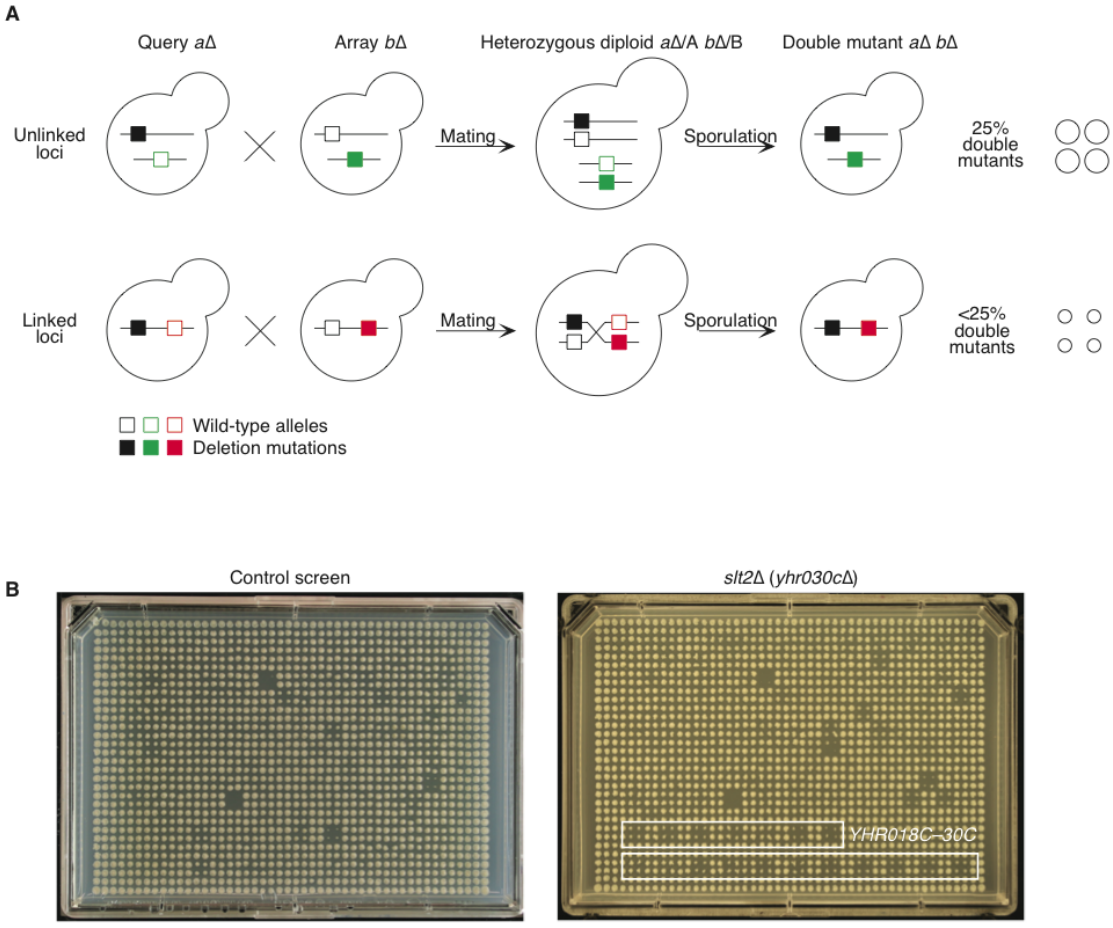


Figure 4.2. Segregation of unlinked and linked loci during meiosis.

(**A**) Unlinked query and array mutations (filled black and filled green squares, respectively) segregate independently, such that on average 25% of all spores contain both mutations. However, when the query and the array mutations are located on the same chromosome (filled black and filled red squares, respectively), the frequency of double mutant spores depends on the frequency of recombination between the loci and is in general lower than 25%, resulting fewer double mutants and smaller colonies.
(**B**) A typical SGA experiment, showing (on the right) a set of array mutants (*yhr018c*Δ–*yhr030c*Δ) linked to the query mutation (*yhr030c*Δ) and thus producing smaller double mutant colonies compared to a control screen (on the left) which uses a genetically independent query mutation.

4.4 Recovering genetic interactions from linked gene pairs

Compared to loci that segregate independently, double mutants involving genetically linked loci result in considerably smaller colonies and, thus, can be mistaken for negative genetic interactions (Figure 4.2B). To avoid introducing false positives, most SGA-based stud-

ies discard gene pairs located within ~50–200 kb from each other, resulting in the loss of ~140 genes per genome-wide screen on average or ~4% of all tested gene pairs (Baryshnikova *et al.*, 2010b; Collins *et al.*, 2010) (Chapter 2.4.6). As a result, linked gene pairs are significantly underrepresented for the number of genetic interactions reported in the literature (2.5-fold underrepresentation in BioGRID, $p < 10^{-21}$; Materials & methods 4.9.1), and their functional relationships are much less well understood.

To examine the genetic interactions among linked gene pairs, I developed a normalization approach to correct for linkage effects and extract fitness information. My strategy is based on the observation that genetic interactions are rare (Chapter 3.6) and the linkage effect is relatively constant within a given genomic region (Chapter 4.6). Consequently, the average colony size of double mutants located in the same genomic region should reflect the strength of the linkage effect in that region, while deviations from average indicate the presence of a genetic interaction (Figure 4.3A; Materials & methods 4.9.2).

I applied this correction strategy to ~1,700 SGA genetic interaction screens, described in Chapter 3, and recovered 7,958 novel negative genetic interactions, which had been discarded by the previous linkage filtering method (Chapter 2.4.6). I focused on detecting negative, rather than positive, genetic interactions as they are less prone to noise and can be identified with higher confidence (Chapter 2.7), especially in the context of genetic linkage where most of the variation in colony size is due to the efficiency of double mutant production and not double mutant fitness. The recovered interactions represent a ~7% increase over the total number of negative genetic interactions reported previously for this dataset (Chapter 3) and show a similar SGA score distribution (Figure 4.3B).

To assess the quality of the recovered genetic interactions, I performed a precision/recall analysis in which I evaluated the number of genetic interactions, at a range of significance thresholds that connected genes annotated to the same Gene Ontology (GO) term or genes encoding physically interacting proteins (Figure 4.3C–D). The precision/recall analysis showed that, similarly to genetic interactions among unlinked gene pairs, genetic interactions recovered from linkage are highly enriched for co-annotations and protein-protein interactions (Figure 4.3C–D) and are therefore predictive of shared function.

While being highly enriched, the frequency of co-annotations among genetically interacting linked genes was ~2-fold lower than the frequency of co-annotations among unlinked gene pairs (Figure 4.3C), suggesting a higher false positive rate among the recovered interactions and advising against their integration with the main genetic interaction dataset without further normalization. To reduce the incidence of false positives in the recovered genetic interactions, I restricted the analysis to only gene pairs that were tested reciprocally (query mutant A-array mutant B and query mutant B-array mutant A) and produced consistent genetic interaction scores. The set of 1,292 genetic interactions that satisfy these criteria overlap with co-annotated gene pairs or protein-protein interactions as often as genetic interactions obtained from unlinked gene pairs (Figure 4.3C–D), thus indicating that filtering for reciprocal AB/BA consistency is a valid strategy for improving the quality of the recovered interactions.

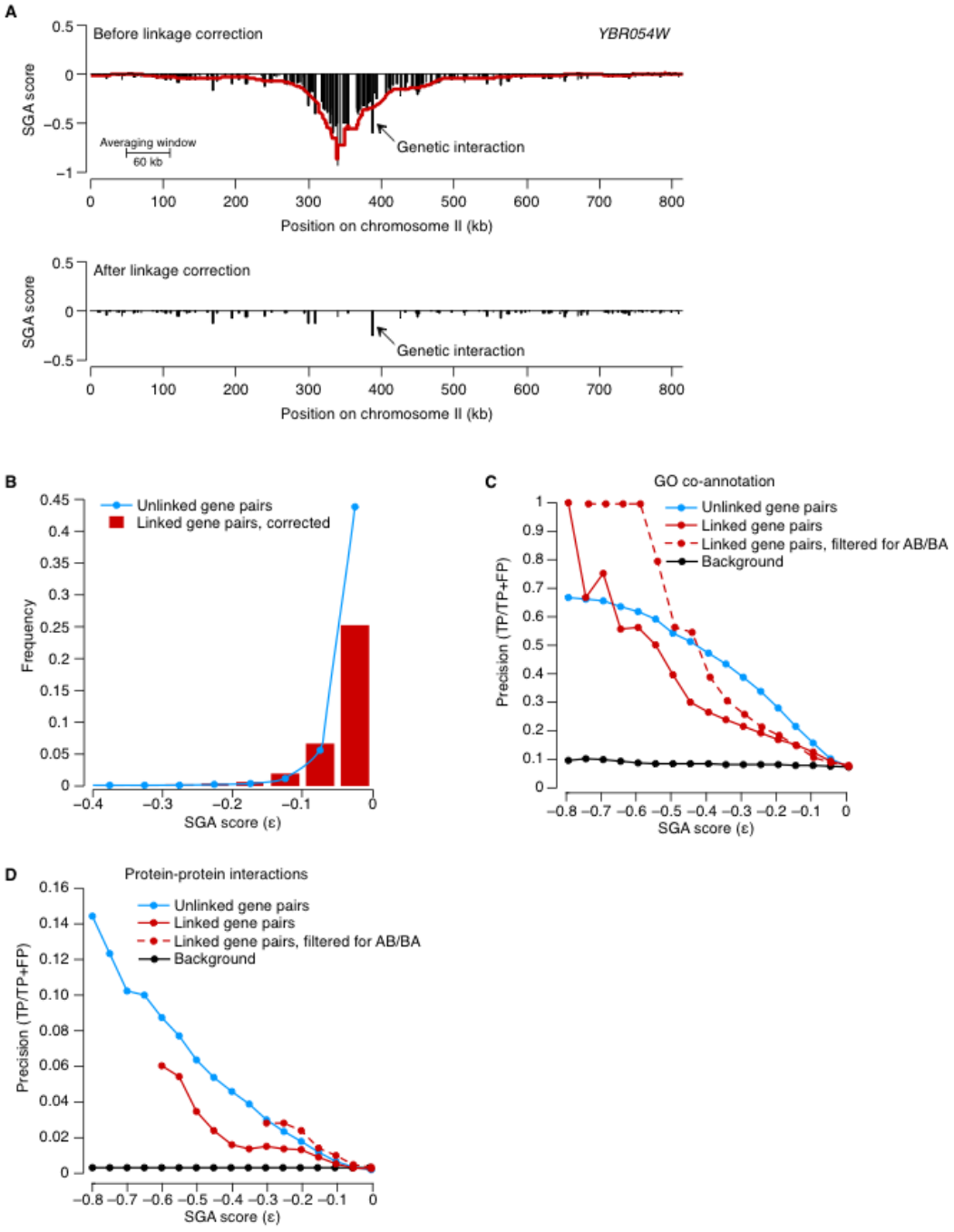


Figure 4.3. Recovering genetic interactions from genetically linked gene pairs.

(**A**) A genetic linkage profile for query mutant *ybr054w*Δ is shown where the average contribution of linkage in a given chromosomal region is indicated by the red line (Materials & methods 4.9.2). After subtracting the contribution of linkage, genetic interactions can be recovered from colony sizes of genetically linked double mutants. Only negative genetic interactions are visualized.
(**B**) Distribution of negative SGA scores for unlinked gene pairs (blue line) and linked gene pairs (red bars) after correction for linkage is applied (Materials & methods 4.9.2).

(**C**) Functional evaluation of the recovered negative genetic interactions against the Gene Ontology (GO) co-annotation standard. A set of gold standard GO terms was chosen using previously described criteria (Myers *et al.*, 2006). True positive interactions (TP) were defined as those involving both genes annotated to the same GO term. Precision and recall were calculated as described in Materials & methods 2.10.2.
(**D**) Functional evaluation of the recovered negative genetic interactions against a set of physical protein-protein interactions. Precision and recall were calculated as described in Materials & methods 2.10.2.

An alternative strategy for reducing false positives is to only retain genetic interactions that are supported by additional lines of evidence. For example, twelve of the 7,958 recovered genetic interactions connected genes encoding members of the same protein complex. Ten of these complexes are essential or contain redundant components and thus are expected to show a negative genetic interaction (Figure 4.4A–I; Chapter 3.9). For example, I recovered a known synthetic lethal relationship between two interchangeable catalytic subunits of the protein phosphatase 2A, Pph21 and Pph22 (Figure 4.4I) (Sneddon *et al.*, 1990). In addition, I detected a novel interaction between putative tyrosine phosphatases *OCA2* and *OCA3* (Figure 4.4J), which are likely to share at least partially redundant functions given that increased dosage of *OCA3* suppresses the growth defect associated with the deletion of *OCA2* (Roma-Mateo *et al.*, 2011). The extensive overlap between protein-protein binding, dosage suppression and negative genetic interactions was noted in a recent study from our laboratory (Magtanong *et al.*, 2011), thus supporting the newly found relationship between *OCA2* and *OCA3*. One additional negative interaction between members of the nonessential Rpd3S complex (*UME1* and *EAF3*; Figure 4.4K) was not supported by previous findings and thus I cannot rule out the possibility that this represents a potential false positive interaction associated with my correction method. Further experimental validation will be necessary to characterize this interaction and determine its biological relevance.

Overall, these findings indicate that the majority of genetic interactions recovered from linked gene pairs are biologically informative and contribute towards a comprehensive coverage of all double mutant combinations in yeast.

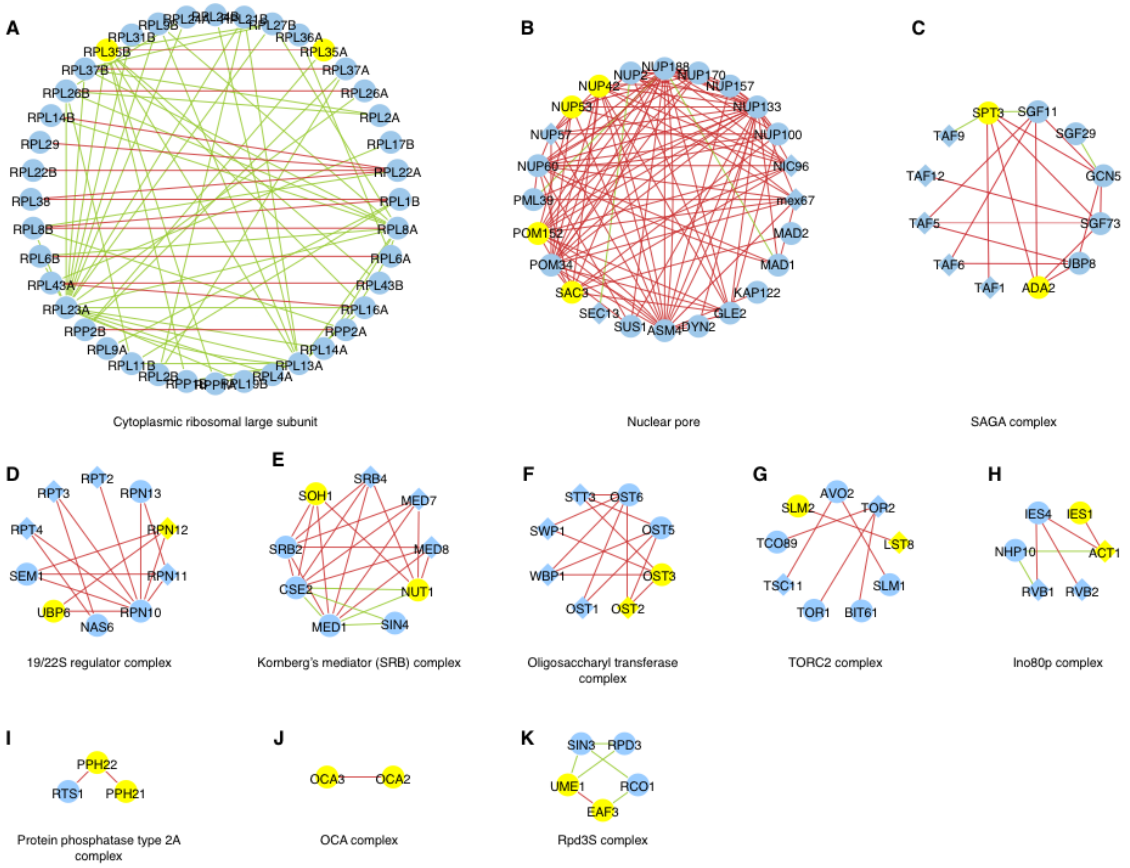


Figure 4.4. Intra-complex negative genetic interactions involving linked gene pairs.

Twelve intra-complex negative genetic interactions were recovered from the set of linked gene pairs (**A**–**K**). Two pairs of linked genes belonged to the same protein complex (**B**), and thus a total of eleven complexes are shown. Red and green edges correspond to negative and positive genetic interactions, respectively. Yellow nodes indicate the gene pairs that were recovered after linkage normalization.

4.5 Estimating genetic distances from colony size data

Recombination rate between two loci is often expressed in terms of their genetic distance, which is measured in centimorgans (cM) and refers to the average number of crossover events occurring between the loci in a single meiosis (Griffiths *et al.*, 2000). Given a population of cells undergoing meiotic division, genetic distance between two loci can be computed using the frequency of recombinant progeny relative to the total number of meiotic products (Griffiths *et al.*, 2000). By definition, a genetic distance of 50 cM corresponds to a 50% frequency of recombinants and thus is indicative of genomic loci that segregate independently. A compilation of 40 years of genetic analyses in *S. cerevisiae* determined that on average 1 kb corresponds to approximately 0.36 cM, and thus most gene pairs located at ~139 kb and farther should be genetically independent [50 cM/(0.36 cM/kb) = 139 kb] (Cherry *et al.*, 1997). In contrast, recombination occurs less frequently in *S. pombe*, in which 1 kb corresponds to 0.18 cM (Munz, 1994; Dixon *et al.*, 2008).

In theory, recombination frequency can also be estimated from SGA experiments by measuring double mutant colony size, which reflects the relative frequency of recombinant meiotic progeny (Chapter 4.3). However, multiple crossovers often generate parental arrangements of alleles that remain undetected during assessment of recombinant frequency and thus result in an underestimate of the genetic distance between loci. A correction method, such as, for example, the Haldane mapping function, must be applied to account for multiple crossover events and obtain more accurate genetic distances (Materials & methods 4.9.1).

We screened ~3,800 query strains against the array of ~4,800 non-essential deletion mutants (Costanzo *et al.*, manuscript in preparation) and obtained double mutant colony sizes for ~1.2 million gene pairs located on the same chromosome. For every gene pair, I used the corresponding single and double mutant colony sizes to estimate the relative fraction of recombinant meiotic products and derive their genetic distance (Materials & methods 4.9.1).

For every gene in our dataset, I generated a genetic linkage profile, defined as the set of genetic distances between that gene and all the other genes located on the same chromosome (Figure 4.5A–B). As expected, genetic distance increased linearly with physical distance (Figure 4.5A), indicating that crossover frequency is relatively constant within a given genomic region and recombination hotspots are rare. Occasionally, genetic linkage profiles on the left and the right hand side of a gene were visibly different (Figure 4.5B), likely reflecting brusque changes in recombination rate in the two regions flanking the gene's locus.

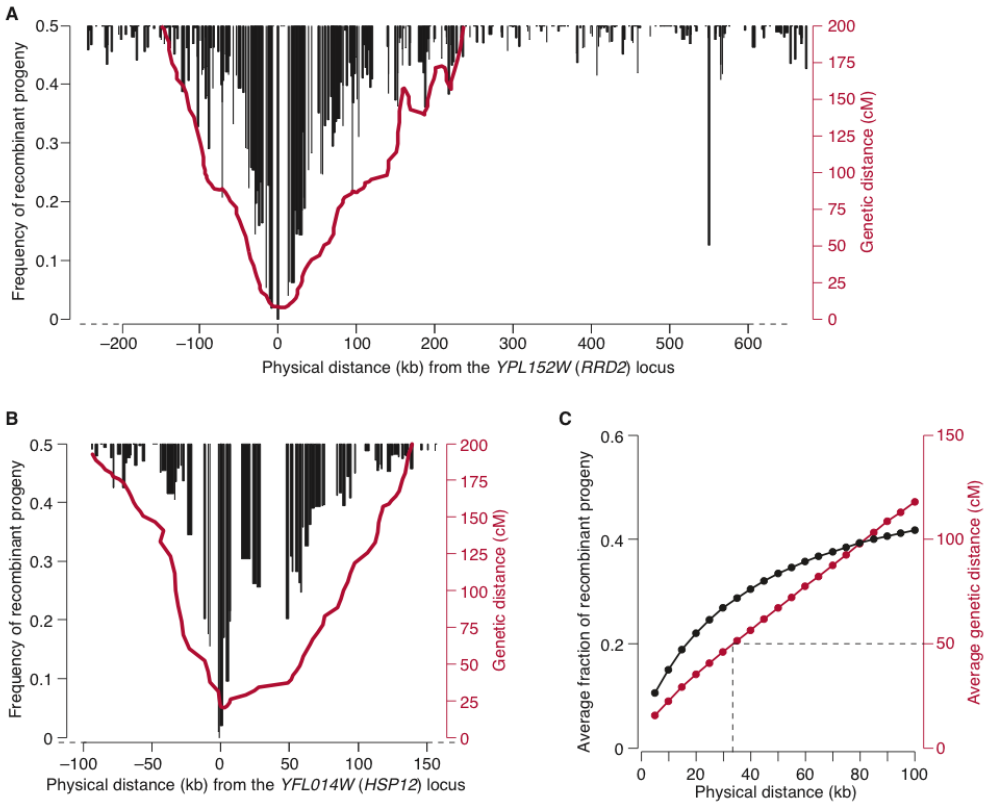


Figure 4.5. Relationship between genetic and physical distances.

(**A**) The genetic linkage profile of the query mutant *ypl152w*Δ (*rrd2*Δ) is shown. Black bars correspond to the frequency of recombinant progeny involving the query mutation and an array mutation located at the position indicated on the x-axis. The red line indicates the corresponding genetic distance, calculated as described in Materials & methods 4.9.3.
(**B**) A genetic linkage profile similar to (**A**) but showing a visible difference between the slope of the left and the right genetic distance estimates relative to the physical distance from the query mutant locus. A trend of this kind might indicate a significant difference in local recombination rates between the regions to the right and to the left of the query.
(**C**) Relationship between physical distance and recombination estimates, as measured by the fraction of recombinant progeny (black line) or genetic distance measured in centimorgans (red line). The dotted line indicates the 50 cM threshold, which corresponds to the shortest genetic distance between two unlinked loci and on average is equivalent to a physical distance of ~35 kb.

My SGA-based analysis suggested that, on average, genes located more than 35 kb apart tended to segregate independently, as their genetic distance was 50 cM or greater (Figure 4.5C). In contrast, early genetic analyses (see above) estimated that unlinked loci are normally positioned at ~139 kb, a distance ~4-fold greater than my estimate of 35 kb, sug-

gesting that meiotic recombination in SGA is occurring at a considerably higher frequency. This hypothesis is consistent with the observation that the selectable *kanMX* cassette, which replaces every array locus in an SGA experiment, increases local frequency of meiotic recombination (Valerie Borde, personal communication) and thus may be responsible for the systematic increase in recombination rate amongst deletion alleles. Despite this limitation, every chromosomal region in our dataset should be equally affected by this phenomenon, which therefore is expected to have a minor impact on the results of my comparative analyses.

While most pairs of unlinked loci were positioned ~35 kb away from each other, the extent of genetic linkage varied considerably across the genome, ranging from ~0 to 400 kb (Figure 4.6A). This variability, however, was not randomly distributed. Genetic linkage profiles of genes located within the same chromosomal region were highly consistent with each other (Figure 4.6B). Indeed, for any pair of loci within a 10 kb distance, the median correlation of their genetic linkage profiles was $R = 0.94$ and the distance to the nearest unlinked locus differed by 5 kb (median; n = 2,407; Materials & methods 4.9.4). Similarly, linkage profiles of query and array strains mutated for the same gene were highly reproducible (Figure 4.6C; median $R = 0.95$; Materials & methods 4.9.4). These observations indicate that variation in SGA-derived genetic linkage profiles reflects the underlying variation in local recombination rates and thus can be used to analyze the distribution of meiotic recombination events across the yeast genome.

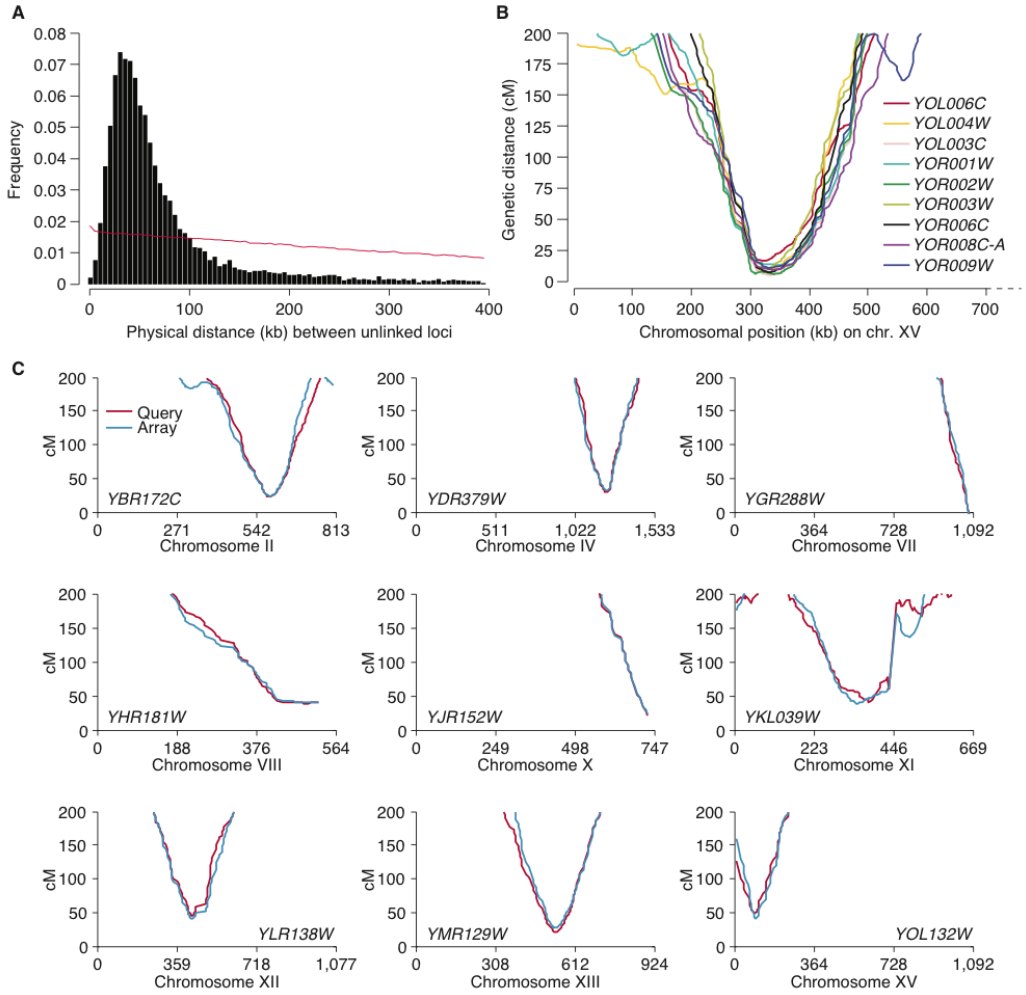


Figure 4.6. Reproducibility of genetic linkage profiles.

(**A**) Distribution of physical distances between pairs of closest unlinked loci across the entire genome.
(**B**) Genetic linkage profiles for 9 consecutive query mutants located on chromosome XV. Genetic profiles of neighboring genes are highly consistent (median $R = 0.94$).
(**C**) Genetic linkage profiles for 9 pairs of query-array mutants involving the same genomic locus. Genetic profiles involving reciprocal double mutants (e.g., query mutant *a*-array mutant *b*, and vice versa, query mutant *b*-array mutant *a*) are highly consistent (median $R = 0.95$).

4.6 Validation of recombination patterns

I constructed a consolidated genetic linkage map for every yeast chromosome by averaging the genetic linkage profiles of query and array mutants for the same genomic locus, as well as profiles of neighboring loci (Figure 4.7; Materials & methods 4.9.5). These genetic

linkage maps revealed a highly structured chromosomal organization, with large blocks of strong linkage separated by shorter regions of unlinked loci, likely corresponding to recombination hotspots (Figure 4.7).

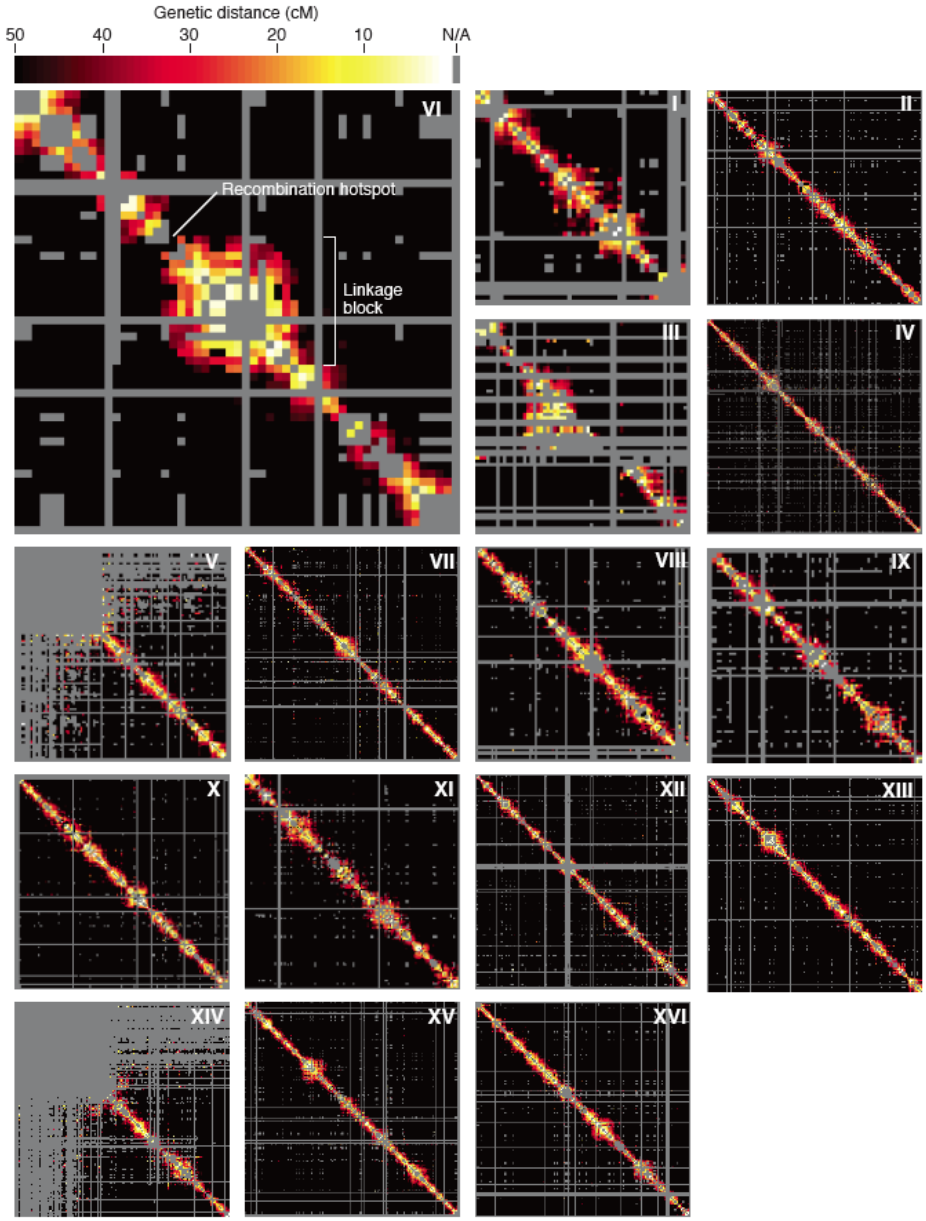


Figure 4.7. Genetic linkage maps for all yeast chromosomes.

A consolidated genetic linkage map for each chromosome was constructed as described in Materials & methods 4.9.5) and visualized as a heatmap, where the x- and the y-axis correspond to chromosomal positions and the color at the intersection of two positions corresponds to the average genetic distance between gene pairs located

in those regions of the chromosome.

I identified 222 putative recombination hotspots defined as local peaks of recombination frequency, measured in centimorgans per kb, that were higher than the average recombination frequency for a given chromosome (Materials & methods 4.9.6). The distribution of these peaks coincided well with known genomic loci associated with high levels of meiotic recombination (Lichten and Goldman, 1995). For example, 8 out of 9 most characterized recombination hotspots were identified in my analysis (*HIS2*, *HIS4*, *ARG4*, *CYS3*, *DED81*, *ARE1–IMG1*, *CDC19* and *THR4*; Figure 4.8).

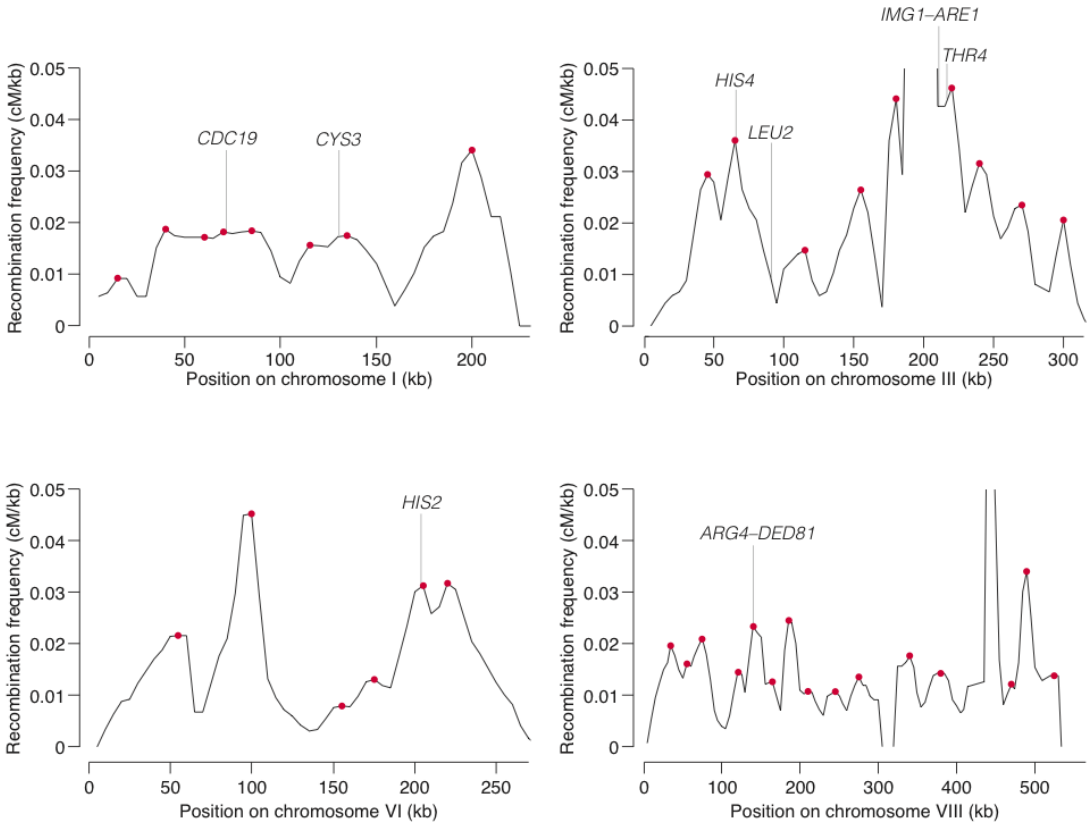


Figure 4.8. Validation of known recombination hotspots.

Profiles of recombination frequency for chromosomes I, II, VI and VIII were computed as described in Materials & methods 4.9.6, and plotted relative to the position on the corresponding chromosome. Red dots represent local maxima of recombination frequency. The positions of the nine characterized recombination hotspots are

indicated by vertical lines.

Meiotic recombination in yeast has been investigated on a genome-wide scale using a variety of experimental approaches (Baudat and Nicolas, 1997; Borde *et al.*, 2004; Blitzblau *et al.*, 2007; Buhler *et al.*, 2007; Mancera *et al.*, 2008). I compared the recombination profiles obtained from these studies to those derived from our SGA-based genetic linkage analysis, and found them to be highly consistent (Figure 4.9).

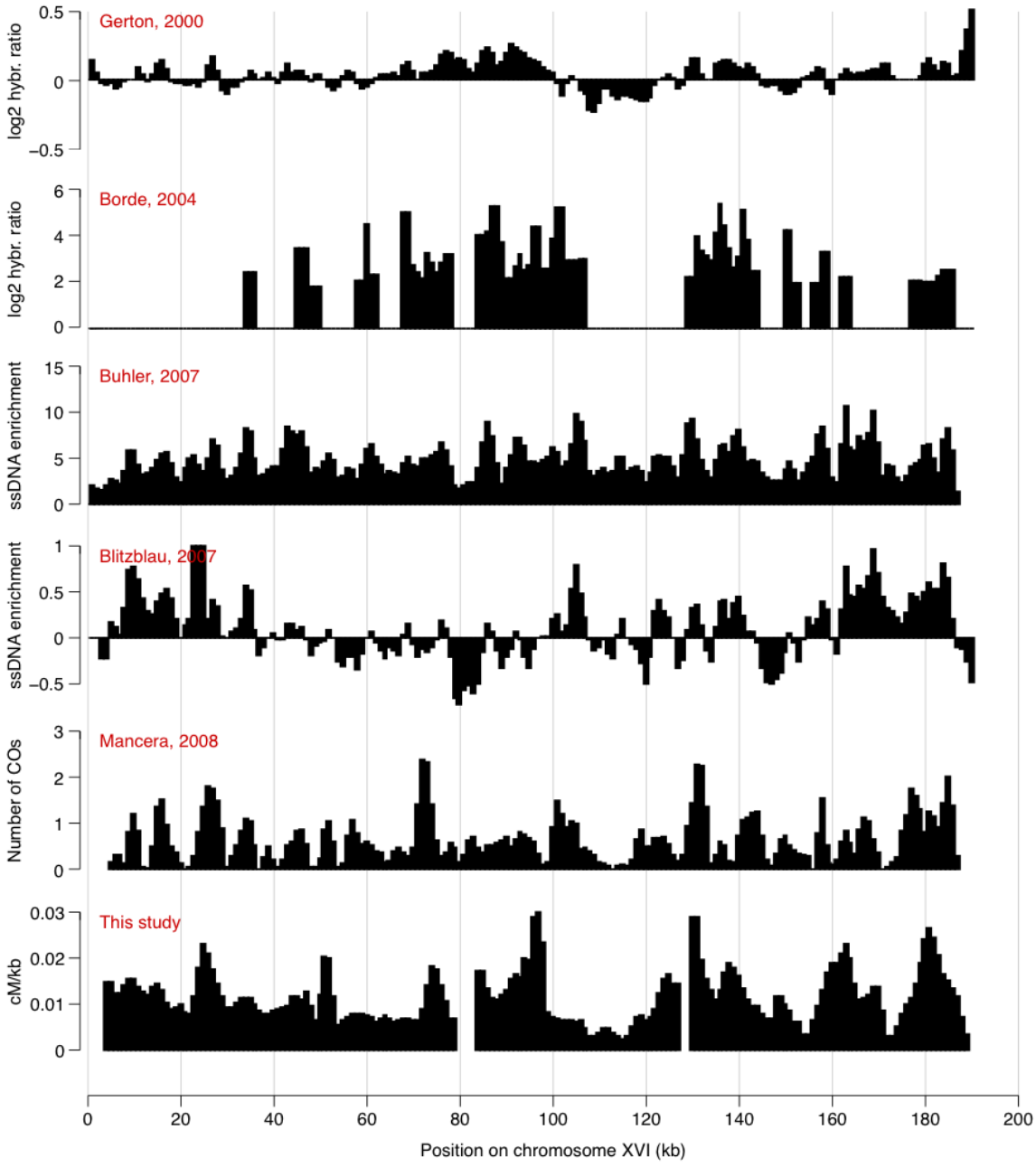


Figure 4.9. Comparison of recombination profiles across various experimental approaches.

Profiles of recombination frequency for chromosome XVI were derived using six different approaches: mapping of Spo11 (Gerton *et al.*, 2000) and Mre11 (Borde *et al.*, 2004) binding, analysis of ssDNA-enriched regions (Blitzblau *et al.*, 2007; Buhler *et al.*, 2007), SNP mapping (Mancera *et al.*, 2008) and SGA-derived genetic linkage analysis (this study). Data relative to the published studies were obtained from the corresponding supplementary datasets.

On average, 61% of local recombination peaks identified in my analysis were located within or adjacent to a recombination peak identified in a previously published study (Figure 4.10; Materials & methods 4.9.6). This overlap is highly significant ($p < 0.01$) and comparable to the overlap observed between other published datasets (Figure 4.10), thus supporting the utility of the SGA-based approach for examining genetic linkage.

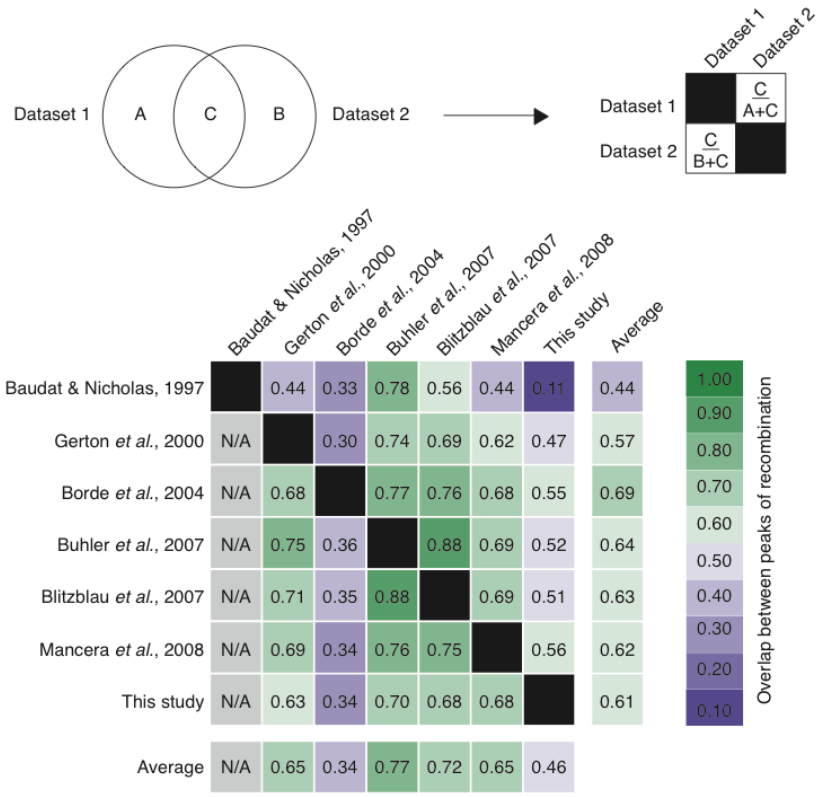


Figure 4.10. Overlap between recombination peaks across various experimental approaches.

Local peaks of recombination frequency were mapped for all the indicated datasets. The overlap between any two datasets is visualized as a heatmap/table, where colors and numbers reflect the fraction of peaks identified by study A (listed on the y-axis) that are located within 15 kb from a peak identified by study B (listed on the x-axis). For example, of all the peaks identified by Gerton *et al.* (y-axis), 30% were also identified by Borde *et al.* (x-axis), and, vice versa, of all the peaks identified by Borde *et al*. (y-axis), 68% were also identified by Gerton *et al.* (x-axis).

Furthermore, I detected 46% of all recombination hotspots reported in other studies, which, in turn, recall 34–77% of each other's findings (Figure 4.10). Our lower sensitivity, which ranks fifth among the six analyzed studies, likely stems from the lower resolution of our experimental design, which is based on systematic crosses of gene deletion mutants and, as a result, can only map recombination events between genomic loci, most of which are located at an average distance of ~3 kb.

Similarly to hotspots, my analysis also identified known recombination coldspots. As expected, a large ~85 kb block of strong genetic linkage corresponded to the rDNA locus on chromosome XII (Figure 4.7, panel XII; Figure 4.11A). Moreover, linkage blocks between 60 and 150 kb in size overlapped the pericentromeric regions of each chromosome (Figure 4.7, panels I–XVI; Figure 4.11B). Recombination frequency near telomeres could not be quantified in our data due to the low gene density in the corresponding regions.

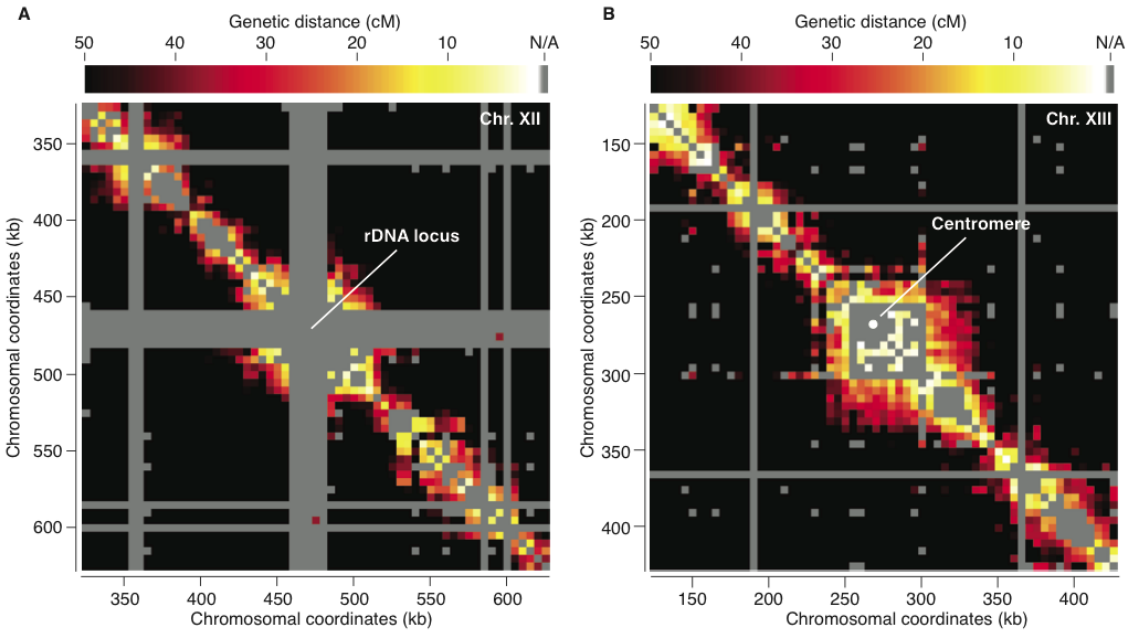


Figure 4.11. Coldspots of recombination.

(**A**) A section of the genetic linkage map of chromosome XII, highlighting the linkage block surrounding the rDNA locus.
(**B**) A section of the genetic linkage map of chromosome XIII, highlighting the linkage block surrounding the centromere.

4.7 Characterization of genetic linkage patterns in pericentric regions

Centromeres are chromosomal regions that mediate the attachment of chromosomes to spindle microtubules through large protein structures known as kinetochores. In *S. cerevisiae*, unlike other organisms, centromeres are defined by a single highly conserved ~125 bp sequence, with little, if any, repetitive DNA surrounding it (Fitzgerald-Hayes *et al.*, 1982a; Fitzgerald-Hayes *et al.*, 1982b). Numerous studies have reported a reduction of recombination frequency within pericentric regions in yeast (Lambie and Roeder, 1988; Baudat and Nicolas, 1997; Gerton *et al.*, 2000; Borde *et al.*, 2004; Buhler *et al.*, 2007; Chen *et al.*, 2008; Mancera *et al.*, 2008) and in higher organisms (Hassold *et al.*, 1996; Koehler *et al.*, 1996b).

While the mechanism of repression is still poorly understood, it is clear that recombination events near centromeres are extremely deleterious, often result in catastrophic missegregation phenotypes and are the primary cause of severe human genetic defects, such as the Down syndrome (Sherman *et al.*, 2006) (Chapter 4.2).

Global analyses of meiotic recombination in yeast found that DSBs, which initiate recombination, are significantly underrepresented within 8–20 kb surrounding centromeres, although the exact extent of repression is dependent on the strain genetic background and the experimental approach (Gerton *et al.*, 2000; Borde *et al.*, 2004; Buhler *et al.*, 2007). Interestingly, our SGA-derived data shows that the reduction of inter-homolog recombination events extends much farther than 8–20 kb from the centromere, suggesting that centromere-proximal DSBs are repaired via alternative pathways. In SGA, genes located on opposite arms of a chromosome, must be separated by at least 40–90 kb in order to segregate independently (Figure 4.12A,C). This distance is 15–160% higher than the overall genomic average for unlinked loci (~35 kb, Figure 4.5, Figure 4.12B,C), reflecting the strong negative effect on recombination frequency exerted by centromeres.

In addition, I found that the variation in the extent of centromere-related linkage showed a strong correlation with variation in chromosome length ($R = 0.77$, Figure 4.12C), a trend that has not been noted previously. This correlation was specific to linkage involving the centromere because distance between unlinked loci positioned on the same arm of the chromosome was generally constant with respect to chromosome length and matched the overall genomic average of ~35 kb (Figure 4.12B–C). The only exceptions were the four

smallest chromosomes (I, III, VI and IX), which are known to have particularly high levels of recombination and therefore show smaller genetic linkage effects (Figure 4.12B,C).

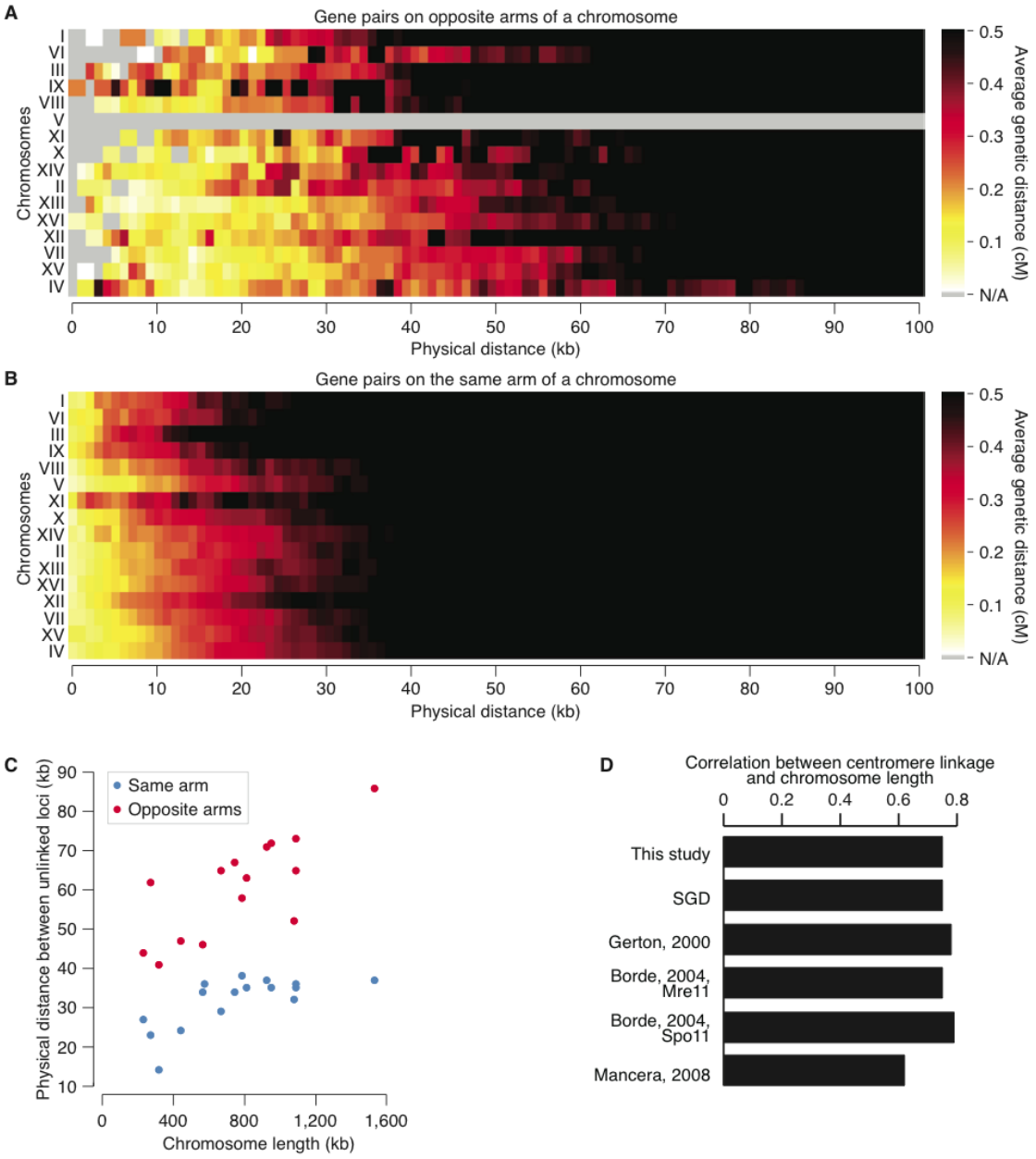


Figure 4.12. Relationship between linkage and chromosome size.

(**A–B**) The distribution of genetic distances for gene pairs located at the indicated physical distances is shown. Gene pairs located on the same chromosome (y-axis) were grouped according to their physical distance (x-axis). Within each group, genetic distances were averaged and visualized as a heatmap, where brighter colors correspond to shorter genetic distances and stronger linkage. Gene pairs located on opposite arms of a chromosome (**A**) or on the same arm of the chromosome (**B**) were analyzed separately.

(**C**) Physical distance at which gene pairs were on average unlinked (50 cM) was estimated from (**A**) and (**B**) for each chromosome, and plotted against chromosome length.
(**D**) Pearson correlation was computed between chromosome length and centromere-related recombination, as defined by the shortest physical distance between loci unlinked from the centromere (This study, SGD) or distance between centromere-adjacent hotspots of meiotic recombination (Gerton *et al.*, 2000; Borde *et al.*, 2004; Mancera *et al.*, 2008).

The analysis of other published datasets strongly supports the patterns of centromere-related recombination uncovered in SGA data (Figure 4.12D). For example, genetic data from the *Saccharomyces* Genome Database (SGD; (Cherry *et al.*, 2012)), which were compiled from classical tetrad analysis carried out by a number of different studies, indicate that the shortest distance between two unlinked loci bridging the centromere is larger for longer chromosomes ($R = 0.77$, Figure 4.12D). Similarly, the distance between centromere-adjacent DSB hotspots, as measured by increased Mre11 or Spo11 binding (Gerton *et al.*, 2000; Borde *et al.*, 2004), correlates with chromosome length ($R = 0.77–0.8$; Figure 4.12D). The consistency of this trend across several independent datasets is a strong indication of its robustness and suggests that indeed centromere-related recombination is established by an underlying mechanism involving chromosome size.

4.8 Discussion

An explanation for the unique role of centromeres in determining local rates of meiotic recombination may be related to the structural organization of pericentric regions during various phases of meiotic division (Brar and Amon, 2008). In meiosis, one round of DNA replication is followed by two consecutive rounds of chromosome segregation, in which homologous chromosomes separate from each other during meiosis I and sister chromatids separate during meiosis II. Protein complexes, known as cohesins, form a heterotetrameric ring

around sister chromatids, thus ensuring that they remain attached during meiosis I and are released in a timely manner during meiosis II (Figure 4.13A). Cohesins are normally distributed along the entire length of the chromosome, but are removed from chromosome arms at the metaphase I-anaphase I transition to facilitate the resolution of inter-homolog chiasmata that mediate recombination and link homologous chromosomes at metaphase I (Buonomo *et al.*, 2000) (Figure 4.13A). Cohesins bound to the centromeres, however, are maintained beyond anaphase I and remain attached to the centromeres until the metaphase II-anaphase II transition, when sister chromatid segregation occurs (Miyazaki and Orr-Weaver, 1994) (Figure 4.13A). It is therefore possible that the longer persistence of cohesin at pericentric regions causes a reduction in the number of successfully completed recombination events and the formation of larger genetic linkage groups in those regions.

The formation of large cohesin clusters at centromeres is thought to be necessary to resist the mechanical tension generated by microtubules bound to sister kinetochores and to prevent premature chromatid separation (Brar and Amon, 2008). It is not known how chromosome length relates to the resistance a chromosome must oppose to microtubule action, but the presence of larger cohesin clusters around the centromeres of larger chromosomes might play a role in this process. In at least one study of cohesin binding (Glynn *et al.*, 2004), the extent of cohesin clustering at pericentric regions of meiotic chromosomes shows a significant correlation to chromosome length ($R = 0.51$; Figure 4.13B); although this trend does not hold for other studies (Kiburz *et al.*, 2005) and thus whether it is true or not remains inconclusive.

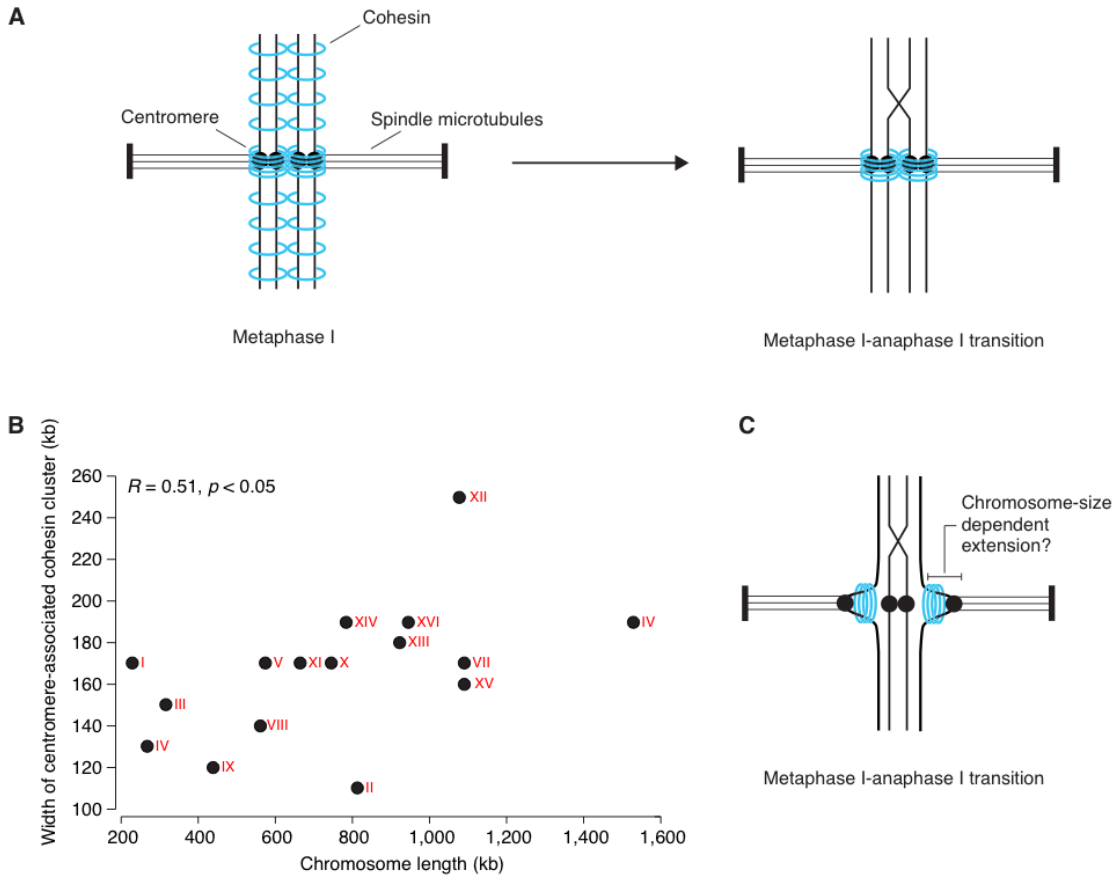


Figure 4.13. Cohesin-dependent linkage formation around centromeres.

(**A**) Cohesin protein complexes (blue rings) are normally distributed along the entire length of the chromosome, but they are removed from chromosome arms at the metaphase I-anaphase I transition to facilitate the resolution of inter-homolog chiasmata. The persistence of cohesin binding to centromeres might explain the formation of larger linkage groups in the corresponding regions.
(**B**) The width of centromere-centered cohesin clusters was measured using data from (Glynn *et al.*, 2004) and plotted against chromosome length.
(**C**) Pericentric DNA is known to extend towards cell poles under the action of microtubule-generated tension. Such extension might be more pronounced in larger chromosomes and thus explain the relationship between chromosome length and the extent of centromere-related linkage.

In addition, recent findings indicate that in *S. cerevisiae*, similarly to other organisms, mitotic centromeres and their flanking DNA form large intra-molecular loop structures that extend from the chromosome axes towards the cell poles, driven by the tension of spindle microtubules (Yeh *et al.*, 2008) (Figure 4.13C). These loop structures are stabilized by intra-

chromosomal cohesin links and, in the case of chromosome III, involve ~25 kb of pericentric DNA (Yeh *et al.*, 2008) (Figure 4.13C). While the size of these loops has not been quantified yet for other chromosomes nor have similar analyses been performed during meiosis, it is possible that tension mechanisms separate sister chromatids to a degree that is proportional to chromosome size and the loop structures formed in this manner prevent recombination from occurring within pericentric regions of varying length.

An intriguing hypothesis is that the *S. cerevisiae* loop-forming centromeres and the associated chromosome size-dependent repression of recombination may reflect a common mechanism for chromosome segregation shared by all eukaryotes, despite drastic differences in their centromere structures (Brar and Amon, 2008). Human centromeres, composed by 0.3–5 Mb of highly repetitive α-satellite DNA sequences, correlate in size with their respective chromosomes (Sanchez *et al.*, 1991; Martorell *et al.*, 2000) and, similarly to *S. cerevisiae*, elongate towards cell poles during mitosis (Zinkowski *et al.*, 1991). A positive correlation between centromere size and chromosome length is also typical of several plant species (Bennett *et al.*, 1981; Jenkins and Bennett, 1981). Interestingly, this trend may also extend to *C. elegans* chromosomes, which lack monocentric centromeres and make contact with spindle microtubules throughout the entire lengths of their chromosomes (Albertson and Thomson, 1982). A recent analysis uncovered an extensive low recombination domain centered in the middle of the each worm chromosome (Rockman and Kruglyak, 2009). The extent of this area is directly proportional to chromosome length (Rockman and Kruglyak, 2009) and therefore may reflect the influence of a segregation mechanism similar to *S. cerevisiae*, plants and humans.

In summary, the analysis presented in this Chapter uses SGA-derived genetic linkage data to uncover a previously unappreciated trend between chromosome length and centromere-related frequency of meiotic recombination. Despite the uniqueness of yeast centromere structure, our data, together with other reported evidence, suggest the existence of a common eukaryotic mechanism regulating inter-homolog recombination within pericentric regions, potentially responsible for proper chromosome segregation and accurate transmission of genetic information to the progeny.

4.9 Materials & methods

4.9.1 Overlap with previous genetic interactions reported in BioGRID

Genetic interactions were downloaded from BioGRID (www.thebiogrid.org) on April 18, 2012, and analysis was restricted to the following data types: negative genetic, phenotypic enhancement, synthetic growth defect, synthetic lethality. Underrepresentation for genetic interactions among linked gene pairs was evaluated using the following fold change and significance hypergeometric p-value calculations:

$$fold = \frac{X}{N}\frac{M}{K} \qquad p = \sum_{i=0}^{X} \frac{\binom{K}{i}\binom{M-K}{N-i}}{\binom{M}{N}} \qquad \textbf{Eq. 4.1}$$

where M is the total number of unlinked gene pairs in the dataset, K is the number of unlinked gene pairs that have a genetic interaction in BioGRID, N is the total number of

linked gene pairs in the dataset and X is the number of linked gene pairs that have a genetic interaction in BioGRID.

4.9.2 Recovering genetic interactions from genetically linked gene pairs

The contribution of linkage to the colony size of a double mutant *ab* was calculated as the mean colony size of all double mutants, sharing the same query mutation *a* and having the array mutation located within 60 kb from *b*. This estimated contribution was then subtracted from each double mutant's colony size to produce a corrected estimate of double mutant fitness. Genetic interaction scores were recalculated as the difference between the corrected double mutant fitness and the product of the two single mutants fitnesses. Confidence p-values relative to all genetic interactions were recomputed accordingly (Materials & methods 2.4.9).

4.9.3 Calculation of genetic distances

To quantify the extent of genetic linkage, I estimated the relative fraction of recombinant progeny for each double mutant as:

$$R = \frac{1}{2}\frac{f_{ij}}{f_i f_j} \qquad \textbf{Eq. 4.2}$$

where f_i, f_j and f_{ij} are the colony size-based fitnesses of the two single and of the double mutant, respectively, calculated as described in Chapter 2.4. The idea is that, in the absence of genetic linkage, just like in the absence of genetic interactions (Chapter 1.1), the fitness of a double mutant should equal the product of the two single mutant fitnesses. Therefore, the

relative size defect of the double mutant colony (f_{ij}) compared to expectation ($f_i f_j$) should relate to the effect of genetic linkage or genetic interactions. Because genetic interactions are rare and do not normally depend on physical distance between loci, their contribution to the shape and the width of genetic linkage profiles can be considered negligible. The ratio of observed and expected double mutant fitness is multiplied by a factor of 0.5 such that $R = 0.5$ for genetically independent loci ($f_{ij} = f_i f_j$).

Recombinant fraction R counts the number of recombinant spores relative to all spores generated by meiosis, and can be used as a measure of recombination among genomic loci. However, R often underestimates the true recombination rate as it does not account for recombination events that do not produce double mutants. For example, multiple crossover events result in a parental arrangement of alleles and the resulting meiotic products do not contribute to recombinant frequency.

Several normalization methods have been proposed to correct for this inaccuracy (Morton, 2004). One of the simplest methods, known as Haldane's map function (Haldane, 1919), is based on the assumption that recombination events are distributed randomly within a given region and thus the probability of observing i recombination events follows the Poisson distribution:

$$f(i) = \frac{e^{-m} m^i}{i!} \qquad \textbf{Eq. 4.3}$$

where m is the mean number of recombination events in the region per meiosis.

Recombinants make up 50% of the spores derived from meioses in which at least one recombination event occurred in the region. Thus, recombinant fraction *R* can be expressed as:

$$R = \frac{1}{2}\left(1 - e^{-m}\right) \qquad \textbf{Eq. 4.4}$$

where e^{-m} is the number of meioses with zero recombination events:

$$f(0) = \frac{e^{-m}m^{0}}{0!} = e^{-m} \qquad \textbf{Eq. 4.5}$$

From Eq. 4.4, we can derive *m*, which would equal to the probability of observing a recombination event in a given region:

$$m = -\ln(1 - 2R) \qquad \textbf{Eq. 4.6}$$

Recombination rate between loci is often expressed in terms of their genetic distance, measured in Morgans, such that 1 Morgan equals 1 recombinant product per meiosis. Since each recombination event produces two recombinant products, genetic distance *M* is derived from Eq. 4.6 as:

$$M = \frac{1}{2}m = -\frac{1}{2}\ln(1 - 2R) \qquad \textbf{Eq. 4.7}$$

1 centimorgan equals 0.01 Morgans.

4.9.4 Reproducibility of genetic linkage profiles

A genetic linkage profile is defined as the set of genetic distances between a given gene and all other genes located on the same chromosome. Genetic linkage profiles for query and array mutants involving the same gene were computed independently and their similarity was quantified using Pearson correlation coefficient. The same approach was used for neighboring gene pairs. In addition, the distance between every gene and its closest unlinked partner was measured by computing a smoothed genetic linkage profile and identifying chromosomal positions where the smoothed linkage profile reaches 50 cM.

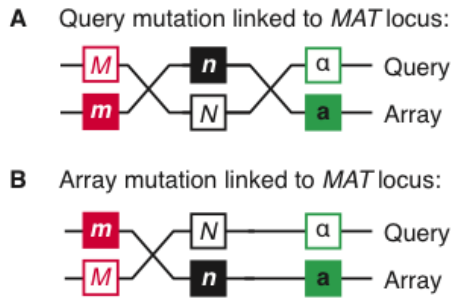
4.9.5 Construction of consolidated genetic linkage maps

Each chromosome was split into a set of consecutive non-overlapping bins of 5 kb each. Query and array genes were assigned to bins based on their chromosomal position, as reported by the *Saccharomyces* Genome Database (SGD, www.yeastgenome.org, accessed in February, 2011). A genetic linkage map G for chromosome C was generated as a $n \times n$ matrix, where n is the total number of bins in C, and the value G_{ij} corresponds to the average of genetic distances for query-array gene pairs where queries belong to bin i and arrays belong to bin j. The consolidated genetic linkage map G' was calculated by averaging query-array and array-query genetic linkage maps for the same positions:

$$G' = \frac{G + G^T}{2} \qquad \textbf{Eq. 4.8}$$

Averaging of query and array genetic linkage maps was applied to all chromosomes, with the exception of chromosome III, because of its peculiar role in the SGA selection pro-

cess. Chromosome III carries the yeast mating type locus (*MAT*), which determines the sex of a yeast haploid cell and distinguishes array and query mutants, which are *MAT***a** and *MAT*α, respectively. Since SGA selects specifically for *MAT***a** meiotic progeny, a query mutant *n* (filled black square in part A of the figure below), located between gene *M* (empty red square) and the *MAT* locus (empty green square), requires a double recombination event to acquire both the array mutation *m* (filled red square) and the *MAT***a** information (filled green square). The array mutant *n*, however, (filled black square in part B of the figure below), being already linked to *MAT***a**, only requires one recombination event to acquire query mutation *m* (filled red square). As a result, query and array mutants linked to the *MAT* locus show different degrees of linkage to nearby genes, and averaging their genetic linkage maps would conceal the true recombination activity occurring within the region.



4.9.6 Hotspot analysis

Using the consolidated genetic map for each chromosome (Materials & methods 4.9.5), I estimated the recombination frequency at each position x along the chromosome by computing the average genetic distance within the interval $[x - 2, x + 2]$, corresponding to a 25 kb interval centered on position x (5 bins × 5 kb/bin). The resulting profile of recombination frequency was smoothed using a moving average filter over 5 consecutive positions.

5 Conclusions, discussion & future directions

5.1 Thesis summary

Mapping all pair-wise genetic interactions in an organism is essential for understanding functional connections between genes and their role in determining cellular phenotypes. In this thesis, I described the construction and the analysis of the genome-wide genetic interaction network for budding yeast *Saccharomyces cerevisiae*, which covers ~30% of all possible gene-gene combinations and represents the largest genetic interaction network to date. I also showed that systematic analysis of yeast double mutants enables investigation of genetic linkage on a global scale and uncovers novel patterns of meiotic recombination.

Large-scale analysis of genetic interactions in yeast is made possible by the Synthetic Genetic Array (SGA) technology, which generates double mutants by mating and sporulation of yeast single mutant arrays. Similar to other genomics technologies, SGA suffers from several systematic experimental biases, which complicate colony size measurements and interfere with our ability to obtain accurate estimates of fitness. In Chapter 2, I described a novel computational method, an SGA scoring system (Baryshnikova *et al.*, 2010b), for identifying and removing these systematic experimental biases. I applied the SGA score to a set of control SGA screens and generated a comprehensive catalog of colony-size based single mutant fitness, which rivals the accuracy of other published datasets and serves as a reference for numerous genomic analyses. Next, I applied the SGA score to double mutant colony sizes to derive quantitative fitness-based genetic interactions, and showed that their quality and gene function prediction capacity provide a several fold improvement over previously available methods.

In Chapter 3, I presented the large-scale application of the SGA score to ~1,700 genome-wide SGA screens that resulted in a quantitative genetic interaction network based on 5.4 million yeast double mutants (Costanzo *et al.*, 2010). I developed a novel network-based approach for visualizing large-scale genetic interaction data that led to an unprecedented view of the functional organization of the cell. In this view, genes participating in similar biological processes clustered together in coherent subsets, which in turn mapped next to each other if they shared numerous functional interconnections. We discovered several physiological and evolutionary gene properties that characterize genetic interaction hubs, and analyzed the relationship between genetic and protein-protein interaction networks, particularly those involving stable protein complexes. Indeed, the strong relationships identified in our study motivated the development of computational models for predicting genetic interaction hubs in complex organisms that are less amenable to high throughput genetic analyses (Koch *et al.*, submitted). In addition, we used genetic suppression, a type of positive genetic interactions, to connect protein complexes whose loss compensates for the disruption of other complexes, and constructed the first global network of genetic suppression (Baryshnikova *et al.*, 2010b). I also integrated genetic and chemical-genetic interaction networks to identify a potential molecular target of a novel chemical compound.

In Chapter 4, I showed that, in addition to genetic interactions, systematic SGA experiments produce a genome-wide genetic linkage map for *S. cerevisiae*, which can be used to estimate local frequencies of meiotic recombination on a global scale. I demonstrated that the SGA-derived genetic linkage maps are accurate in recapitulating previously identified recombination patterns, including inhibition of recombination at pericentromeric chromosomal

regions. I found that the extent of centromere-related recombination repression is directly proportional to chromosome length. The relationship between chromosome size and genetic linkage may possibly result from a chromosome size-dependent mechanism that counteracts the tension generated by microtubules during meiosis I, and highlights SGA methodology as a unique approach for systematic analysis of the global landscape of yeast meiotic recombination.

5.2 Next steps in genetic interaction analysis

As we approach the completion of all pair-wise genetic interactions in yeast, we can envision what needs to be done next to further advance our understanding of the yeast functional organization. Our next steps should make the best use of the available data and put forward new experimental and analytical strategies to fill in the gaps still remaining in our knowledge.

5.2.1 Decoding genetic redundancy

Similarly to individual gene mutations, the vast majority of yeast double mutants (~98%) do not show any fitness defect (Chapter 3.6), suggesting that many more genetic interactions remain to be uncovered. For example, some genes may be buffered at a much higher level of redundancy or be inactive under the investigated growth conditions. Moreover, the effect of a genetic interaction may be manifested through a phenotype other than fitness and thus remain undetected in our standard SGA analysis. Examining each of these scenarios in a systematic manner is necessary to build truly comprehensive genetic interaction

networks, uncover the full spectrum of yeast genetic robustness and draw precise functional connections between all genes in the genome.

5.2.1.1 Condition specific network analysis

The yeast deletion collection has been exposed to hundreds of environmental and chemical growth conditions, revealing that virtually every gene is required for optimal growth in at least one of the tested environments (Hillenmeyer *et al.*, 2008). Similarly, we expect a much higher frequency of genetic interactions when cells are grown in conditions that require normally dispensable genes and pathways. Indeed, treatment with sublethal concentrations of the DNA damaging agent methyl methanesulfonate (MMS) nearly doubles the number of genetic interactions detected among members of DNA repair and recombination pathways (St Onge *et al.*, 2007; Bandyopadhyay *et al.*, 2010), whereas growth in high salt concentration or other stress conditions increases interaction rate among duplicate pairs by ~16% (Musso *et al.*, 2008).

Expanding the analysis of condition-specific genetic interaction networks to a genome-wide scale is a challenging task, given the high number of double mutant combinations and the near infinite range of testable growth environments. It is therefore necessary to develop radically new genetic mapping technologies that would enable rapid assessment of double mutant growth in multiple parallel conditions. Analogously to the single mutant collection, a library of yeast double mutants would greatly facilitate this task. Such library, constructed via SGA, dSLAM or GIM approaches (Chapter 1.3.2), could be stored in a pooled format (for example, ~6,000 pools of ~6,000 double mutants each) and examined under various growth

conditions using large-scale barcode analysis, made affordable by recent advances in high-throughput sequencing (Smith *et al.*, 2009). Alternatively, the double mutant collection can be stored as a single pool of ~36 million mutants, although this approach would require novel strategies for identifying each double mutant within a mixed population, for example by stitching together barcodes associated with the deletion of each locus (Roth *et al.*, 2009).

The scope of condition-specific genetic network analysis could be greatly reduced by prioritizing genetic interaction screens based on their expected functional outcome. We showed, for example, that in at least one growth condition the genetic interaction degree of a query mutant is highly correlated to its single mutant fitness (Chapter 3). As a result, single mutant fitness can be used as a robust predictor of genetic network density under any condition of interest. Specifically, mutants presenting condition-specific fitness defects are the most likely to produce condition-specific genetic interaction networks and thus provide unique insights into the structure of yeast buffering systems.

5.2.1.2 Higher order genetic interactions

In addition to condition-specific genetic interaction networks, yeast robustness to perturbations must be further explored using higher order mutant combinations, which involve three or more genes. Complex genetic interaction networks are key for modeling natural genotype-to-phenotype relationships, where numerous genetic variants interact in determining a particular phenotype (Hartman *et al.*, 2001; Zuk *et al.*, 2012). Moreover, multi-locus genetic interactions uncover the functional roles of highly redundant gene groups, such as those derived from recent gene duplication events (Tong *et al.*, 2004). The analysis of our pair-wise

genetic interaction network showed that duplicate genes exhibit fewer genetic interactions with the rest of the genome compared to non-duplicated genes, likely reflecting the ability of each duplicate to compensate for the absence of its partner (Costanzo *et al.*, 2010; VanderSluis *et al.*, 2010). Pilot triple mutant experiments, using double mutant query strains deleted for both duplicates, uncovered numerous novel genetic interactions, which were specific to triple mutant combinations and did not appear during the analysis of double mutants (Tong *et al.*, 2004) (E. Kuzmin, C. Boone, unpublished data). Large-scale examination of triple genetic interactions will enable the identification of shared versus unique biological functions of gene duplicates, thus shedding light onto the mechanisms of functional divergence and the evolution of the yeast genome.

Given the size of the yeast genome (~6,000 genes), the number of possible triple mutant combinations (6,000 x 6,000 x 6,000 = 2.2 x 10^{11}) is extremely high and their systematic analysis using current genetic mapping technologies is nearly impossible. We must therefore develop novel experimental methods that would accelerate construction and phenotypic analysis of yeast mutants by at least an order of magnitude.

5.2.1.3 Expanding phenotypic assays

Investigating phenotypes other than fitness will greatly increase the sensitivity of genetic interaction analysis. A recent survey of spindle morphology phenotypes uncovered four times more genetic interactions than the equivalent fitness-based study (Vizeacoumar *et al.*, 2010), forecasting an explosion of data as soon as the costs and times of high-throughput microscopy become more affordable. Similar results were obtained from investigating genetic

interactions affecting yeast filamentous growth (Drees *et al.*, 2005), receptor endocytosis (Burston *et al.*, 2009) and protein folding pathways (Jonikas *et al.*, 2009). An in-depth analysis of subcellular phenotypes may also provide mechanistic explanations for fitness-based genetic interactions by linking double mutant growth defect to the malfunction of a specific molecular pathway.

5.2.2 Mapping complex phenotypes in yeast and in higher organisms

As a model system for genetic interactions, yeast plays an important role in answering the genotype-to-phenotype question, which attempts to associate quantitative traits, such as disease risk, to specific combinations of genomic variants carried by an individual. Recent advancements in genome sequence analysis and genome-wide association studies have recognized that interactions between genomic variants account for a substantial fraction of disease heritability that is not explained by the variants' additive contributions (Chapter 1.6). As a result, genetic interactions must be taken into account when estimating loci contribution to the heritability of disease.

The complexity of common phenotypes, however, likely extends well beyond single gene contributions and their pair-wise interactions, and is governed by an intricate network connecting multiple genomic loci. In *S. cerevisiae*, for example, most cases of conditional essentiality, where a gene is essential in one laboratory strain but completely dispensable in another, appear to be modulated by four or more modifiers (Dowell *et al.*, 2010). Similarly, yeast chemical resistance traits are regulated by as many as 40–50 loci (Ehrenreich *et al.*, 2010), with variation depending on the strain's genetic background and the trait under exam-

ination (Ehrenreich *et al.*, 2012). These findings indicate that even an organism as simple as yeast has an extremely intricate genotype-to-phenotype relationship and that systematic mapping of pairwise genetic interactions reveals just the tip of the iceberg of its complex genetic architecture. Approaching this problem in higher organisms, including humans, faces numerous additional challenges, including reduced sample sizes and heterogeneous genetic backgrounds. However, since the divergence between the two laboratory yeast strains is often comparable to the divergence of two human individuals (Wang *et al.*, 2008; Dowell *et al.*, 2010), yeast is an ideal model organism to pursue the investigation of complex phenotypes and will help uncovering the general rules connecting an individual's phenotype to its genotype.

5.2.3 Stochastic effects in genotype to phenotype mapping

While increasing evidence shows that genetic interactions contribute a substantial fraction of disease heritability (Zuk *et al.*, 2012), for most diseases the total extent of heritability is still undetermined, as we don't know what fraction of phenotype variance is controlled by genetic factors (e.g., genomic variants and their interactions) and what fraction is instead due to environmental differences or stochastic variability at the intracellular level. A recent analysis of medical histories of monozygotic twins suggests that the non-genetic components of disease risk may be more prominent than anticipated: for 13 of the 27 analyzed conditions, the majority of the affected individuals would result negative to a genetic test based on the phenotypes of their twins, i.e. individuals with an identical genotype (Roberts *et al.*, 2012). This suggests that the environmental and stochastic components of disease risk are as important as their genetic counterparts and must be investigated in a systematic manner.

Recent findings in *C. elegans* have started to unravel the potential mechanisms of non-genetic phenotypic variation (Burga *et al.*, 2011; Casanueva *et al.*, 2012). It was shown, for example, that genetic variation in one gene can be buffered by stochastic or environmentally induced changes in the expression level of a buffering gene, such as a duplicate (Burga *et al.*, 2011) or a molecular chaperone (Burga *et al.*, 2011; Casanueva *et al.*, 2012).

In yeast, large populations of genetically identical cells can be easily monitored using robotic manipulations and high-throughput imaging approaches. Phenotype penetrance, i.e. the proportion of genetically identical individuals expressing a given phenotype, can be scored using single-cell microscopy or, at a higher level, colony growth variation. For example, a set of deletion mutants can be examined for growth to determine the distribution of relative fitness values within an isogenic population. The width and shape of this distribution will be very informative for identifying mutations that predispose yeast to genotype-unrelated phenotypic variation and might be indicative of its underlying cause. For example, secondary site mutations are expected to arise at a low frequency in the population and will likely result in isolated peaks of unexpected fitness. In contrast, stochastic changes in gene expression levels are more likely to cause a multi-modal distribution of fitness and will involve genes whose dosage suppresses or exacerbates the phenotype of the original genetic background. The availability of yeast large-scale genetic interaction data, involving both loss-of-function and gain-of-function mutations (Costanzo *et al.*, 2010; Magtanong *et al.*, 2011; Sharifpoor *et al.*, 2012), will greatly facilitate the interpretation of this phenotypic variability and help elucidating the natural buffering mechanisms of eukaryotic cells.

5.2.4 The global functional map of *S. cerevisiae*

Despite the importance of exploring yeast robustness to perturbation using a variety of experimental approaches, the next big challenge in yeast functional genomics is integration of all the accumulated data into a single model of yeast biology. In the ten years that have elapsed since the completion of the yeast deletion collection (Giaever *et al.*, 2002), nearly all non-essential genes have been scrutinized in hundreds, if not thousands, of different ways (Botstein and Fink, 2011). However, despite the availability of this incredible wealth of data, we are still lacking a global unified picture of the yeast functional organization, which would serve as a repository for our collective knowledge and a model for more complex cellular systems. Numerous studies have firmly established that different types of genomic data provide complementary information for defining gene function (Beyer *et al.*, 2007), but the exact relationships between data types are still elusive to our understanding.

For example, the analysis of our global genetic interaction network showed that genetic interactions, while often connecting members of the same pathway or protein complex, can also span across very distant biological processes, describing the degree to which they share common functionality (Chapter 3.7). In contrast, protein-protein interactions define strictly local relationships and, for example, cannot link genes acting in different cellular compartments. The integration of physical and genetic interactions has been very successful in exploring the local neighborhoods of numerous genes (Tewari *et al.*, 2004; Kelley and Ideker, 2005; Ye *et al.*, 2005; Zhao *et al.*, 2005; Pan *et al.*, 2006). However, unification of the two data types on a global level still remains to be accomplished. Similarly, the integration of physical and genetic networks with gene expression or protein abundance profiles has the

potential to add a spatio-temporal dimension to the network, i.e. uncover network domains that are active in a particular growth condition or at a particular stage of the cell cycle. In *C. elegans*, this strategy helped elucidating the coordination of early embryogenesis processes and showed that static and dynamic networks are highly informative when combined (Gunsalus *et al.*, 2005). In yeast, the identification of subnetworks active under specific environmental conditions may help prioritizing experiments for constructing condition-specific genetic interaction networks, thus reducing the number of testable combinations (Chapter 5.2.1.1).

Importantly, the scientific community is making remarkable efforts to facilitate integration of genome-wide datasets. For example, it provides easy access to genomic data by adopting common data exchange formats, such as BioPAX (Demir *et al.*, 2010), and releasing open-source data visualization software, such as Cytoscape (Smoot *et al.*, 2011). Unprecedented progress in computer infrastructure development in the last decade has revolutionized genomic data storage, transfer and analysis, introducing cloud computing and parallel computer architectures (Schadt *et al.*, 2010). In addition, numerous computational algorithms have been developed to minimize the noise and maximize the sensitivity of genomic datasets, often through integration of heterogeneous data which compensates for their reciprocal false positive and false negative rates (Beyer *et al.*, 2007).

5.3 Yeast as a model for genome organization

Meiotic division is a key reproductive mechanism that promotes biological diversity by re-shuffling genetic information through homologous recombination (Chapter 4). While most

genetic variants assort at random relative to their maternal and paternal configurations, a smaller portion of the genome tends to remain unchanged across generations due to physical and genetic linkage among loci (Chapter 4). An important question that remains unanswered is what determines such an uneven distribution of recombination events. It is possible, for example, that within certain genomic regions recombination is harmful and thus has been reduced by natural selection. Alternatively, the positioning of recombination events could be independent from selective pressure and merely reflect the local properties of the DNA sequence or chromatin structure. A comparative examination of functional, structural and evolutionary features of genes located at hotspots and coldspots of recombination will likely provide an answer to this question and help elucidating the history of current genomes.

In theory, preserving certain combinations of gene variants could be evolutionary advantageous if, for example, they share a strong positive interaction for fitness (Peters and Lively, 2000). More generally, tight linkage between genes sharing a common biological function would prevent recombination from combining mutations arising independently in each gene, thus precluding potential deleterious effects. Supporting this hypothesis, yeast genes located closely on the same chromosome (but not necessarily adjacent to one another) tend to show similar expression profiles throughout the cell cycle, suggesting a common regulatory mechanism (Cohen *et al.*, 2000). Several other studies have reported similar non-random distributions of functionally related genes within yeast, worm and human genomes (Nei, 2003; Michalak, 2008). However, the degree to which functional clustering relates to frequency of meiotic recombination is still unclear.

A preliminary analysis of the genome-wide genetic linkage map for *S. cerevisiae*, presented in Chapter 4, indicates that linked gene pairs share co-annotation to the same Gene Ontology term as often as gene pairs segregating independently, thus suggesting that linkage does not correlate with co-function (A. Baryshnikova, unpublished data). However, an annotation-based approach in the analysis of co-segregating gene pairs bears several limitations that may hinder our ability to detect a strong relationship. First, an annotation bias against closely located gene pairs is not unexpected, given the experimental challenges associated with the investigation of their relationships (Chapter 4.4). Second, phenotypic analyses of closely located mutants might be affected by the so-called neighboring-gene effect, where the phenotype of a deletion mutant strain reflects the combined biological consequence of both inactivating the gene of interest and perturbing its neighbors (Baryshnikova and Andrews, 2012; Ben-Shitrit *et al*., 2012). As a result, the assessment of functional relationships between linked gene pairs must be carried out using annotation-independent and possibly deletion-independent approaches. For example, it must be determined how often linked gene pairs share direct protein-protein interactions or belong to the same protein complex, whether they are similarly expressed in response to the same chemical and environmental stimuli, and whether their protein abundances are comparable.

Selection for a reduced recombination frequency within certain genomic regions may also be unrelated from gene function. Homologous recombination is a potentially mutagenic mechanism and is normally suppressed whenever its consequences might be particularly deleterious (San Filippo *et al.*, 2008). For example, recombination rarely occurs within highly repetitive genomic regions, such as telomeres and the ribosomal DNA locus (Chapter 4.2), to

prevent substantial chromosomal rearrangements. Moreover, yeast essential genes tend to cluster away from double-strand breaks (DSB) which initiate meiotic recombination (Pal and Hurst, 2003), suggesting that DSB repair involving essential gene sequences may cause lethality. Pal and Hurst (2003) suggested that, if mutagenic risk is the cause of essential gene clustering, silent substitutions should be particularly rare in regions of low recombination, a hypothesis that can be tested by examining the sequences of *S. cerevisiae* close relatives. High quality genomes of at least three *Saccharomyces sensu stricto* yeasts (*S. bayanus*, *S. kudriavzevii* and *S. mikatae*) have been assembled recently (Scannell *et al.*, 2011), thus enabling the assessment of substitution rates in proximity of essential genes.

The preservation of linkage may also be unrelated from selection mechanisms and be simply the result of a random distribution of recombination hotspots, determined by local properties of DNA sequence or chromatin configuration. In this case, linked gene pairs will not display any functional relationship, but may share a number of additional features reflecting their common history. For example, linked gene pairs would be more likely to be conserved in other species, both in terms of their sequence and relative ordering. The evolution rates of linked genes would be expected to be comparable, as would be the frequencies of gain or loss of additional gene copies. Because of the availability of many sequenced genomes, we can now test these hypotheses and elucidate the contribution of meiotic recombination to the evolution of eukaryotic genomes.

5.4 Mechanisms of recombination inhibition within pericentric regions

Genetic linkage surrounding yeast centromeres is especially interesting because its extent appears to be proportional to chromosome length (Chapter 4.7). Centromeres play an essential role during meiotic division and are responsible for the correct segregation of homologous chromosomes (Brar and Amon, 2008). Centromeres associate with complex protein structures, which mediate the attachment of spindle microtubules and assist the timely separation of homologous chromosomes and sister chromatids (Chapter 4.2). It may not be surprising that larger chromosomes require larger protein structures for segregation; however, our knowledge of chromosome-size related meiotic mechanisms is quite limited and experimental validation is necessary to confirm this hypothesis.

One potential explanation for the chromosome-size dependent inhibition of recombination around the centromeres is the formation of chromosomal loops, extending towards the cell poles in response to the tension generated by the spindle microtubules (Chapter 4.7). A number of experimental tests can be performed to validate this hypothesis. First, a fluorescently marked histone H3 variant, as well as other centromere-specific proteins (Chen *et al.*, 2000; Goshima and Yanagida, 2000; Pearson *et al.*, 2001), can be used to assess whether the distance between homologous centromeres oscillates at the metaphase I-anaphase I transition, similarly to the oscillation of sister centromeres prior to mitotic anaphase (Yeh *et al.*, 2008). At meiosis I, sister kinetochores show monopolar attachment to the same spindle pole and are not under tension. Therefore, intracellular fluorescent signal should localize to one or two defined spots, depending on the relative distance between homologous centromeres. In addi-

tion, repeated arrays of specific protein binding sequences can be inserted into the yeast genome to examine oscillation amplitude at varying distances from the centromere. For example, introduction of the *E. coli* lac operator (lacO) enabled monitoring of lac repressor (lacR)-GFP localization to various chromosomal domains during mitotic division (Straight *et al.*, 1997; Goshima and Yanagida, 2000; He *et al.*, 2000; Tanaka *et al.*, 2000; Pearson *et al.*, 2001). Similarly, tetO and tetR-GFP constructs have been used to map sister kinetochore bi-orientation during mitosis (Tanaka *et al.*, 2002). The simultaneous use of two constructs with two different fluorescent markers, for example, by mating a strain carrying lacO/lacR-GFP construct and a strain carrying tetO/tetR-RFP construct, will allow to observe the dynamics of the two homologous centromeres and their flanking regions relative to one another. The insertion of these constructs into pericentric regions of small and large chromosomes will help determining whether chromosome length affects the degree to which centromeres extend towards the cell poles during metaphase I-anaphase I transition, and thus will provide a potential explanation for the chromosome size-dependent genetic linkage variation we observe in SGA data.

The relationship between chromosome size and pericentric linkage could also be the result of a centromere-specific distribution of recombination proteins and their regulators, which determine the number of crossover events occurring within a given genomic region. In yeast, about 150 DSBs are generated during each meiosis, but only ~20 are resolved as crossovers (Hyppa and Smith, 2010), suggesting that a regulatory mechanism directs most DSBs towards alternative repair pathways. One of the factors determining the fate of a DSB could be its physical location, intrinsically associated with local chromatin structure and

availability of recombination effectors. It is possible, for example, that recombination proteins bind DNA at a much lower frequency within pericentric regions than anywhere else in the genome and thus Spo11-mediated DSBs, which occur at a relatively normal rate up to 8–20 kb from the centromere (Gerton *et al.*, 2000; Borde *et al.*, 2004; Buhler *et al.*, 2007), cannot be resolved via a crossover mechanism. To investigate this hypothesis, we must compare our linkage map to the genome-wide distribution of recombination protein binding sites that can be systematically obtained using chromatin immunoprecipitation (ChIP) coupled with microarray hybridization or high-throughput sequencing (Park, 2009).

Among the first candidates for localization mapping is Dmc1, the meiosis-specific RecA homolog that binds ssDNA and initiates strand invasion (Krejci *et al.*, 2012). Dmc1 is an essential component of the inter-homolog crossover process and, when mutated, causes a dramatic reduction of recombination intermediates and spore inviability (Krejci *et al.*, 2012). The function of Dmc1 is assisted by several auxiliary proteins, such as the Mei5-Sae3 complex, required for Dmc1 recruitment and loading, and the Hop2-Mnd1 complex, which stabilizes Dmc1-filaments and promotes strand invasion and D-loop formation (Krejci *et al.*, 2012). Mapping the location of these and other meiosis-specific recombination effectors will be invaluable for understanding the non-random distribution of recombination events, particularly near centromeres.

Another important component of the meiotic recombination process is the synaptonemal complex (SC), a large protein structure that mediates inter-homolog pairing during meiotic prophase (Lynn *et al.*, 2007). Members of the SC include Hop1, Red1, Zip1, Zip2, Zip3 and Mek1 (Rockmill and Roeder, 1988; Hollingsworth *et al.*, 1990; Sym *et al.*, 1993;

Chua and Roeder, 1998; Agarwal and Roeder, 2000). Zip1, in particular, has been associated with regulating recombination around centromeres: in a *zip1*Δ mutant, pericentric repression of recombination is entirely abolished, while frequency of DSBs is unaltered (Sym and Roeder, 1994), thus suggesting that Zip1 normally channels pericentric DSBs into a different resolution pathway. The genome-wide localization of Zip1, together with other SC components Hop1 and Red1, has been recently made available via a large-scale ChIP-chip experiment (Panizza *et al.*, 2011), thus enabling the analysis of the dataset in relation to our genetic linkage data.

5.5 Conclusions

The awesome power of yeast genetics remains the leading force for understanding the inner workings of a simple eukaryotic cell and charting the general rules governing the life of other organisms. In this Thesis, I have shown that systematic genetic approaches, coupled with rigorous data processing, provide a formidable tool for exploring the functional organization of a living cell and generate information-rich datasets that will be revealing countless layers of novel biology for the next several years. Our results have set up the stage for numerous follow-up experiments, several of which are already under way and promise to achieve higher levels of speed, throughput and functional content than possible ever before. These data will provide us with a wider and deeper understanding of eukaryotic organisms, including humans.

6 Appendix

6.1 Theoretical evidence for adopting the multiplicative model of genetic interactions for fitness

It is not always clear which model of genetic interactions (additive or multiplicative) is the most appropriate for a particular phenotype. In case of fitness, there is a theoretical reason for adopting the multiplicative model: linkage disequilibrium will not arise in a population at equilibrium if genes are independent under the multiplicative model.

Linkage disequilibrium is defined as the non-random association of alleles at two or more loci in a population. Given a two-locus two-allele model, the population is in linkage equilibrium when the frequencies of the four genotypes (A_1B_1, A_1B_2, A_2B_1 and A_2B_2) equal to the products of the frequencies of the relative alleles.

$$
\begin{aligned}
f(A_1B_1) &= p_A p_B \\
f(A_1B_2) &= p_A(1-p_B) \\
f(A_2B_1) &= (1-p_A)p_B \\
f(A_2B_2) &= (1-p_A)(1-p_B)
\end{aligned}
$$

where p_A and p_B are the frequencies of the alleles A_1 and B_1, respectively.

Following Lewontin and Kojima (Lewontin and Kojima, 1960), deviations from equilibrium are usually measured by the coefficient of linkage disequilibrium D:

$$D = f(A_1B_1)f(A_2B_2) - f(A_1B_2)f(A_2B_1)$$

The frequencies of the four genotypes can then be expressed as a function of the frequencies of the individual alleles and D:

$$f(A_1B_1) = p_A p_B + D$$
$$f(A_1B_2) = p_A(1-p_B) - D$$
$$f(A_2B_1) = (1-p_A)p_B - D$$
$$f(A_2B_2) = (1-p_A)(1-p_B) + D$$

To provide theoretical evidence in favor of the multiplicative model of genetic interactions for fitness, we can calculate the linkage disequilibrium coefficient D' in a population after one round of selection and show that, if the population was originally at equilibrium (D = 0), it will stay so after selection ($D' = 0$) only if allele fitness defects combine multiplicatively.

After selection, the frequency of each genotype equals the product of its frequency before selection times its relative fitness, which equals its genotypic fitness divided by the mean population fitness, W. Assuming an additive model of genetic interactions, the fitness and the relative frequency of each genotype can be expressed as follows:

Genotype	Fitness (additive)	Genotype frequency before selection	Genotype frequency after selection
A_1B_1	1	$p_A p_B + D$	$(p_A p_B + D)\frac{1}{W}$
A_1B_2	$1-t$	$p_A(1-p_B) - D$	$\left[p_A(1-p_B) - D\right]\frac{1-t}{W}$
A_2B_1	$1-s$	$(1-p_A)p_B - D$	$\left[(1-p_A)p_B - D\right]\frac{1-s}{W}$
A_2B_2	$1-t-s$	$(1-p_A)(1-p_B) + D$	$\left[(1-p_A)(1-p_B) + D\right]\frac{1-t-s}{W}$

The coefficient of linkage disequilibrium D' can be computed as:

$$D' = \frac{1}{W}(p_A p_B + D)\frac{1-t-s}{W}\left[(1-p_A)(1-p_B)+D\right] - \frac{1-t}{W}\frac{1-s}{W}\left[p_A(1-p_B)-D\right]\left[p_B(1-p_A)-D\right]$$
$$D' = -\frac{1}{W^2}\left\{ts\left[p_A(1-p_B)-D\right]\left[p_B(1-p_A)-D\right]+(1-t-s)\left(D-2D^2-2p_A p_B D\right)\right\}$$

When the initial population is in linkage equilibrium ($D = 0$), D' simplifies to:

$$D' = -\frac{1}{W^2}p_A p_B(1-p_A)(1-p_B)ts$$

Under the assumption of a multiplicative model of genetic interactions, the fitness and the relative frequency of each genotype will follow the following rules:

Genotype	Fitness (multiplicative)	Genotype frequency before selection	Genotype frequency after selection
A_1B_1	1	$p_A p_B + D$	$(p_A p_B + D)\frac{1}{W}$
A_1B_2	$1-t$	$p_A(1-p_B)-D$	$\left[p_A(1-p_B)-D\right]\frac{1-t}{W}$
A_2B_1	$1-s$	$(1-p_A)p_B-D$	$\left[(1-p_A)p_B-D\right]\frac{1-s}{W}$
A_2B_2	$(1-t)(1-s)$	$(1-p_A)(1-p_B)+D$	$\left[(1-p_A)(1-p_B)+D\right]\frac{(1-t)(1-s)}{W}$

The coefficient of linkage disequilibrium D' can be computed as:

$$D' = \frac{1}{W}(p_A p_B + D)\frac{(1-t)(1-s)}{W}\left[(1-p_A)(1-p_B)+D\right] - \frac{1-t}{W}\frac{1-s}{W}\left[p_A(1-p_B)-D\right]\left[p_B(1-p_A)-D\right]$$
$$D' = \frac{(1-t)(1-s)}{W^2}D$$

Therefore, if the population is initially in linkage equilibrium ($D = 0$), it will remain so ($D' = 0$) despite selection, under the assumption of the multiplicative (but not additive) model of genetic interactions.

Bibliography

Agarwal, S., and G.S. Roeder. 2000. Zip3 provides a link between recombination enzymes and synaptonemal complex proteins. *Cell*. 102:245-255.

Aguilar, P.S., F. Frohlich, M. Rehman, M. Shales, I. Ulitsky, A. Olivera-Couto, H. Braberg, R. Shamir, P. Walter, M. Mann, C.S. Ejsing, N.J. Krogan, and T.C. Walther. 2010. A plasma-membrane E-MAP reveals links of the eisosome with sphingolipid metabolism and endosomal trafficking. *Nat Struct Mol Biol*. 17:901-908.

Alberti, S., A.D. Gitler, and S. Lindquist. 2007. A suite of Gateway cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae. *Yeast*. 24:913-919.

Albertson, D.G., and J.N. Thomson. 1982. The kinetochores of Caenorhabditis elegans. *Chromosoma*. 86:409-428.

Avery, L., and S. Wasserman. 1992. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet*. 8:312-316.

Bandyopadhyay, S., R. Kelley, N.J. Krogan, and T. Ideker. 2008. Functional maps of protein complexes from quantitative genetic interaction data. *PLoS Comput Biol*. 4:e1000065.

Bandyopadhyay, S., M. Mehta, D. Kuo, M.K. Sung, R. Chuang, E.J. Jaehnig, B. Bodenmiller, K. Licon, W. Copeland, M. Shales, D. Fiedler, J. Dutkowski, A. Guenole, H. van Attikum, K.M. Shokat, R.D. Kolodner, W.K. Huh, R. Aebersold, M.C. Keogh, N.J. Krogan, and T. Ideker. 2010. Rewiring of genetic networks in response to DNA damage. *Science*. 330:1385-1389.

Barwell, K.J., J.H. Boysen, W. Xu, and A.P. Mitchell. 2005. Relationship of DFG16 to the Rim101p pH response pathway in Saccharomyces cerevisiae and Candida albicans. *Eukaryot Cell*. 4:890-899.

Baryshnikova, A., and B. Andrews. 2012. Neighboring-gene effect: a genetic uncertainty principle. *Nat Methods*. 9:341, 343.

Baryshnikova, A., M. Costanzo, S. Dixon, F.J. Vizeacoumar, C.L. Myers, B. Andrews, and C. Boone. 2010a. Synthetic genetic array (SGA) analysis in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Methods Enzymol*. 470:145-179.

Baryshnikova, A., M. Costanzo, Y. Kim, H. Ding, J. Koh, K. Toufighi, J.Y. Youn, J. Ou, B.J. San Luis, S. Bandyopadhyay, M. Hibbs, D. Hess, A.C. Gingras, G.D. Bader, O.G. Troyanskaya, G.W. Brown, B. Andrews, C. Boone, and C.L. Myers. 2010b. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Methods*. 7:1017-1024.

Bateson, W. 1909. Mendel's Principles of Heredity. Cambridge University Press, London, England.

Baudat, F., and A. Nicolas. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc Natl Acad Sci U S A*. 94:5213-5218.

Bellaoui, M., M. Chang, J. Ou, H. Xu, C. Boone, and G.W. Brown. 2003. Elg1 forms an alternative RFC complex important for DNA replication and genome integrity. *Embo J*. 22:4304-4313.

Bellay, J., S. Han, M. Michaut, T. Kim, M. Costanzo, B.J. Andrews, C. Boone, G.D. Bader, C.L. Myers, and P.M. Kim. 2011. Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol*. 12:R14.

Ben-Aroya, S., C. Coombes, T. Kwok, K.A. O'Donnell, J.D. Boeke, and P. Hieter. 2008. Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. *Mol Cell*. 30:248-258.

Ben-Aroya, S., X. Pan, J.D. Boeke, and P. Hieter. 2010. Making temperature-sensitive mutants. *Methods Enzymol*. 470:181-204.

Ben-Shitrit, T., N. Yosef, K. Shemesh, R. Sharan, E. Ruppin, and M. Kupiec. 2012. Systematic identification of gene annotation errors in the widely used yeast mutation collections. *Nat Methods*.

Bender, A., and J.R. Pringle. 1991. Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in Saccharomyces cerevisiae. *Mol Cell Biol*. 11:1295-1305.

Bennett, M.D., J.B. Smith, J. Ward, and G. Jenkins. 1981. The relationship between nuclear DNA content and centromere volume in higher plants. *J Cell Sci*. 47:91-115.

Beyer, A., S. Bandyopadhyay, and T. Ideker. 2007. Integrating physical and genetic maps: from genomes to interaction networks. *Nat Rev Genet*. 8:699-710.

Blitzblau, H.G., G.W. Bell, J. Rodriguez, S.P. Bell, and A. Hochwagen. 2007. Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. *Curr Biol*. 17:2003-2012.

Borde, V., W. Lin, E. Novikov, J.H. Petrini, M. Lichten, and A. Nicolas. 2004. Association of Mre11p with double-strand break sites during yeast meiosis. *Mol Cell*. 13:389-401.

Botstein, D., and G.R. Fink. 2011. Yeast: an experimental organism for 21st Century biology. *Genetics*. 189:695-704.

Boyle, E.I., S. Weng, J. Gollub, H. Jin, D. Botstein, J.M. Cherry, and G. Sherlock. 2004. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*. 20:3710-3715.

Brar, G.A., and A. Amon. 2008. Emerging roles for centromeres in meiosis I chromosome segregation. *Nat Rev Genet*. 9:899-910.

Brauer, M.J., C. Huttenhower, E.M. Airoldi, R. Rosenstein, J.C. Matese, D. Gresham, V.M. Boer, O.G. Troyanskaya, and D. Botstein. 2008. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Molecular biology of the cell*. 19:352-367.

Breslow, D.K., D.M. Cameron, S.R. Collins, M. Schuldiner, J. Stewart-Ornstein, H.W. Newman, S. Braun, H.D. Madhani, N.J. Krogan, and J.S. Weissman. 2008. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods*. 5:711-718.

Bryant, H.E., N. Schultz, H.D. Thomas, K.M. Parker, D. Flower, E. Lopez, S. Kyle, M. Meuth, N.J. Curtin, and T. Helleday. 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*. 434:913-917.

Buhler, C., V. Borde, and M. Lichten. 2007. Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in Saccharomyces cerevisiae. *PLoS Biol*. 5:e324.

Buonomo, S.B., R.K. Clyne, J. Fuchs, J. Loidl, F. Uhlmann, and K. Nasmyth. 2000. Disjunction of homologous chromosomes in meiosis I depends on proteolytic cleavage of the meiotic cohesin Rec8 by separin. *Cell*. 103:387-398.

Burga, A., M.O. Casanueva, and B. Lehner. 2011. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature*. 480:250-253.

Burston, H.E., L. Maldonado-Baez, M. Davey, B. Montpetit, C. Schluter, B. Wendland, and E. Conibear. 2009. Regulators of yeast endocytosis identified by systematic quantitative analysis. *J Cell Biol*. 185:1097-1110.

Butland, G., M. Babu, J.J. Diaz-Mejia, F. Bohdana, S. Phanse, B. Gold, W. Yang, J. Li, A.G. Gagarinova, O. Pogoutse, H. Mori, B.L. Wanner, H. Lo, J. Wasniewski, C. Christopolous, M. Ali, P. Venn, A. Safavi-Naini, N. Sourour, S. Caron, J.Y. Choi, L. Laigle, A. Nazarians-Armavil, A. Deshpande, S. Joe, K.A. Datsenko, N. Yamamoto, B.J. Andrews, C. Boone, H. Ding, B. Sheikh, G. Moreno-Hagelseib, J.F. Greenblatt, and A. Emili. 2008. eSGA: E. coli synthetic genetic array analysis. *Nat Methods*. 5:789-795.

Byrne, A.B., M.T. Weirauch, V. Wong, M. Koeva, S.J. Dixon, J.M. Stuart, and P.J. Roy. 2007. A global analysis of genetic interactions in Caenorhabditis elegans. *J Biol*. 6:8.

Byrne, K.P., and K.H. Wolfe. 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res*. 15:1456-1461.

Casanueva, M.O., A. Burga, and B. Lehner. 2012. Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. *Science*. 335:82-85.

Chen, E.J., and C.A. Kaiser. 2003. LST8 negatively regulates amino acid biosynthesis as a component of the TOR pathway. *J Cell Biol*. 161:333-347.

Chen, S.Y., T. Tsubouchi, B. Rockmill, J.S. Sandler, D.R. Richards, G. Vader, A. Hochwagen, G.S. Roeder, and J.C. Fung. 2008. Global analysis of the meiotic crossover landscape. *Dev Cell*. 15:401-415.

Chen, Y., R.E. Baker, K.C. Keith, K. Harris, S. Stoler, and M. Fitzgerald-Hayes. 2000. The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. *Mol Cell Biol*. 20:7037-7048.

Cherry, J.M., C. Ball, S. Weng, G. Juvik, R. Schmidt, C. Adler, B. Dunn, S. Dwight, L. Riles, R.K. Mortimer, and D. Botstein. 1997. Genetic and physical maps of Saccharomyces cerevisiae. *Nature*. 387:67-73.

Cherry, J.M., E.L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E.T. Chan, K.R. Christie, M.C. Costanzo, S.S. Dwight, S.R. Engel, D.G. Fisk, J.E. Hirschman, B.C. Hitz, K. Karra, C.J. Krieger, S.R. Miyasato, R.S. Nash, J. Park, M.S. Skrzypek, M. Simison, S. Weng, and E.D. Wong. 2012. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res*. 40:D700-705.

Cheung, H.W., G.S. Cowley, B.A. Weir, J.S. Boehm, S. Rusin, J.A. Scott, A. East, L.D. Ali, P.H. Lizotte, T.C. Wong, G. Jiang, J. Hsiao, C.H. Mermel, G. Getz, J. Barretina, S. Gopal, P. Tamayo, J. Gould, A. Tsherniak, N. Stransky, B. Luo, Y. Ren, R. Drapkin, S.N. Bhatia, J.P. Mesirov, L.A. Garraway, M. Meyerson, E.S. Lander, D.E. Root, and W.C. Hahn. 2011. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc Natl Acad Sci U S A*. 108:12372-12377.

Chua, G., Q.D. Morris, R. Sopko, M.D. Robinson, O. Ryan, E.T. Chan, B.J. Frey, B.J. Andrews, C. Boone, and T.R. Hughes. 2006. Identifying transcription factor functions and targets by phenotypic activation. *Proc Natl Acad Sci U S A*. 103:12045-12050.

Chua, P.R., and G.S. Roeder. 1998. Zip2, a meiosis-specific protein required for the initiation of chromosome synapsis. *Cell*. 93:349-359.

Cleveland, W.S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*. 74.

Cohen, B.A., R.D. Mitra, J.D. Hughes, and G.M. Church. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet*. 26:183-186.

Cokol, M., H.N. Chua, M. Tasan, B. Mutlu, Z.B. Weinstein, Y. Suzuki, M.E. Nergiz, M. Costanzo, A. Baryshnikova, G. Giaever, C. Nislow, C.L. Myers, B.J. Andrews, C. Boone, and F.P. Roth. 2011. Systematic exploration of synergistic drug pairs. *Mol Syst Biol*. 7:544.

Collins, S.R., K.M. Miller, N.L. Maas, A. Roguev, J. Fillingham, C.S. Chu, M. Schuldiner, M. Gebbia, J. Recht, M. Shales, H. Ding, H. Xu, J. Han, K. Ingvarsdottir, B. Cheng, B. Andrews, C. Boone, S.L. Berger, P. Hieter, Z. Zhang, G.W. Brown, C.J. Ingles, A. Emili, C.D. Allis, D.P. Toczyski, J.S. Weissman, J.F. Greenblatt, and N.J. Krogan. 2007. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature*. 446:806-810.

Collins, S.R., A. Roguev, and N.J. Krogan. 2010. Quantitative genetic interaction mapping using the E-MAP approach. *Methods Enzymol*. 470:205-231.

Collins, S.R., M. Schuldiner, N.J. Krogan, and J.S. Weissman. 2006. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol*. 7:R63.

Costanzo, M., A. Baryshnikova, J. Bellay, Y. Kim, E.D. Spear, C.S. Sevier, H. Ding, J.L. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R.P. St Onge, B. VanderSluis, T. Makhnevych, F.J. Vizeacoumar, S. Alizadeh, S. Bahr, R.L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z.Y. Lin, W. Liang, M. Marback, J. Paw, B.J. San Luis, E. Shuteriqi, A.H. Tong, N. van Dyk, I.M. Wallace, J.A. Whitney, M.T. Weirauch, G. Zhong, H. Zhu, W.A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pal, F.P. Roth, G. Giaever, C. Nislow, O.G. Troyanskaya, H. Bussey, G.D. Bader, A.C. Gingras, Q.D. Morris, P.M. Kim, C.A. Kaiser, C.L. Myers, B.J. Andrews, and C. Boone. 2010. The genetic landscape of a cell. *Science*. 327:425-431.

Costanzo, M., A. Baryshnikova, C.L. Myers, B. Andrews, and C. Boone. 2011. Charting the genetic interaction map of a cell. *Curr Opin Biotechnol*. 22:66-74.

Costanzo, M., A. Baryshnikova, B. Vandersluis, B. Andrews, C.L. Myers, and C. Boone. 2012. Genetic networks. *In* Systems Biology, a Handbook. J. Dekker, M. Vidal, and A.J. Walhout, editors, in press.

Costanzo, M., and C. Boone. 2009. SGAM: an array-based approach for high-resolution genetic mapping in Saccharomyces cerevisiae. *Methods Mol Biol*. 548:37-53.

Costanzo, M., J.L. Nishikawa, X. Tang, J.S. Millman, O. Schub, K. Breitkreuz, D. Dewar, I. Rupes, B. Andrews, and M. Tyers. 2004. CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. *Cell*. 117:899-913.

Daly, M.J., J.D. Rioux, S.F. Schaffner, T.J. Hudson, and E.S. Lander. 2001. High-resolution haplotype structure in the human genome. *Nat Genet*. 29:229-232.

Davierwala, A.P., J. Haynes, Z. Li, R.L. Brost, M.D. Robinson, L. Yu, S. Mnaimneh, H. Ding, H. Zhu, Y. Chen, X. Cheng, G.W. Brown, C. Boone, B.J. Andrews, and T.R. Hughes. 2005. The synthetic genetic interaction spectrum of essential genes. *Nat Genet*. 37:1147-1152.

de Visser, J.A., J. Hermisson, G.P. Wagner, L. Ancel Meyers, H. Bagheri-Chaichian, J.L. Blanchard, L. Chao, J.M. Cheverud, S.F. Elena, W. Fontana, G. Gibson, T.F. Hansen, D. Krakauer, R.C. Lewontin, C. Ofria, S.H. Rice, G. von Dassow, A. Wagner, and M.C. Whitlock. 2003. Perspective: Evolution and detection of genetic robustness. *Evolution*. 57:1959-1972.

Decourty, L., C. Saveanu, K. Zemam, F. Hantraye, E. Frachon, J.C. Rousselle, M. Fromont-Racine, and A. Jacquier. 2008. Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles. *Proc Natl Acad Sci U S A*. 105:5821-5826.

Demir, E., M.P. Cary, S. Paley, K. Fukuda, C. Lemer, I. Vastrik, G. Wu, P. D'Eustachio, C. Schaefer, J. Luciano, F. Schacherer, I. Martinez-Flores, Z. Hu, V. Jimenez-Jacinto, G. Joshi-Tope, K. Kandasamy, A.C. Lopez-Fuentes, H. Mi, E. Pichler, I. Rodchenkov, A. Splendiani, S. Tkachev, J. Zucker, G. Gopinath, H. Rajasimha, R. Ramakrishnan, I. Shah, M. Syed, N. Anwar, O. Babur, M. Blinov, E. Brauner, D. Corwin, S. Donaldson, F. Gibbons, R. Goldberg, P. Hornbeck, A. Luna, P. Murray-Rust, E. Neumann, O. Reubenacker, M. Samwald, M. van Iersel, S. Wimalaratne, K. Allen, B. Braun, M. Whirl-Carrillo, K.H. Cheung, K. Dahlquist, A. Finney, M. Gillespie, E. Glass, L. Gong, R. Haw, M. Honig, O. Hubaut, D. Kane, S. Krupa, M. Kutmon, J. Leonard, D. Marks, D. Merberg, V. Petri, A. Pico, D. Ravenscroft, L. Ren, N. Shah, M. Sunshine, R. Tang, R. Whaley, S. Letovksy, K.H. Buetow, A. Rzhetsky, V. Schachter, B.S. Sobral, U. Dogrusoz, S. McWeeney, M. Aladjem, E. Birney, J. Collado-Vides, S. Goto, M. Hucka, N. Le Novere, N. Maltsev, A. Pandey, P. Thomas, E. Wingender, P.D. Karp, C. Sander, and G.D. Bader. 2010. The BioPAX community standard for pathway data sharing. *Nat Biotechnol*. 28:935-942.

Deutschbauer, A.M., D.F. Jaramillo, M. Proctor, J. Kumm, M.E. Hillenmeyer, R.W. Davis, C. Nislow, and G. Giaever. 2005. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics*. 169:1915-1925.

Dixon, S.J., M. Costanzo, A. Baryshnikova, B. Andrews, and C. Boone. 2009. Systematic mapping of genetic interaction networks. *Annu Rev Genet*. 43:601-625.

Dixon, S.J., Y. Fedyshyn, J.L. Koh, T.S. Prasad, C. Chahwan, G. Chua, K. Toufighi, A. Baryshnikova, J. Hayles, K.L. Hoe, D.U. Kim, H.O. Park, C.L. Myers, A. Pandey, D. Durocher, B.J. Andrews, and C. Boone. 2008. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc Natl Acad Sci U S A*. 105:16653-16658.

Dobzhansky, T. 1946. Genetics of Natural Populations. Xiii. Recombination and Variability in Populations of Drosophila Pseudoobscura. *Genetics*. 31:269-290.

Dowell, R.D., O. Ryan, A. Jansen, D. Cheung, S. Agarwala, T. Danford, D.A. Bernstein, P.A. Rolfe, L.E. Heisler, B. Chin, C. Nislow, G. Giaever, P.C. Phillips, G.R. Fink, D.K. Gifford, and C. Boone. 2010. Genotype to phenotype: a complex problem. *Science*. 328:469.

Drees, B.L., V. Thorsson, G.W. Carter, A.W. Rives, M.Z. Raymond, I. Avila-Campillo, P. Shannon, and T. Galitski. 2005. Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol*. 6:R38.

Dunham, M.J., H. Badrane, T. Ferea, J. Adams, P.O. Brown, F. Rosenzweig, and D. Botstein. 2002. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 99:16144-16149.

Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 5:113.

Ehrenreich, I.M., J. Bloom, N. Torabi, X. Wang, Y. Jia, and L. Kruglyak. 2012. Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet*. 8:e1002570.

Ehrenreich, I.M., N. Torabi, Y. Jia, J. Kent, S. Martis, J.A. Shapiro, D. Gresham, A.A. Caudy, and L. Kruglyak. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature*. 464:1039-1042.

Elias, J.E., and S.P. Gygi. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*. 4:207-214.

Esberg, A., B. Huang, M.J. Johansson, and A.S. Bystrom. 2006. Elevated levels of two tRNA species bypass the requirement for elongator complex in transcription and exocytosis. *Mol Cell*. 24:139-148.

Farmer, H., N. McCabe, C.J. Lord, A.N. Tutt, D.A. Johnson, T.B. Richardson, M. Santarosa, K.J. Dillon, I. Hickson, C. Knights, N.M. Martin, S.P. Jackson, G.C. Smith, and A. Ashworth. 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*. 434:917-921.

Fiedler, D., H. Braberg, M. Mehta, G. Chechik, G. Cagney, P. Mukherjee, A.C. Silva, M. Shales, S.R. Collins, S. van Wageningen, P. Kemmeren, F.C. Holstege, J.S. Weissman, M.C. Keogh, D. Koller, K.M. Shokat, and N.J. Krogan. 2009. Functional organization of the S. cerevisiae phosphorylation network. *Cell*. 136:952-963.

Fillingham, J., P. Kainth, J.P. Lambert, H. van Bakel, K. Tsui, L. Pena-Castillo, C. Nislow, D. Figeys, T.R. Hughes, J. Greenblatt, and B.J. Andrews. 2009. Two-color cell array screen reveals interdependent roles for histone chaperones and a chromatin boundary regulator in histone gene repression. *Mol Cell*. 35:340-351.

Finger, F.P., and P. Novick. 2000. Synthetic interactions of the post-Golgi sec mutations of Saccharomyces cerevisiae. *Genetics*. 156:943-951.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Proceedings of the Royal Society of Edinburgh*. 52:399.

Fitzgerald-Hayes, M., J.M. Buhler, T.G. Cooper, and J. Carbon. 1982a. Isolation and subcloning analysis of functional centromere DNA (CEN11) from Saccharomyces cerevisiae chromosome XI. *Mol Cell Biol*. 2:82-87.

Fitzgerald-Hayes, M., L. Clarke, and J. Carbon. 1982b. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell*. 29:235-244.

Fong, P.C., D.S. Boss, T.A. Yap, A. Tutt, P. Wu, M. Mergui-Roelvink, P. Mortimer, H. Swaisland, A. Lau, M.J. O'Connor, A. Ashworth, J. Carmichael, S.B. Kaye, J.H. Schellens, and J.S. de Bono. 2009. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med*. 361:123-134.

Fraser, H.B., D.P. Wall, and A.E. Hirsh. 2003. A simple dependence between protein evolution rate and the number of protein-protein interactions. *BMC Evol Biol*. 3:11.

Frazer, K.A., S.S. Murray, N.J. Schork, and E.J. Topol. 2009. Human genetic variation and its contribution to complex traits. *Nat Rev Genet*. 10:241-251.

Friesen, H., C. Humphries, Y. Ho, O. Schub, K. Colwill, and B. Andrews. 2006. Characterization of the yeast amphiphysins Rvs161p and Rvs167p reveals roles for the Rvs heterodimer in vivo. *Molecular biology of the cell*. 17:1306-1321.

Gasch, A.P., P.T. Spellman, C.M. Kao, O. Carmel-Harel, M.B. Eisen, G. Storz, D. Botstein, and P.O. Brown. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell*. 11:4241-4257.

Gavin, A.C., P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L.J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M.A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J.M. Rick,

B. Kuster, P. Bork, R.B. Russell, and G. Superti-Furga. 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature*. 440:631-636.

Gerton, J.L., J. DeRisi, R. Shroff, M. Lichten, P.O. Brown, and T.D. Petes. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 97:11383-11390.

Giaever, G., A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A.P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.D. Entian, P. Flaherty, F. Foury, D.J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J.H. Hegemann, S. Hempel, Z. Herman, D.F. Jaramillo, D.E. Kelly, S.L. Kelly, P. Kotter, D. LaBonte, D.C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S.L. Ooi, J.L. Revuelta, C.J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D.D. Shoemaker, S. Sookhai-Mahadeo, R.K. Storms, J.N. Strathern, G. Valle, M. Voet, G. Volckaert, C.Y. Wang, T.R. Ward, J. Wilhelmy, E.A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J.D. Boeke, M. Snyder, P. Philippsen, R.W. Davis, and M. Johnston. 2002. Functional profiling of the Saccharomyces cerevisiae genome. *Nature*. 418:387-391.

Giaever, G., D.D. Shoemaker, T.W. Jones, H. Liang, E.A. Winzeler, A. Astromoff, and R.W. Davis. 1999. Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat Genet*. 21:278-283.

Gibson, G. 2011. Rare and common variants: twenty arguments. *Nat Rev Genet*. 13:135-145.

Gingras, A.C., M. Caballero, M. Zarske, A. Sanchez, T.R. Hazbun, S. Fields, N. Sonenberg, E. Hafen, B. Raught, and R. Aebersold. 2005. A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol Cell Proteomics*. 4:1725-1740.

Glynn, E.F., P.C. Megee, H.G. Yu, C. Mistrot, E. Unal, D.E. Koshland, J.L. DeRisi, and J.L. Gerton. 2004. Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. *PLoS Biol*. 2:E259.

Goshima, G., and M. Yanagida. 2000. Establishing biorientation occurs with precocious separation of the sister kinetochores, but not the arms, in the early spindle of budding yeast. *Cell*. 100:619-633.

Goudreault, M., L.M. D'Ambrosio, M.J. Kean, M.J. Mullin, B.G. Larsen, A. Sanchez, S. Chaudhry, G.I. Chen, F. Sicheri, A.I. Nesvizhskii, R. Aebersold, B. Raught, and A.C. Gingras. 2009. A PP2A phosphatase high density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein. *Mol Cell Proteomics*. 8:157-171.

Griffiths, A.J.F., J.H. Miller, D.T. Suzuki, R.C. Lewontin, and W.M. Gelbart. 2000. Accurate calculation of large map distances. *In* An Introduction to Genetic Analysis, 7th edition. W. H. Freeman, New York.

Gsponer, J., M.E. Futschik, S.A. Teichmann, and M.M. Babu. 2008. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science*. 322:1365-1368.

Gu, Z., A. Cavalcanti, F.C. Chen, P. Bouman, and W.H. Li. 2002. Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. *Mol Biol Evol*. 19:256-262.

Guarente, L. 1993. Synthetic enhancement in gene interaction: a genetic tool come of age. *Trends Genet*. 9:362-366.

Gunsalus, K.C., H. Ge, A.J. Schetter, D.S. Goldberg, J.D. Han, T. Hao, G.F. Berriz, N. Bertin, J. Huang, L.S. Chuang, N. Li, R. Mani, A.A. Hyman, B. Sonnichsen, C.J. Echeverri, F.P. Roth, M. Vidal, and F. Piano. 2005. Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. *Nature*. 436:861-865.

Gurley, K.E., and C.J. Kemp. 2001. Synthetic lethality between mutation in Atm and DNA-PK(cs) during murine embryogenesis. *Curr Biol*. 11:191-194.

Haldane, J.B.S. 1919. The combination of linkage values, and the calculation of distances between the loci of linked factors. *Journal of genetics*. 8:299-209.

Hartman, J.L.t., B. Garvik, and L. Hartwell. 2001. Principles for the buffering of genetic variation. *Science*. 291:1001-1004.

Hartman, J.L.t., and N.P. Tippery. 2004. Systematic quantification of gene interactions by phenotypic array analysis. *Genome Biol*. 5:R49.

Hartman, P.E., and J.R. Roth. 1973. Mechanisms of suppression. *Adv Genet*. 17:1-105.

Hartwell, L. 2004. Genetics. Robust interactions. *Science*. 303:774-775.

Hartwell, L.H., P. Szankasi, C.J. Roberts, A.W. Murray, and S.H. Friend. 1997. Integrating genetic approaches into the discovery of anticancer drugs. *Science*. 278:1064-1068.

Hassold, T., M. Abruzzo, K. Adkins, D. Griffin, M. Merrill, E. Millie, D. Saker, J. Shen, and M. Zaragoza. 1996. Human aneuploidy: incidence, origin, and etiology. *Environ Mol Mutagen*. 28:167-175.

Hayashi, M., T. Fukuzawa, H. Sorimachi, and T. Maeda. 2005. Constitutive activation of the pH-responsive Rim101 pathway in yeast mutants defective in late steps of the MVB/ESCRT pathway. *Mol Cell Biol*. 25:9478-9490.

He, X., S. Asthana, and P.K. Sorger. 2000. Transient sister chromatid separation and elastic deformation of chromosomes during mitosis in budding yeast. *Cell*. 101:763-775.

Hillenmeyer, M.E., E. Fung, J. Wildenhain, S.E. Pierce, S. Hoon, W. Lee, M. Proctor, R.P. St Onge, M. Tyers, D. Koller, R.B. Altman, R.W. Davis, C. Nislow, and G. Giaever. 2008. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science*. 320:362-365.

Hinnebusch, A.G., and G.R. Fink. 1983. Positive regulation in the general amino acid control of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 80:5374-5378.

Ho, C.H., J. Piotrowski, S.J. Dixon, A. Baryshnikova, M. Costanzo, and C. Boone. 2011. Combining functional genomics and chemical biology to identify targets of bioactive compounds. *Curr Opin Chem Biol*. 15:66-78.

Ho, Y., A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys, and M. Tyers. 2002. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*. 415:180-183.

Hollingsworth, N.M., L. Goetsch, and B. Byers. 1990. The HOP1 gene encodes a meiosis-specific component of yeast chromosomes. *Cell*. 61:73-84.

Holstege, F.C., E.G. Jennings, J.J. Wyrick, T.I. Lee, C.J. Hengartner, M.R. Green, T.R. Golub, E.S. Lander, and R.A. Young. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*. 95:717-728.

Hoppins, S., S.R. Collins, A. Cassidy-Stone, E. Hummel, R.M. Devay, L.L. Lackner, B. Westermann, M. Schuldiner, J.S. Weissman, and J. Nunnari. 2011. A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *J Cell Biol*.

Horn, T., T. Sandmann, B. Fischer, E. Axelsson, W. Huber, and M. Boutros. 2011. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat Methods*. 8:341-346.

Hu, Y., A. Rolfs, B. Bhullar, T.V. Murthy, C. Zhu, M.F. Berger, A.A. Camargo, F. Kelley, S. McCarron, D. Jepson, A. Richardson, J. Raphael, D. Moreira, E. Taycher, D. Zuo, S. Mohr, M.F. Kane, J. Williamson, A. Simpson, M.L. Bulyk, E. Harlow, G. Marsischky, R.D. Kolodner, and J. LaBaer. 2007. Approaching a complete repository

of sequence-verified protein-encoding clones for Saccharomyces cerevisiae. *Genome Res*. 17:536-543.

Huber, A., B. Bodenmiller, A. Uotila, M. Stahl, S. Wanka, B. Gerrits, R. Aebersold, and R. Loewith. 2009. Characterization of the rapamycin-sensitive phosphoproteome reveals that Sch9 is a central coordinator of protein synthesis. *Genes Dev*. 23:1929-1943.

Hughes, T.R., M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey, H. Dai, Y.D. He, M.J. Kidd, A.M. King, M.R. Meyer, D. Slade, P.Y. Lum, S.B. Stepaniants, D.D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, and S.H. Friend. 2000. Functional discovery via a compendium of expression profiles. *Cell*. 102:109-126.

Huh, W.K., J.V. Falvo, L.C. Gerke, A.S. Carroll, R.W. Howson, J.S. Weissman, and E.K. O'Shea. 2003. Global analysis of protein localization in budding yeast. *Nature*. 425:686-691.

Hyppa, R.W., and G.R. Smith. 2010. Crossover invariance determined by partner choice for meiotic DNA break repair. *Cell*. 142:243-255.

Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A*. 98:4569-4574.

Jasnos, L., and R. Korona. 2007. Epistatic buffering of fitness loss in yeast double deletion strains. *Nat Genet*. 39:550-554.

Jenkins, G., and M.D. Bennett. 1981. The intranuclear relationship between centromere volume and chromosome size in Festuca scariosa X drymeja. *J Cell Sci*. 47:117-125.

Jin, R., C.J. Dobry, P.J. McCown, and A. Kumar. 2008. Large-scale analysis of yeast filamentous growth by systematic gene disruption and overexpression. *Mol Biol Cell*. 19:284-296.

Johansen, L.D., T. Naumanen, A. Knudsen, N. Westerlund, I. Gromova, M. Junttila, C. Nielsen, T. Bottzauw, A. Tolkovsky, J. Westermarck, E.T. Coffey, M. Jaattela, and T. Kallunki. 2008. IKAP localizes to membrane ruffles with filamin A and regulates actin cytoskeleton organization and cell migration. *J Cell Sci*. 121:854-864.

Jonikas, M.C., S.R. Collins, V. Denic, E. Oh, E.M. Quan, V. Schmid, J. Weibezahn, B. Schwappach, P. Walter, J.S. Weissman, and M. Schuldiner. 2009. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science*. 323:1693-1697.

Jorgensen, P., B. Nelson, M.D. Robinson, Y. Chen, B. Andrews, M. Tyers, and C. Boone. 2002. High-resolution genetic mapping with ordered arrays of Saccharomyces cerevisiae deletion mutants. *Genetics*. 162:1091-1099.

Kaliraman, V., J.R. Mullen, W.M. Fricke, S.A. Bastin-Shanower, and S.J. Brill. 2001. Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. *Genes Dev*. 15:2730-2740.

Kapitzky, L., P. Beltrao, T.J. Berens, N. Gassner, C. Zhou, A. Wuster, J. Wu, M.M. Babu, S.J. Elledge, D. Toczyski, R.S. Lokey, and N.J. Krogan. 2010. Cross-species chemogenomic profiling reveals evolutionarily conserved drug mode of action. *Mol Syst Biol*. 6:451.

Keeney, S., C.N. Giroux, and N. Kleckner. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*. 88:375-384.

Keeney, S., and M.J. Neale. 2006. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem Soc Trans*. 34:523-525.

Kelley, R., and T. Ideker. 2005. Systematic interpretation of genetic interactions using protein networks. *Nat Biotechnol*. 23:561-566.

Kemp, H.A., and G.F. Sprague, Jr. 2003. Far3 and five interacting proteins prevent premature recovery from pheromone arrest in the budding yeast Saccharomyces cerevisiae. *Mol Cell Biol*. 23:1750-1763.

Kendall, M.G., A. Stuart, J.K. Ord, S.F. Arnold, and A. O'Hagan. 1994. Kendall's advanced theory of statistics. Halsted Press, London, New York.

Kiburz, B.M., D.B. Reynolds, P.C. Megee, A.L. Marston, B.H. Lee, T.I. Lee, S.S. Levine, R.A. Young, and A. Amon. 2005. The core centromere and Sgo1 establish a 50-kb cohesin-protected domain around centromeres during meiosis I. *Genes Dev*. 19:3017-3030.

Kim, P.M., L.J. Lu, Y. Xia, and M.B. Gerstein. 2006. Relating three-dimensional structures to protein networks provides evolutionary insights. *Science*. 314:1938-1941.

Koehler, K.E., C.L. Boulton, H.E. Collins, R.L. French, K.C. Herman, S.M. Lacefield, L.D. Madden, C.D. Schuetz, and R.S. Hawley. 1996a. Spontaneous X chromosome MI and MII nondisjunction events in Drosophila melanogaster oocytes have different recombinational histories. *Nat Genet*. 14:406-414.

Koehler, K.E., R.S. Hawley, S. Sherman, and T. Hassold. 1996b. Recombination and nondisjunction in humans and flies. *Hum Mol Genet*. 5 Spec No:1495-1504.

Krejci, L., V. Altmannova, M. Spirek, and X. Zhao. 2012. Homologous recombination and its regulation. *Nucleic Acids Res*.

Krogan, N.J., G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A.P. Tikuisis, T. Punna, J.M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M.D. Robinson, A. Paccanaro, J.E. Bray, A. Sheung, B. Beattie, D.P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M.M. Canete, J. Vlasblom, S. Wu, C. Orsi, S.R. Collins, S. Chandran, R. Haw, J.J. Rilstone, K. Gandi, N.J. Thompson, G. Musso, P. St Onge, S. Ghanny, M.H. Lam, G. Butland, A.M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J.S. Weissman, C.J. Ingles, T.R. Hughes, J. Parkinson, M. Gerstein, S.J. Wodak, A. Emili, and J.F. Greenblatt. 2006. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature*. 440:637-643.

Kroll, E.S., K.M. Hyland, P. Hieter, and J.J. Li. 1996. Establishing genetic interactions by a synthetic dosage lethality phenotype. *Genetics*. 143:95-102.

Kunin, V., and C.A. Ouzounis. 2003. GeneTRACE-reconstruction of gene content of ancestral species. *Bioinformatics*. 19:1412-1416.

Lamb, N.E., S.B. Freeman, A. Savage-Austin, D. Pettay, L. Taft, J. Hersey, Y. Gu, J. Shen, D. Saker, K.M. May, D. Avramopoulos, M.B. Petersen, A. Hallberg, M. Mikkelsen, T.J. Hassold, and S.L. Sherman. 1996. Susceptible chiasmate configurations of chromosome 21 predispose to non-disjunction in both maternal meiosis I and meiosis II. *Nat Genet*. 14:400-405.

Lambeth, L.S., N.J. Van Hateren, S.A. Wilson, and V. Nair. 2010. A direct comparison of strategies for combinatorial RNA interference. *BMC Mol Biol*. 11:77.

Lambie, E.J., and G.S. Roeder. 1988. A yeast centromere acts in cis to inhibit meiotic gene conversion of adjacent sequences. *Cell*. 52:863-873.

Lander, E.S. 2011. Initial impact of the sequencing of the human genome. *Nature*. 470:187-197.

Lehar, J., B.R. Stockwell, G. Giaever, and C. Nislow. 2008. Combination chemical genetics. *Nat Chem Biol*. 4:674-681.

Lehner, B., C. Crombie, J. Tischler, A. Fortunato, and A.G. Fraser. 2006. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet*. 38:896-903.

Leidel, S., P.G. Pedrioli, T. Bucher, R. Brost, M. Costanzo, A. Schmidt, R. Aebersold, C. Boone, K. Hofmann, and M. Peter. 2009. Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. *Nature*. 458:228-232.

Levy, S.F., and M.L. Siegal. 2008. Network hubs buffer environmental variation in Saccharomyces cerevisiae. *PLoS Biol*. 6:e264.

Lewontin, R.C. 1974. The genetic basis of evolutionary change. Columbia University Press, New York.

Lewontin, R.C., and K. Kojima. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution*. 14:458-472.

Li, Z., F.J. Vizeacoumar, S. Bahr, J. Li, J. Warringer, F.S. Vizeacoumar, R. Min, B. Vandersluis, J. Bellay, M. Devit, J.A. Fleming, A. Stephens, J. Haase, Z.Y. Lin, A. Baryshnikova, H. Lu, Z. Yan, K. Jin, S. Barker, A. Datti, G. Giaever, C. Nislow, C. Bulawa, C.L. Myers, M. Costanzo, A.C. Gingras, Z. Zhang, A. Blomberg, K. Bloom, B. Andrews, and C. Boone. 2011. Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat Biotechnol*. 29:361-367.

Lichten, M., and A.S. Goldman. 1995. Meiotic recombination hotspots. *Annu Rev Genet*. 29:423-444.

Lin, Y.Y., S. Kiihl, Y. Suhail, S.Y. Liu, Y.H. Chou, Z. Kuang, J.Y. Lu, C.N. Khor, C.L. Lin, J.S. Bader, R. Irizarry, and J.D. Boeke. 2012. Functional dissection of lysine deacetylases reveals that HDAC1 and p300 regulate AMPK. *Nature*. 482:251-255.

Lin, Y.Y., J.Y. Lu, J. Zhang, W. Walter, W. Dang, J. Wan, S.C. Tao, J. Qian, Y. Zhao, J.D. Boeke, S.L. Berger, and H. Zhu. 2009. Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell*. 136:1073-1084.

Lin, Y.Y., Y. Qi, J.Y. Lu, X. Pan, D.S. Yuan, Y. Zhao, J.S. Bader, and J.D. Boeke. 2008. A comprehensive synthetic genetic interaction network governing yeast histone acetylation and deacetylation. *Genes Dev*. 22:2062-2074.

Luo, B., H.W. Cheung, A. Subramanian, T. Sharifnia, M. Okamoto, X. Yang, G. Hinkle, J.S. Boehm, R. Beroukhim, B.A. Weir, C. Mermel, D.A. Barbie, T. Awad, X. Zhou, T. Nguyen, B. Piqani, C. Li, T.R. Golub, M. Meyerson, N. Hacohen, W.C. Hahn, E.S. Lander, D.M. Sabatini, and D.E. Root. 2008. Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A*. 105:20380-20385.

Lynn, A., R. Soucek, and G.V. Borner. 2007. ZMM proteins during meiosis: crossover artists at work. *Chromosome Res*. 15:591-605.

Magtanong, L., C.H. Ho, S.L. Barker, W. Jiao, A. Baryshnikova, S. Bahr, A.M. Smith, L.E. Heisler, J.S. Choy, E. Kuzmin, K. Andrusiak, A. Kobylianski, Z. Li, M. Costanzo, M.A. Basrai, G. Giaever, C. Nislow, B. Andrews, and C. Boone. 2011. Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. *Nat Biotechnol*. 29:505-511.

Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L.M. Steinmetz. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*. 454:479-485.

Mani, R., R.P. St Onge, J.L.t. Hartman, G. Giaever, and F.P. Roth. 2008. Defining genetic interaction. *Proc Natl Acad Sci U S A*. 105:3461-3466.

Marcotte, R., K.R. Brown, F. Suarez, A. Sayad, K. Karamboulas, P.M. Krzyzanowski, F. Sircoulomb, M. Medrano, Y. Fedyshyn, J.L.Y. Koh, D. van Dyk, B. Fedyshyn, M. Luhova, G.C. Brito, F.J. Vizeacoumar, F.S. Vizeacoumar, A. Datti, D. Kasimer, A. Buzina, P. Mero, C. Misquitta, J. Normand, M. Haider, T. Ketela, J.L. Wrana, R. Rottapel, B.G. Neel, and J. Moffat. 2012. Essential Gene Profiles in Breast, Pancreatic, and Ovarian Cancer Cells. *Cancer Discovery*. 2:172-189.

Marsh, L., A.M. Neiman, and I. Herskowitz. 1991. Signal transduction during pheromone response in yeast. *Annu Rev Cell Biol*. 7:699-728.

Martorell, M.R., J. Benet, C. Marquez, J. Egozcue, and J. Navarro. 2000. Correlation between centromere and chromosome length in human male pronuclear chromosomes: ultrastructural analysis. *Zygote*. 8:79-85.

McClellan, A.J., Y. Xia, A.M. Deutschbauer, R.W. Davis, M. Gerstein, and J. Frydman. 2007. Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell*. 131:121-135.

Meiklejohn, C.D., and D.L. Hartl. 2002. A single mode of canalization. *Trends Ecol. Evol.* 17:468-473.

Menne, T.F., B. Goyenechea, N. Sanchez-Puig, C.C. Wong, L.M. Tonkin, P.J. Ancliff, R.L. Brost, M. Costanzo, C. Boone, and A.J. Warren. 2007. The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nat Genet*. 39:486-495.

Metzger, M.B., and S. Michaelis. 2009. Analysis of quality control substrates in distinct cellular compartments reveals a unique role for Rpn4p in tolerating misfolded membrane proteins. *Molecular biology of the cell*. 20:1006-1019.

Michalak, P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics*. 91:243-248.

Michaut, M., and G.D. Bader. 2012. Multiple genetic interaction experiments provide complementary information useful for gene function prediction. *PLoS Comput Biol*. 8:e1002559.

Miller, J.P., R.S. Lo, A. Ben-Hur, C. Desmarais, I. Stagljar, W.S. Noble, and S. Fields. 2005. Large-scale identification of yeast integral membrane protein interactions. *Proc Natl Acad Sci U S A*. 102:12123-12128.

Mitchell, A.P. 2008. A VAST staging area for regulatory proteins. *Proc Natl Acad Sci U S A*. 105:7111-7112.

Miyazaki, W.Y., and T.L. Orr-Weaver. 1994. Sister-chromatid cohesion in mitosis and meiosis. *Annu Rev Genet*. 28:167-187.

Mnaimneh, S., A.P. Davierwala, J. Haynes, J. Moffat, W.T. Peng, W. Zhang, X. Yang, J. Pootoolal, G. Chua, A. Lopez, M. Trochesset, D. Morse, N.J. Krogan, S.L. Hiley, Z. Li, Q. Morris, J. Grigull, N. Mitsakakis, C.J. Roberts, J.F. Greenblatt, C. Boone, C.A. Kaiser, B.J. Andrews, and T.R. Hughes. 2004. Exploration of essential gene functions via titratable promoter alleles. *Cell*. 118:31-44.

Moffat, J., D.A. Grueneberg, X. Yang, S.Y. Kim, A.M. Kloepfer, G. Hinkle, B. Piqani, T.M. Eisenhaure, B. Luo, J.K. Grenier, A.E. Carpenter, S.Y. Foo, S.A. Stewart, B.R. Stockwell, N. Hacohen, W.C. Hahn, E.S. Lander, D.M. Sabatini, and D.E. Root. 2006. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*. 124:1283-1298.

Morton, N.E. 2004. History of genetic mapping. *In* Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics.

Mostafavi, S., D. Ray, D. Warde-Farley, C. Grouios, and Q. Morris. 2008. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*. 9 Suppl 1:S4.

Munz, P. 1994. An analysis of interference in the fission yeast Schizosaccharomyces pombe. *Genetics*. 137:701-707.

Musso, G., M. Costanzo, M. Huangfu, A.M. Smith, J. Paw, B.J. San Luis, C. Boone, G. Giaever, C. Nislow, A. Emili, and Z. Zhang. 2008. The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Res*. 18:1092-1099.

Myers, C.L., D.R. Barrett, M.A. Hibbs, C. Huttenhower, and O.G. Troyanskaya. 2006. Finding function: evaluation methods for functional genomic data. *BMC Genomics*. 7:187.

Naumanen, T., L.D. Johansen, E.T. Coffey, and T. Kallunki. 2008. Loss-of-function of IKAP/ELP1: could neuronal migration defect underlie familial dysautonomia? *Cell Adh Migr*. 2:236-239.

Nei, M. 2003. Genome evolution: let's stick together. *Heredity (Edinb)*. 90:411-412.

Novick, P., and D. Botstein. 1985. Phenotypic analysis of temperature-sensitive yeast actin mutants. *Cell*. 40:405-416.

O'Shaughnessy, J., C. Osborne, J.E. Pippen, M. Yoffe, D. Patt, C. Rocha, I.C. Koo, B.M. Sherman, and C. Bradley. 2011. Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *N Engl J Med*. 364:205-214.

Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani, T.L. Saito, A. Saka, T. Fukuda, S. Ishihara, S. Oka, G. Suzuki, M. Watanabe, A. Hirata, M. Ohtani, H. Sawai, N. Fraysse, J.P. Latge, J.M. Francois, M. Aebi, S. Tanaka, S. Muramatsu, H. Araki, K. Sonoike, S. Nogami, and S. Morishita. 2005. High-dimensional and large-scale phenotyping of yeast mutants. *Proc Natl Acad Sci U S A*. 102:19015-19020.

Online Mendelian Inheritance in Man, http://omim.org (March 24, 2012)

Pal, C., and L.D. Hurst. 2003. Evidence for co-evolution of gene order and recombination rate. *Nat Genet*. 33:392-395.

Pal, C., B. Papp, and L.D. Hurst. 2001. Highly expressed genes in yeast evolve slowly. *Genetics*. 158:927-931.

Pan, J., M. Sasaki, R. Kniewel, H. Murakami, H.G. Blitzblau, S.E. Tischfield, X. Zhu, M.J. Neale, M. Jasin, N.D. Socci, A. Hochwagen, and S. Keeney. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell*. 144:719-731.

Pan, X., P. Ye, D.S. Yuan, X. Wang, J.S. Bader, and J.D. Boeke. 2006. A DNA integrity network in the yeast Saccharomyces cerevisiae. *Cell*. 124:1069-1081.

Pan, X., D.S. Yuan, D. Xiang, X. Wang, S. Sookhai-Mahadeo, J.S. Bader, P. Hieter, F. Spencer, and J.D. Boeke. 2004. A robust toolkit for functional profiling of the yeast genome. *Mol Cell*. 16:487-496.

Panizza, S., M.A. Mendoza, M. Berlinger, L. Huang, A. Nicolas, K. Shirahige, and F. Klein. 2011. Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. *Cell*. 146:372-383.

Park, P.J. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 10:669-680.

Parsons, A.B., R.L. Brost, H. Ding, Z. Li, C. Zhang, B. Sheikh, G.W. Brown, P.M. Kane, T.R. Hughes, and C. Boone. 2004. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat Biotechnol*. 22:62-69.

Parsons, A.B., A. Lopez, I.E. Givoni, D.E. Williams, C.A. Gray, J. Porter, G. Chua, R. Sopko, R.L. Brost, C.H. Ho, J. Wang, T. Ketela, C. Brenner, J.A. Brill, G.E. Fernandez, T.C. Lorenz, G.S. Payne, S. Ishihara, Y. Ohya, B. Andrews, T.R. Hughes, B.J. Frey, T.R. Graham, R.J. Andersen, and C. Boone. 2006. Exploring the mode-of-

action of bioactive compounds by chemical-genetic profiling in yeast. *Cell*. 126:611-625.

Pearson, C.G., P.S. Maddox, E.D. Salmon, and K. Bloom. 2001. Budding yeast chromosome structure and dynamics during mitosis. *J Cell Biol*. 152:1255-1266.

Peng, J., J.E. Elias, C.C. Thoreen, L.J. Licklider, and S.P. Gygi. 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res*. 2:43-50.

Perkins, D.N., D.J. Pappin, D.M. Creasy, and J.S. Cottrell. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 20:3551-3567.

Peters, A.D., and C.M. Lively. 2000. Epistasis and the maintenance of sex. *In* Epistasis and the evolutionary process. J.B. Wolf, E.D.I. Brodie, and M.J. Wade, editors. Oxford University Press, New York.

Petes, T.D. 2001. Meiotic recombination hot spots and cold spots. *Nat Rev Genet*. 2:360-369.

Petes, T.D., and D. Botstein. 1977. Simple Mendelian inheritance of the reiterated ribosomal DNA of yeast. *Proc Natl Acad Sci U S A*. 74:5091-5095.

Phillips, P.C., S.P. Otto, and M.C. Whitlock. 2000. Beyond the average. The evolutionary importance of gene interactions and variability of epistatic effects. *In* Epistasis and the evolutionary process. J.B. Wolf, E.D.I. Brodie, and M.J. Wade, editors. Oxford University Press, New York, NY. 20-38.

Posas, F., J. Clotet, M.T. Muns, J. Corominas, A. Casamayor, and J. Arino. 1993. The gene PPG encodes a novel yeast protein phosphatase involved in glycogen accumulation. *J Biol Chem*. 268:1349-1354.

Prelich, G. 1999. Suppression mechanisms: themes from variations. *Trends Genet*. 15:261-266.

Ptacek, J., G. Devgan, G. Michaud, H. Zhu, X. Zhu, J. Fasolo, H. Guo, G. Jona, A. Breitkreutz, R. Sopko, R.R. McCartney, M.C. Schmidt, N. Rachidi, S.J. Lee, A.S. Mah, L. Meng, M.J. Stark, D.F. Stern, C. De Virgilio, M. Tyers, B. Andrews, M. Gerstein, B. Schweitzer, P.F. Predki, and M. Snyder. 2005. Global analysis of protein phosphorylation in yeast. *Nature*. 438:679-684.

Pu, S., J. Wong, B. Turner, E. Cho, and S.J. Wodak. 2009. Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Res*. 37:825-831.

Rahl, P.B., C.Z. Chen, and R.N. Collins. 2005. Elp1p, the yeast homolog of the FD disease syndrome protein, negatively regulates exocytosis independently of transcriptional elongation. *Mol Cell*. 17:841-853.

Reece-Hoyes, J.S., A. Diallo, B. Lajoie, A. Kent, S. Shrestha, S. Kadreppa, C. Pesyna, J. Dekker, C.L. Myers, and A.J. Walhout. 2011. Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat Methods*. 8:1059-1064.

Risinger, A.L., N.E. Cain, E.J. Chen, and C.A. Kaiser. 2006. Activity-dependent reversible inactivation of the general amino acid permease. *Molecular biology of the cell*. 17:4411-4419.

Roberg, K.J., S. Bickel, N. Rowley, and C.A. Kaiser. 1997. Control of amino acid permease sorting in the late secretory pathway of Saccharomyces cerevisiae by SEC13, LST4, LST7 and LST8. *Genetics*. 147:1569-1584.

Roberts, C.J., B. Nelson, M.J. Marton, R. Stoughton, M.R. Meyer, H.A. Bennett, Y.D. He, H. Dai, W.L. Walker, T.R. Hughes, M. Tyers, C. Boone, and S.H. Friend. 2000. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science*. 287:873-880.

Roberts, N.J., J.T. Vogelstein, G. Parmigiani, K.W. Kinzler, B. Vogelstein, and V.E. Velculescu. 2012. The Predictive Capacity of Personal Genome Sequencing. *Sci Transl Med*.

Rockman, M.V., and L. Kruglyak. 2009. Recombinational landscape and population genomics of Caenorhabditis elegans. *PLoS Genet*. 5:e1000419.

Rockmill, B., and G.S. Roeder. 1988. RED1: a yeast gene required for the segregation of chromosomes during the reductional division of meiosis. *Proc Natl Acad Sci U S A*. 85:6057-6061.

Rockmill, B., K. Voelkel-Meiman, and G.S. Roeder. 2006. Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in Saccharomyces cerevisiae. *Genetics*. 174:1745-1754.

Roguev, A., S. Bandyopadhyay, M. Zofall, K. Zhang, T. Fischer, S.R. Collins, H. Qu, M. Shales, H.O. Park, J. Hayles, K.L. Hoe, D.U. Kim, T. Ideker, S.I. Grewal, J.S. Weissman, and N.J. Krogan. 2008. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*. 322:405-410.

Roma-Mateo, C., A. Sacristan-Reviriego, N.J. Beresford, J.A. Caparros-Martin, F.A. Culianez-Macia, H. Martin, M. Molina, L. Tabernero, and R. Pulido. 2011. Phylogenetic and genetic linkage between novel atypical dual-specificity phosphatases from non-metazoan organisms. *Mol Genet Genomics*. 285:341-354.

Rost, B. 1999. Twilight zone of protein sequence alignments. *Protein Eng*. 12:85-94.

Roth, F.P., Y. Suzuki, and J. Mellor. 2009. Methods and applications for stitched DNA barcodes. Vol. US 2009/0098555 A1. USPTO, editor. President and Fellows of Harvard College, Cambridge, MA, United States.

Rothfels, K., J.C. Tanny, E. Molnar, H. Friesen, C. Commisso, and J. Segall. 2005. Components of the ESCRT pathway, DFG16, and YGR122w are required for Rim101 to act as a corepressor with Nrg1 at the negative regulatory element of the DIT1 gene of Saccharomyces cerevisiae. *Mol Cell Biol*. 25:6772-6788.

San Filippo, J., P. Sung, and H. Klein. 2008. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem*. 77:229-257.

Sanchez, L., P. Martinez, and V. Goyanes. 1991. Analysis of centromere size in human chromosomes 1, 9, 15, and 16 by electron microscopy. *Genome*. 34:710-713.

Scannell, D.R., O.A. Zill, A. Rokas, C. Payen, M.J. Dunham, M.B. Eisen, J. Rine, M. Johnston, and C.T. Hittinger. 2011. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3 (Bethesda)*. 1:11-25.

Schacherer, J., D.M. Ruderfer, D. Gresham, K. Dolinski, D. Botstein, and L. Kruglyak. 2007. Genome-wide analysis of nucleotide-level variation in commonly used Saccharomyces cerevisiae strains. *PLoS One*. 2:e322.

Schadt, E.E., M.D. Linderman, J. Sorenson, L. Lee, and G.P. Nolan. 2010. Computational solutions to large-scale data management and analysis. *Nat Rev Genet*. 11:647-657.

Schuldiner, M., S.R. Collins, N.J. Thompson, V. Denic, A. Bhamidipati, T. Punna, J. Ihmels, B. Andrews, C. Boone, J.F. Greenblatt, J.S. Weissman, and N.J. Krogan. 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell*. 123:507-519.

Segre, D., A. Deluna, G.M. Church, and R. Kishony. 2005. Modular epistasis in yeast metabolism. *Nat Genet*. 37:77-83.

Sevier, C.S., and C.A. Kaiser. 2006. Disulfide transfer between two conserved cysteine pairs imparts selectivity to protein oxidation by Ero1. *Molecular biology of the cell*. 17:2256-2266.

Sevier, C.S., H. Qu, N. Heldman, E. Gross, D. Fass, and C.A. Kaiser. 2007. Modulation of cellular disulfide-bond formation and the ER redox environment by feedback regulation of Ero1. *Cell*. 129:333-344.

Sharifpoor, S., D. van Dyk, M. Costanzo, A. Baryshnikova, H. Friesen, A.C. Douglas, J.Y. Youn, B. Vandersluis, C.L. Myers, B. Papp, C. Boone, and B.J. Andrews. 2012. Functional wiring of the yeast kinome revealed by global analysis of genetic network motifs. *Genome Res*. 22:791-801.

Sherman, S.L., N.E. Lamb, and E. Feingold. 2006. Relationship of recombination patterns and maternal age among non-disjoined chromosomes 21. *Biochem Soc Trans*. 34:578-580.

Shor, E., J. Weinstein, and R. Rothstein. 2005. A genetic screen for top3 suppressors in Saccharomyces cerevisiae identifies SHU1, SHU2, PSY3 and CSM2: four genes involved in error-free DNA repair. *Genetics*. 169:1275-1289.

Silva, J.M., K. Marran, J.S. Parker, J. Silva, M. Golding, M.R. Schlabach, S.J. Elledge, G.J. Hannon, and K. Chang. 2008. Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science*. 319:617-620.

Smith, A.M., L.E. Heisler, J. Mellor, F. Kaper, M.J. Thompson, M. Chee, F.P. Roth, G. Giaever, and C. Nislow. 2009. Quantitative phenotyping via deep barcode sequencing. *Genome Res*. 19:1836-1842.

Smoot, M.E., K. Ono, J. Ruscheinski, P.L. Wang, and T. Ideker. 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 27:431-432.

Sneddon, A.A., P.T. Cohen, and M.J. Stark. 1990. Saccharomyces cerevisiae protein phosphatase 2A performs an essential cellular function and is encoded by two genes. *Embo J*. 9:4339-4346.

Sopko, R., D. Huang, N. Preston, G. Chua, B. Papp, K. Kafadar, M. Snyder, S.G. Oliver, M. Cyert, T.R. Hughes, C. Boone, and B. Andrews. 2006. Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell*. 21:319-330.

Spellman, P.T., G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Molecular biology of the cell*. 9:3273-3297.

St Onge, R.P., R. Mani, J. Oh, M. Proctor, E. Fung, R.W. Davis, C. Nislow, F.P. Roth, and G. Giaever. 2007. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat Genet*. 39:199-206.

Stark, C., B.J. Breitkreutz, A. Chatr-Aryamontri, L. Boucher, R. Oughtred, M.S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Reguly, J.M. Rust, A. Winter, K. Dolinski, and M. Tyers. 2011. The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res*. 39:D698-704.

Stevenson, B.J., N. Rhodes, B. Errede, and G.F. Sprague, Jr. 1992. Constitutive mutants of the protein kinase STE11 activate the yeast pheromone response pathway in the absence of the G protein. *Genes Dev*. 6:1293-1304.

Straight, A.F., W.F. Marshall, J.W. Sedat, and A.W. Murray. 1997. Mitosis in living budding yeast: anaphase A but no metaphase plate. *Science*. 277:574-578.

Su, Y., A.B. Barton, and D.B. Kaback. 2000. Decreased meiotic reciprocal recombination in subtelomeric regions in Saccharomyces cerevisiae. *Chromosoma*. 109:467-475.

Sung, P. 1997. Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes Dev*. 11:1111-1121.

Sunnerhagen, P. 2002. Prospects for functional genomics in Schizosaccharomyces pombe. *Curr Genet*. 42:73-84.

Sym, M., J.A. Engebrecht, and G.S. Roeder. 1993. ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. *Cell*. 72:365-378.

Sym, M., and G.S. Roeder. 1994. Crossover interference is abolished in the absence of a synaptonemal complex protein. *Cell*. 79:283-292.

Szappanos, B., K. Kovacs, B. Szamecz, F. Honti, M. Costanzo, A. Baryshnikova, G. Gelius-Dietrich, M.J. Lercher, M. Jelasity, C.L. Myers, B.J. Andrews, C. Boone, S.G. Oliver, C. Pal, and B. Papp. 2011. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet*. 43:656-662.

Tanaka, T., J. Fuchs, J. Loidl, and K. Nasmyth. 2000. Cohesin ensures bipolar attachment of microtubules to sister centromeres and resists their precocious separation. *Nat Cell Biol*. 2:492-499.

Tanaka, T.U., N. Rachidi, C. Janke, G. Pereira, M. Galova, E. Schiebel, M.J. Stark, and K. Nasmyth. 2002. Evidence that the Ipl1-Sli15 (Aurora kinase-INCENP) complex promotes chromosome bi-orientation by altering kinetochore-spindle pole connections. *Cell*. 108:317-329.

Tarailo, M., S. Tarailo, and A.M. Rose. 2007. Synthetic lethal interactions identify phenotypic "interologs" of the spindle assembly checkpoint components. *Genetics*. 177:2525-2530.

Tarassov, K., V. Messier, C.R. Landry, S. Radinovic, M.M. Serna Molina, I. Shames, Y. Malitskaya, J. Vogel, H. Bussey, and S.W. Michnick. 2008. An in vivo map of the yeast protein interactome. *Science*. 320:1465-1470.

Tewari, M., P.J. Hu, J.S. Ahn, N. Ayivi-Guedehoussou, P.O. Vidalain, S. Li, S. Milstein, C.M. Armstrong, M. Boxem, M.D. Butler, S. Busiguina, J.F. Rual, N. Ibarrola, S.T. Chaklos, N. Bertin, P. Vaglio, M.L. Edgley, K.V. King, P.S. Albert, J. Vandenhaute, A. Pandey, D.L. Riddle, G. Ruvkun, and M. Vidal. 2004. Systematic interactome mapping and genetic perturbation analysis of a C. elegans TGF-beta signaling network. *Mol Cell*. 13:469-482.

The International HapMap Consortium. 2003. The International HapMap Project. *Nature*. 426:789-796.

Tischler, J., B. Lehner, and A.G. Fraser. 2008. Evolutionary plasticity of genetic interaction networks. *Nat Genet*. 40:390-391.

Tong, A.H., M. Evangelista, A.B. Parsons, H. Xu, G.D. Bader, N. Page, M. Robinson, S. Raghibizadeh, C.W. Hogue, H. Bussey, B. Andrews, M. Tyers, and C. Boone. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*. 294:2364-2368.

Tong, A.H., G. Lesage, G.D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G.F. Berriz, R.L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D.S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J.N. Levinson, H. Lu, P. Menard, C. Munyana, A.B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A.M. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S.L. Wong, L.V. Zhang, H. Zhu, C.G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F.P. Roth, G.W. Brown, B. Andrews, H. Bussey, and C. Boone. 2004. Global mapping of the yeast genetic interaction network. *Science*. 303:808-813.

Typas, A., R.J. Nichols, D.A. Siegele, M. Shales, S.R. Collins, B. Lim, H. Braberg, N. Yamamoto, R. Takeuchi, B.L. Wanner, H. Mori, J.S. Weissman, N.J. Krogan, and C.A. Gross. 2008. High-throughput, quantitative analyses of genetic interactions in E. coli. *Nat Methods*. 5:781-787.

Ulrich, H.D., and S. Jentsch. 2000. Two RING finger proteins mediate cooperation between ubiquitin-conjugating enzymes in DNA repair. *Embo J*. 19:3388-3397.

Ungar, D., T. Oka, M. Krieger, and F.M. Hughson. 2006. Retrograde transport on the COG railway. *Trends Cell Biol*. 16:113-120.

van Dongen, S. 2000. Graph Clustering by Flow Simulation. Vol. PhD. University of Utrecht.

Van Driessche, N., J. Demsar, E.O. Booth, P. Hill, P. Juvan, B. Zupan, A. Kuspa, and G. Shaulsky. 2005. Epistasis analysis with global transcriptional phenotypes. *Nat Genet*. 37:471-477.

VanderSluis, B., J. Bellay, G. Musso, M. Costanzo, B. Papp, F.J. Vizeacoumar, A. Baryshnikova, B. Andrews, C. Boone, and C.L. Myers. 2010. Genetic interactions reveal the evolutionary trajectories of duplicate genes. *Mol Syst Biol*. 6:429.

Vizeacoumar, F.J., N. van Dyk, S.V. F, V. Cheung, J. Li, Y. Sydorskyy, N. Case, Z. Li, A. Datti, C. Nislow, B. Raught, Z. Zhang, B. Frey, K. Bloom, C. Boone, and B.J. Andrews. 2010. Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. *J Cell Biol*. 188:69-81.

Waddington, C.H. 1957. The Strategy of the Gene. Allen & Unwin, London.

Wang, J., W. Wang, R. Li, Y. Li, G. Tian, L. Goodman, W. Fan, J. Zhang, J. Li, Y. Guo, B. Feng, H. Li, Y. Lu, X. Fang, H. Liang, Z. Du, D. Li, Y. Zhao, Y. Hu, Z. Yang, H. Zheng, I. Hellmann, M. Inouye, J. Pool, X. Yi, J. Zhao, J. Duan, Y. Zhou, J. Qin, L. Ma, G. Li, G. Zhang, B. Yang, C. Yu, F. Liang, W. Li, S. Li, P. Ni, J. Ruan, Q. Li, H. Zhu, D. Liu, Z. Lu, N. Li, G. Guo, J. Ye, L. Fang, Q. Hao, Q. Chen, Y. Liang, Y. Su, A. San, C. Ping, S. Yang, F. Chen, L. Li, K. Zhou, Y. Ren, L. Yang, Y. Gao, G. Yang, Z. Li, X. Feng, K. Kristiansen, G.K. Wong, R. Nielsen, R. Durbin, L. Bolund, X. Zhang, and H. Yang. 2008. The diploid genome sequence of an Asian individual. *Nature*. 456:60-65.

Wapinski, I., A. Pfeffer, N. Friedman, and A. Regev. 2007. Natural history and evolutionary principles of gene duplication in fungi. *Nature*. 449:54-61.

Ward, J.J., L.J. McGuffin, K. Bryson, B.F. Buxton, and D.T. Jones. 2004. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. 20:2138-2139.

Warringer, J., E. Ericson, L. Fernandez, O. Nerman, and A. Blomberg. 2003. High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc Natl Acad Sci U S A*. 100:15724-15729.

Westfall, P.H., and S.S. Young. 1993. Resampling-based Multiple Testing: Examples and Methods for p-value Adjustment. Wiley, New York.

White, M.A., P. Detloff, M. Strand, and T.D. Petes. 1992. A promoter deletion reduces the rate of mitotic, but not meiotic, recombination at the HIS4 locus in yeast. *Curr Genet*. 21:109-116.

White, M.A., M. Dominska, and T.D. Petes. 1993. Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 90:6621-6625.

Wilmes, G.M., M. Bergkessel, S. Bandyopadhyay, M. Shales, H. Braberg, G. Cagney, S.R. Collins, G.B. Whitworth, T.L. Kress, J.S. Weissman, T. Ideker, C. Guthrie, and N.J. Krogan. 2008. A genetic interaction map of RNA-processing factors reveals links

between Sem1/Dss1-containing complexes and mRNA export and splicing. *Mol Cell*. 32:735-746.

Wu, M., B. Repetto, D.M. Glerum, and A. Tzagoloff. 1995. Cloning and characterization of FAD1, the structural gene for flavin adenine dinucleotide synthetase of Saccharomyces cerevisiae. *Mol Cell Biol*. 15:264-271.

Wu, T.C., and M. Lichten. 1994. Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science*. 263:515-518.

Yang, W.S., and B.R. Stockwell. 2008. Inhibition of casein kinase 1-epsilon induces cancer-cell-selective, PERIOD2-dependent growth arrest. *Genome Biol*. 9:R92.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586-1591.

Ye, P., B.D. Peyser, X. Pan, J.D. Boeke, F.A. Spencer, and J.S. Bader. 2005. Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol Syst Biol*. 1:2005 0026.

Yeh, E., J. Haase, L.V. Paliulis, A. Joglekar, L. Bond, D. Bouck, E.D. Salmon, and K.S. Bloom. 2008. Pericentric chromatin is organized into an intramolecular loop in mitosis. *Curr Biol*. 18:81-90.

Yu, H., P. Braun, M.A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.F. Rual, A. Dricot, A. Vazquez, R.R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.S. de Smet, A. Motyl, M.E. Hudson, J. Park, X. Xin, M.E. Cusick, T. Moore, C. Boone, M. Snyder, F.P. Roth, A.L. Barabasi, J. Tavernier, D.E. Hill, and M. Vidal. 2008. High-quality binary protein interaction map of the yeast interactome network. *Science*. 322:104-110.

Yvert, G., R.B. Brem, J. Whittle, J.M. Akey, E. Foss, E.N. Smith, R. Mackelprang, and L. Kruglyak. 2003. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat Genet*. 35:57-64.

Zhao, R., M. Davey, Y.C. Hsu, P. Kaplanek, A. Tong, A.B. Parsons, N. Krogan, G. Cagney, D. Mai, J. Greenblatt, C. Boone, A. Emili, and W.A. Houry. 2005. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell*. 120:715-727.

Zheng, J., J.J. Benschop, M. Shales, P. Kemmeren, J. Greenblatt, G. Cagney, F. Holstege, H. Li, and N.J. Krogan. 2010. Epistatic relationships reveal the functional organization of yeast transcription factors. *Mol Syst Biol*. 6:420.

Zickler, D., and N. Kleckner. 1998. The leptotene-zygotene transition of meiosis. *Annu Rev Genet*. 32:619-697.

Zickler, D., and N. Kleckner. 1999. Meiotic chromosomes: integrating structure and function. *Annu Rev Genet*. 33:603-754.

Zinkowski, R.P., J. Meyne, and B.R. Brinkley. 1991. The centromere-kinetochore complex: a repeat subunit model. *J Cell Biol*. 113:1091-1110.

Zinman, G.E., S. Zhong, and Z. Bar-Joseph. 2011. Biological interaction networks are conserved at the module level. *BMC Syst Biol*. 5:134.

Zlatanova, J., and A. Thakar. 2008. H2A.Z: view from the top. *Structure*. 16:166-179.

Zuk, O., E. Hechter, S.R. Sunyaev, and E.S. Lander. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A*. 109:1193-1198.