Several Consequences of Aneuploidy

By

Jason Sheltzer
B.A., Molecular Biology
Princeton University

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015 [September 2015]

Signature of Author: Signature redacted

Department of Biology
June 16th, 2015

Certified by: Signature redacted

Angelika Amon
Professor of Biology
Thesis Supervisor

Accepted by: Signature redacted

Michael Hemann
Professor of Biology
Chair, Committee for Graduate Students

Several Consequences of Aneuploidy

By

Jason Sheltzer

Submitted to the Department of Biology
On June 16th, 2015 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in Biology

Abstract

Whole-chromosome aneuploidy, or a karyotype that is not a multiple of the haploid complement, is the most common cause of miscarriage and developmental delay in humans. Aneuploidy is also a hallmark of cancer: greater than 90% of tumors display chromosomal copy number alterations. Thus, understanding the consequences of aneuploidy has broad relevance for human health and development. To that end, I have studied several distinct aspects of aneuploid cell biology. In the budding yeast *Saccharomyces cerevisiae*, I demonstrated that single-chromosome gains are sufficient to induce numerous forms of genomic instability. Aneuploid yeast strains displayed increased rates of forward mutation, mitotic recombination, chromosome loss, and double-strand break formation, which could significantly impact the evolution of tumor genomes. Secondly, I characterized the effects of aneuploidy on gene expression. I established that aneuploidy induced a transcriptional stress response that was independent of the identity of the extra chromosome and was remarkably well-conserved among eukaryotes. This gene expression program was apparent in trisomic primary cells as well as in chromosomally-unstable cancer cells. Thirdly, I compared the tumorigenicity of euploid and trisomic cell lines that were genetically-identical and differed only in karyotype. I discovered that under most circumstances, aneuploidy impeded proliferation, anchorage-independent growth, and tumor formation in xenografts. Thus, single-chromosome aneuploidy actually functions as a tumor suppressor, rather than a tumor-promoting agent. In total, these results shed light on the diverse ways that chromosomal imbalances can alter the physiology of normal cells and of cancer.

Thesis Supervisor: Angelika Amon
Title: Professor of Biology

This thesis is dedicated to my grandparents.

Acknowledgments

First and foremost, I would like to thank my advisor, Angelika Amon. This dissertation would not have been possible without your support and guidance. I took a rather circuitous route through graduate school, working on everything from yeast translesion polymerases to cancer metastasis assays, and I'm thankful that you not just tolerated but actively encouraged my forays into different fields. You are a singular individual, and you run an incredibly unique lab, and I'm deeply grateful that I had the chance to work with you for the past six years. Thanks also to all of the members of the Amon lab, past and present, for taking part in stimulating scientific debates, teaching me new techniques, and also for tolerating my group meetings.

I had a terrific thesis committee during my time at MIT. Thanks to Iain Cheeseman and Mike Hemann for providing me with significant scientific help and career advice both during and outside of our meetings. Thanks also to David Pellman, who has agreed to join my committee for my dissertation defense. Beyond my committee, I'd like to thank Mike Laub, Aviv Regev, Frank Solomon, and Terry Orr-Weaver, who provided me with incredibly helpful feedback on papers that I've given them over the years. I've also had very fruitful collaborations with Maitreya Dunham and Zuzana Storchova, and I'm grateful that I had the chance to work with both of them during my time in grad school.

Thanks to all of the awesome members of Biograds2008: Josh, MK, Ricks, Will (an honorary member), Kaplan, Coppe (another honorary member), Lori, and Jenny. We had some great times together over the years, from the Pit to the ShingleRicks wedding. Thanks for keeping me (mostly) sane.

I'd also like to thank my undergraduate advisor, Mark Rose, for teaching me about the process of scientific inquiry, and for not thinking I was an idiot when I spent a week accidentally culturing *E. coli* instead of yeast. I owe much of what I've learned about how to approach biological research from the time that I spent in the Rose lab.

Finally, to Joan and Priyanka. I am so incredibly lucky to have met you. Joan, from Café Luna with a missing wallet to getting interviewed together by *Science*, it has been an incredible journey. I quite literally could not have done half of the work in this thesis without your assistance. I'm scared about living in separate cities next year, but I'm hopeful that, like most things in our privileged lives, it will work out in the end. Pri, your kindness, courage, and intelligence are such an inspiration to me. I've loved our journeys together, from Conundrum to Ponta Delgada, and I'm looking forward to so many more in the future. Thank you for all of the time that we've spent together and for all of the things we've done.

Table of Contents

Chapter 1: Introduction

Some sections of this chapter have been reproduced from:

Sheltzer, J.M., and Amon, A. "The aneuploidy paradox: costs and benefits of an incorrect karyotype. Trends in Genetics, 2011:446-53.

Introduction

Aneuploidy, an incorrect karyotype

Eukaryotic organisms have evolved robust mechanisms to ensure accurate segregation of the genetic material during mitosis. Cell cycle checkpoints delay chromosome segregation until DNA replication has been completed and sister chromatids are properly aligned at the metaphase plate. However, these safeguards occasionally fail, resulting in daughter cells that have gained or lost portions of the genetic material. Since the seminal observations of von Hansemann and Boveri at the start of the 20th century, these unequal cell divisions have been associated with developmental defects and cancer (1). Yet, determining how exactly the aneuploid state impacts cell physiology has remained elusive. In recent years, several models of aneuploidy have been developed to study the consequences of chromosome segregation errors (2–6). In my graduate work, I have investigated various aspects of aneuploid biology in several species, including yeast, mice, and humans. In this introduction, I will summarize our current understanding of the ways in which aneuploidy impacts cells and organisms, as well as the molecular underpinnings of aneuploidy-associated phenotypes.

For the purposes of this introduction, we will limit our discussion to *aneuploidy*, defined as changes in karyotype that are not whole-number multiples of the haploid complement. Aneuploidy differs from *polyploidy*, in which cells gain a balanced set of all chromosomes. Polyploidy appears to be better-tolerated than aneuploidy at the cellular level, and the specific defects associated with imbalanced karyotypes are not present in polyploid cells (7–9). Additionally, aneuploidy must be differentiated from the closely related concept of chromosomal instability (CIN). Aneuploidy is a description of a cellular *state*; it specifically describes a cell whose karyotype is not a whole-number multiple of the haploid complement. Chromosomal instability is most accurately characterized as a *rate*; it refers to a cell that missegregates chromosomes more frequently than a wild-type cell does. Cells may become aneuploid without CIN, as chromosome missegregation events occur at low levels in normal cells. The distinction

between aneuploidy and CIN may be particularly important for understanding genome alterations in cancer (see Chapters 4 and 5). Though aneuploidy and CIN are not identical, the study of CIN has contributed to our understanding of aneuploidy, and several results from CIN cells will be described below.

Aneuploidy is detrimental to cell and organismal fitness

Aneuploidy impairs normal development and is rare in adult humans

Whole-organism aneuploidy is the most common cause of miscarriage and mental retardation in humans (10,11). All human monosomies and 20 out of the 23 possible autosomal trisomies are embryonic lethal. Of the three trisomies which are viable at birth, only one, Trisomy 21/Down syndrome, can survive until adulthood (12). Similarly, in mice, all autosomal aneuploidies are embryonic lethal, with the exception of Trisomy 19, which dies shortly after birth (13). Aneuploidy has also been associated with developmental defects and lethality in a variety of other model organisms, including maize (14), flies (15), and nematodes (16–18). Thus, aneuploidy presents a considerable barrier toward successful development.

The consequences of somatic cellular aneuploidy are less well understood. It has previously been reported that mammalian hepatocytes and neurons display significant whole-chromosome aneuploidy (19,20). However, careful examination using single cell sequencing has revealed that aneuploidy is present in only about 1-2% of cells in these tissues, similar to the levels observed in skin (21). Thus, in most humans, somatic aneuploidy is rare. When somatic aneuploidy is present, it is typically the result of a clinical condition called Mosaic Variegated Aneuploidy (MVA), which is caused by bi-allelic mutations in either *BUB1B* or *CEP57* (22,23). Individuals with MVA present with random somatic aneuploidies in various tissues, and they typically exhibit growth retardation, developmental delays, and facial dysmorphia. Thus,

somatic aneuploidies with less than 100% cellular penetrance can still have devastating consequences on health and development.

Cellular aneuploidy slows the rate of cell proliferation

In addition to its effects at the organismal level, several lines of evidence demonstrate that aneuploidy decreases the rate of cell proliferation. The first studies on aneuploidy and cell division were performed in fibroblasts from individuals with Down syndrome; these cells were found to divide more slowly than age-matched euploid controls (24). More recently, several labs have investigated the consequences of aneuploidies which result from genetic ablation of the spindle assembly checkpoint (SAC). These cells exhibit CIN and develop highly aneuploid, non-clonal karyotypes. While mouse models of CIN are generally tumor prone (see below), individual cells with a CIN phenotype proliferate more slowly than euploid cells (25–27). An alternate approach, pioneered by the Compton lab, has used transient treatment with spindle poisons to induce random aneuploidy in otherwise diploid cell lines. As with other models of aneuploidy, cells with chemically-induced aberrant karyotypes were found to proliferate more slowly than euploid cells (5,28).

In order to gain a systematic understanding of the effects of aneuploidy on cell physiology, our lab has created 17 haploid strains of the budding yeast *Saccharomyces cerevisiae* which carry two copies of one or a few yeast chromosomes. All disomic strains were found to proliferate more slowly than an isogenic euploid strain (2), though the doubling time varied between the disomes. Similarly, nearly all aneuploid strains derived by sporulation of triploid and pentaploid strains displayed impaired proliferation under normal growth conditions in *S. cerevisiae* (4) and in *S. pombe* (29,30). Additionally, we have used a breeding scheme involving naturally-occurring Robertsonian chromosome fusions which allows us to harvest sibling-matched trisomic and euploid mouse embryonic fibroblasts (3). In all cases, trisomic fibroblasts were found to divide more slowly than euploid controls. Finally, these experiments

have been expanded to human cells by the Storchova lab, who observed reduced proliferative capacity following microcell-mediated chromosome transfer (MMCT) of extra chromosomes into euploid RPE-1 and HCT116 cells (6). The nature of the proliferative delay in aneuploid cells has proven to be elusive, though evidence in yeast and human cells suggests that it predominantly occurs during the G1 phase of the cell cycle (6,31). Thus, in a variety of aneuploid models, an incorrect karyotype reduces the proliferative capacity of cells.

Why is aneuploidy detrimental to cellular fitness?

Several factors may contribute to the detrimental phenotypes associated with aneuploidy. By definition, aneuploid cells contain different quantities of DNA than euploid cells do. However, it is unlikely that extra DNA alone is sufficient to impair cell fitness in most cases. Yeast strains carrying large artificial chromosomes containing human or mouse DNA (which presumably encode few or no genes that are expressed in yeast) do not exhibit proliferation defects (2). An overabundance of specific DNA sequences may confer some toxicity, as cells carrying >10 extra centromeres display a metaphase delay and increased chromosome loss, though these effects are not observed when fewer excess centromeres are present (32,33). In humans, Jiang et al. (2013) introduced the *Xist* non-coding RNA into one copy of chromosome 21 in Down syndrome IPs cells (34). Induction of *Xist* resulted in enhanced proliferation and neural rosette formation in these cells, even though they still carried a now-transcriptionally silent third copy of chromosome 21. Thus, low levels of superfluous DNA appear to be benign at the cellular level. Instead, the most likely explanation for the detrimental phenotypes associated with aneuploidy is the *gene dosage hypothesis*: gains or losses of whole chromosomes immediately alter the dosage of hundreds of genes in a cell, thereby leading to imbalances in critical proteins (Fig. 1A). Several lines of evidence support this hypothesis. First, aneuploid chromosomes appear to be expressed in a variety of contexts. Tissue from

individuals with Down syndrome generally show upregulation of transcripts from chromosome 21 (35,36), and trisomic plants as well as fibroblasts from trisomic mouse embryos display a proportional increase in mRNA levels from the additional chromosomes (3,37). Aneuploidy correlates with altered transcript levels in yeast and human cells (2,6,38), and, importantly, quantitative mass spectrometry demonstrates that aberrant karyotypes cause proportional imbalances in the relative levels of most endogenous proteins as well (4,6,39,40). Secondly, the severity of the phenotypes caused by aneuploidy correlates with the number of genes gained or lost. The three human trisomies which survive until birth have the fewest protein-coding genes on them, while the survival of trisomic mouse embryos *in utero* correlates with the number of genes on the additional chromosome [Fig. 1B; (41)]. Similarly, in disomic yeast, the delay in cell division correlates with the number of ORFs on the additional chromosome (Fig. 1C), though the presence of particularly toxic genes (i.e. β-tubulin) on small chromosomes can cause disproportionate effects (2). Finally, an increasing body of evidence demonstrates that 2-fold and even 1.5-fold changes in gene copy number can have significant effects on cellular and organismal phenotypes. Down syndrome is associated with a decreased frequency of solid tumor formation (42); studies in mice suggest that single extra copies of Ets2 and DSCR1 (both located on chromosome 21) may confer protection from tumorigenesis (43,44). Single extra copies of Sir2 prolong lifespan in yeast (45), while a single extra copy of α-synuclein predisposes individuals to Parkinson's disease (46). However, we note that some organisms, including *Drosophila melanogaster*, appear to maintain mechanisms for dosage compensation which dampen the phenotypic effects of aneuploidy (47,48). In these organisms, gene dosage may not be a robust predictor of protein expression levels, and aneuploid phenotypes may result from the overabundance of specific genes which escape attenuation.

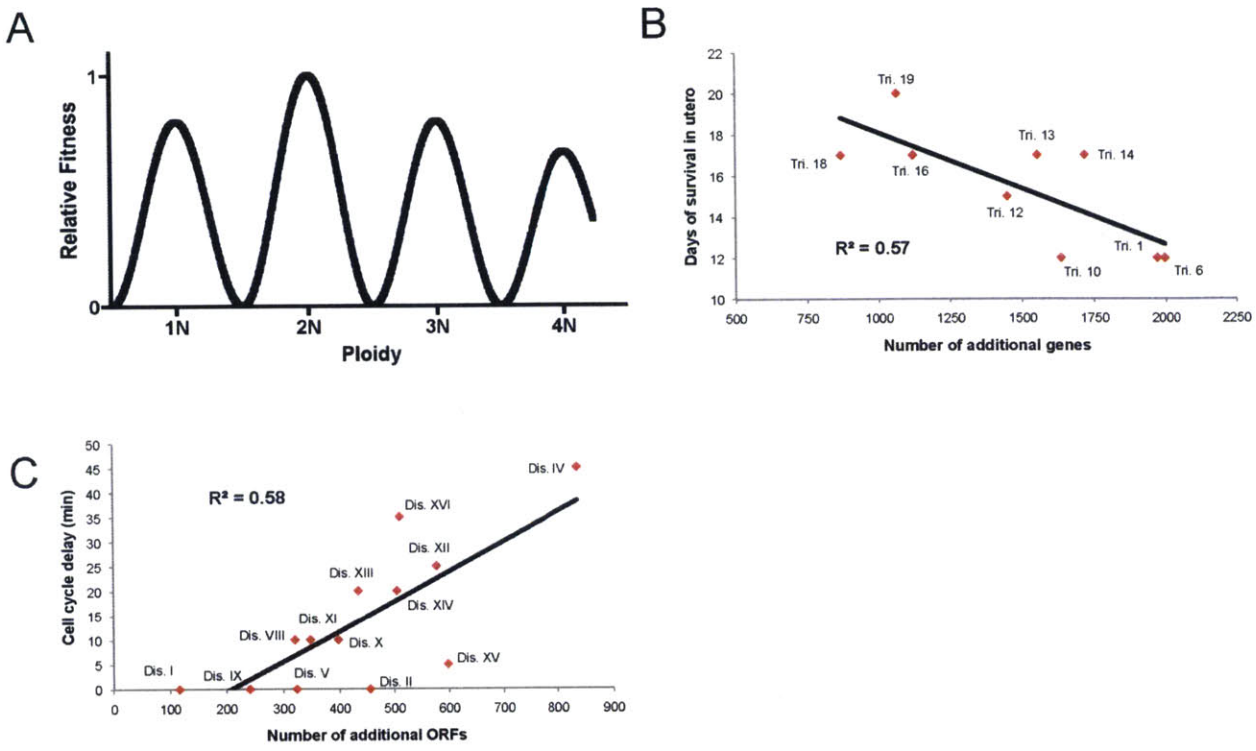


Figure 1. Cellular and organismal fitness vary according to the deviation from euploidy. (A) A model of the effect of ploidy on cell fitness. Cells with balanced sets of all chromosomes are generally healthy, although haploid and polyploid cells are moderately less fit than are diploid cells. As cells move away from euploidy, fitness decreases, and greater karyotypic imbalances typically cause more severe phenotypes. Note that transformed cells are able to tolerate a high degree of aneuploidy via mechanisms that are largely unknown. (B) The survival *in utero* of trisomic mouse embryos correlates with the number of genes encoded by the additional chromosome. A linear correlation is plotted against the data. (C) The cell cycle delay in aneuploid yeast strains correlates with the number of open reading frames (ORFs) encoded by the additional chromosome. A linear correlation is plotted against the data. Adapted from (41) and (2).

How can alterations in gene dosage levels affect cell fitness? First, changes in the concentration of a particular protein can directly modulate the efficiency of that protein's cellular function (Fig. 2A). For instance, *DSCR1* is a negative regulator of angiogenic signaling, and an additional copy of *DSCR1* in a mouse model of Down syndrome may block tumor formation by inhibiting angiogenesis (43). Secondly, changes in gene copy number can affect the formation or function of stoichiometry-sensitive compounds (Fig. 2B). Proper stoichiometry can be disrupted by both under-expression [i.e., haploinsufficiency; (49,50)] and over-expression of proteins in complexes. In *S. cerevisiae*, a single extra copy of the phosphatase *CDC14* delays cell cycle progression, as it is no longer inhibited by its 1:1 stoichiometric binding partner *CFI1/NET1* (51). Finally, an intriguing recent report suggests that a key determinant of dosage sensitivity is the susceptibility of some proteins to make promiscuous molecular interactions (52). Genes which are toxic when over-expressed are enriched for those that have many low-affinity binary interaction partners, suggesting that mass-action driven off-target interactions may impair cellular fitness in aneuploid cells (Fig. 2C). This type of promiscuous interaction may explain the gain-of-function phenotype in signaling pathways in certain cancers (3,52). For instance, when lung cancers with *EGFR*-activating mutations are treated with an *EGFR* inhibitor, some cells acquire drug resistance via amplification of the *MET* oncogene (53,54). Over-expression of *MET* leads to activation of kinases downstream of *EGFR*, which are independent of *MET* signaling in cells which contain normal levels of *MET* (53,54). Thus, promiscuous molecular interactions may significantly alter the cellular interactome when gene copy number is changed.

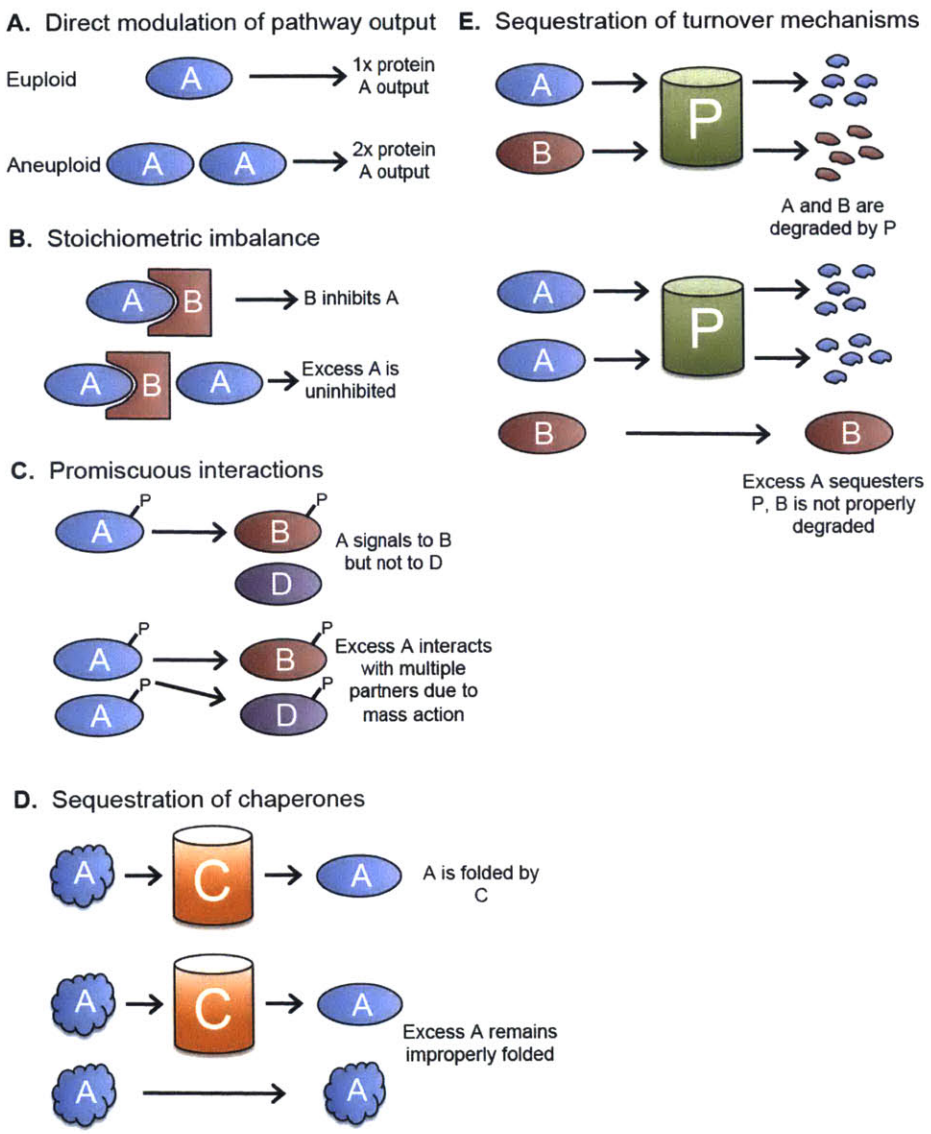


Figure 2. Consequences of changes in gene copy number. (A) An increased dosage of a single gene, such as a rate-limiting enzyme, can increase the output or function of a cellular pathway. (B) Altered gene dosage can interfere with the function of stoichiometry-sensitive complexes. (C) Protein-protein interactions depend on the concentration of each binding partner. Altered expression of some proteins, such as signal-transduction kinases, might cause promiscuous molecular interactions that alter cellular phenotypes. (D) Many proteins require chaperones to fold correctly. If aneuploidy overwhelms cellular chaperones, then misfolded proteins that escape chaperone-dependent folding might form insoluble and potentially cytotoxic

aggregates. It is also possible that other essential clients of these chaperones remain unfolded. (E) Quality-control mechanisms, including the ubiquitin-proteasome pathway, ensure that misfolded or improperly expressed proteins are rapidly turned over. Regulated protein degradation is also utilized to trigger various cellular programs, including cell cycle progression. The overabundance of certain proteins might interfere with the folding or turnover of other client proteins.

The detrimental phenotypes associated with aneuploidy can also result from the synergistic effects of changing the copy number of several hundred genes at once. Aneuploid strains of yeast have constitutively high levels of protein aggregates and are sensitive to drugs that interfere with protein translation, folding, and degradation (2,39,55). Similarly, trisomic MEFs and aneuploid cancer cells are readily killed by the protein-folding inhibitor 17-AAG (56). These sensitivities are largely independent of the identity of the additional chromosome, suggesting that aneuploidy generally challenges a cell's ability to maintain protein homeostasis. While the consequences of over-expressing any one protein may be unique, cells utilize a limited number of quality-control mechanisms for protein folding and turnover. For instance, most proteins which contain a WD40 domain require the eukaryotic chaperonin TRiC/CCT to correctly fold (57). In the absence of sufficient chaperone capacity to accommodate over-expressed proteins, other chaperone clients may remain unfolded, leading to loss-of-function phenotypes and the formation of potentially cytotoxic aggregates (Fig. 2D). In yeast, misfolded proteins exact a fitness cost, even when the misfolded proteins represent less than 0.1% of the total cellular proteome (58). Thus, aneuploidy may impair cell proliferation via an accumulation of improperly folded or aggregated proteins.

Misfolded proteins, as well as properly-folded proteins which are present in excess, can also impinge on cellular mechanisms for protein turnover. It has been demonstrated that cells ensure the integrity of some stoichiometric complexes by rapidly degrading over-expressed free subunits, such as histones (59) and ribosomal proteins (60,61), and we have observed that not all proteins in aneuploid cells are proportionally over-expressed (39,40). Thus, aneuploidy-induced stoichiometric imbalances may severely stress the proteasome (Fig. 2E). In order to understand how cells cope with the anti-proliferative effects of aneuploidy, our lab allowed disomic yeast strains to grow continuously for 14 days, then used whole-genome sequencing to identify genetic alterations which improved their proliferative capacity (39). Interestingly, many different disomic strains developed mutations in the proteasome pathway, and two strains

independently acquired loss-of-function mutations in the ubiquitin-specific protease *UBP6*. Deletion of *UBP6* was found to improve the growth rates of most disomic strains, and quantitative mass spectrometry demonstrated that loss of *UBP6* led to an attenuation in the levels of proteins overproduced due to aneuploidy (39,40). These results suggest that proteotoxic stress is a key source of aneuploidy's anti-proliferative effects.

Lastly, aneuploidy has been found to significantly alter metabolism and increase cellular energy needs. Aneuploid yeast and MEFs are less efficient at converting nutrients into biomass than euploid cells are (2,3). This could result from cells wasting energy by translating and then turning over excess proteins from the additional chromosomes. Consequently, aneuploid fibroblasts and cancer cells display broad sensitivities to drugs that interfere with cellular energy homeostasis (56). We and others have also noted that aneuploid cells produce significantly more lactate during proliferation than euploid cells do (3,27), a phenotype which they share with cancer cells (62). The underlying cause of these metabolic alterations is not known, but the resultant energy stress is a likely limit on the proliferation of aneuploid cells. Additionally, the metabolic similarities between aneuploid primary cells and cancer cells suggest that cancer cells have not fully escaped the stresses associated with aneuploidy.

What proportion of aneuploidy-induced phenotypes are due to changes in the copy number of specific genes and what proportion are a consequence of the additive effects of imbalances across an entire chromosome? Bonney et al. (2015) expressed ~50 genes on CEN plasmids that had previously been found to inhibit growth when fewer than 20 extra copies of the gene were present (63,64). The phenotypes of the CEN plasmid overexpression strains were then compared to the phenotypes of aneuploid yeast strains. While all aneuploid yeast strains divided slowly, very few of these dosage-sensitive genes were found to reduce proliferation when present at only 2-fold increased copy number. Similarly, in humans, it has previously been hypothesized that the phenotypes associated with Down Syndrome (Trisomy 21) are caused by the triplication of a small set of genes which lie within a 5.4Mb "Down

Syndrome Critical Region" on chromosome 21 (65,66). However, evidence from mouse models suggest that this region is not sufficient to recapitulate Down syndrome-like phenotypes, and genetic mapping of individuals with partial trisomies of chromosome 21 has revealed many individuals who lack extra copies of the purported critical region but still present with a Down syndrome-like condition (67–69). In total, this evidence suggests that many phenotypes caused by aneuploidy are driven by imbalances across an entire chromosome, or synergistic interactions between many different genes, rather than extra copies of a few particularly toxic genes.

Aneuploidy is frequently associated with enhanced cell proliferation

Despite the seemingly detrimental effects of the massive gene dosage imbalances which result from aneuploidy, changes in karyotype are associated with proliferative advantages under certain circumstances. Most notably, aneuploidy is a hallmark of cancer, a disease of increased cell proliferation (70). Approximately 90% of solid tumors and 75% of hematopoietic cancers have gained or lost entire chromosomes (71). Whether aneuploidy has a causal role in tumorigenesis or is largely a byproduct of transformation remains a crucial and unsolved question (see Chapter 5). However, several lines of evidence suggest that aneuploidy may be a driver of tumorigenesis. First, aneuploidy appears prior to or coincident with malignant transformation in a variety contexts, including in human patients (72–74) and mouse models of cancer (75,76). Secondly, individuals with Down syndrome develop pediatric leukemias at high rates, suggesting a link between trisomy of chromosome 21 and leukemogenesis (42). Thirdly, individuals with Mosaic Variegated Aneuploidy are prone to developing several different types of cancers (22). Similarly, transgenic mice with elevated rates of chromosome missegregation usually develop cancers at elevated rates. These studies suggest that changes in karyotype

can contribute to, if not also cause, cellular transformation, a process that enhances the proliferative capacity of a cell.

Additionally, certain aneuploidies confer a proliferative advantage in the experimental evolution of microorganisms. *S. cerevisiae* cells grown in nutrient-limiting conditions develop aneuploidies (77,78), as do pathogenic strains of *C. ablicans* grown in the presence of the anti-fungal drug fluconazole (79,80). In the yeast deletion collection, ~8% of all strains have spontaneously gained extra chromosomes (81). Lastly, in a comprehensive analysis of the genetic alterations that result from the evolution of a yeast strain deprived of a cytokinetic motor, Rancati et al. (2008) found that the 10 fittest evolved strains all displayed whole-chromosome aneuploidies (38). Thus, it is clear that alterations in karyotype can improve cellular fitness under certain circumstances.

How is aneuploidy beneficial?

As with the detrimental consequences of aneuploidy, most of the growth-advantageous properties of aneuploidy likely result from changes in gene copy number. However, while the disadvantages of aneuploidy result from additive imbalances across an entire chromosome, many of the benefits of aneuploidy can apparently be explained by the change in copy number of one or a few genes. In studies of nutrient-limited yeast, cells acquire aneuploidies that confer extra copies of the transporters that take up the scarce nutrient (77,78). Fluconazole resistance in *C. albicans* can be caused by the gain of an isochromosome containing *TAC1*, a transcriptional regulator of drug-efflux pumps, and *ERG11*, the target of fluconazole (80). Decreasing the copy number of *TAC1* and *ERG11* abolishes the protective effects of aneuploidy. Among the aneuploid strains which arose during construction of the yeast deletion collection, many of the strains were found to harbor extra copies of a chromosome containing a paralogue of the gene which was deleted, suggesting a dosage-dependent rescue of cell

proliferation (81). And, in the evolution of cytokinesis-defective yeast, improved growth due to aneuploidy can be phenocopied in euploid strains by increasing the dosage of a transcription factor and a signaling kinase that are present on a frequently-gained chromosome (38).

Changes in copy number of a few genes may also explain the tumorigenic properties of aneuploidy. Karyotypic alterations can lead to the gain in copy number of growth-promoting oncogenes and loss of tumor-suppressor loci. For instance, cervical cancers frequently gain extra copies of chromosome 3, which contains the human telomerase gene *TERC* (73,82), while Wilms' tumor, Ewing's sarcoma, and several types of leukemias frequently gain chromosome 8, which contains the *MYC* transcription factor (83–86). It is believed that these genes are largely responsible for the recurrent gain of these chromosomes because in many cancers these loci are focally amplified (87–90). Chromosome loss can also promote cellular transformation by decreasing the dosage of tumor suppressors (91,92). Interestingly, loss of heterozygosity at tumor suppressors is frequently "copy-neutral," i.e. a duplication event occurs to maintain euploid dosage levels of all genes on the affected chromosome (91,93,94). This may arise due to selective pressure to protect cancer cells from the deleterious effects of haploinsufficiency, or from imbalances in other genes required for proliferation. A similar result has been described in yeast: Avihu et al. (2012) grew cells in the presence of thermal stress for several thousand generations (95). The cells rapidly acquired extra copies of chromosome III, which were found to enhance the growth of yeast at high temperatures. However, upon prolonged evolution under thermal stress, the yeast lost the trisomy, and instead upregulated several genes from chromosome III which mimicked the beneficial effects of the trisomy. Thus, these experiments suggest that the gain or loss of whole-chromosomes can function as a blunt "quick-fix" mechanism for a cell to change the copy number of a few select genes. Over time, a cell can evolve more refined methods to accomplish the same ends while avoiding unwanted aneuploidy-induced stresses.

Summary

Aneuploidy represents a gross challenge to cellular homeostasis: by simultaneously changing the copy number of several hundred genes at once, aneuploidy can vastly alter normal cellular functions. The consequences of this dysregulation are apparent in the defects caused by aneuploidy during cell proliferation and development. Yet, it is for this same reason that certain genetic and environmental stresses can be overcome via aneuploidy. When one chromosome contains a gene that has a dosage-dependent effect on cell growth (such as an oncogene or a transporter of a scarce nutrient), chromosome missegregation represents a relatively easy way by which the copy number of that gene can be increased. This aneuploidy exerts a high cost on the cell, particularly by increasing the burden on energy and protein homeostasis. However, cells can evolve to either lose the aneuploidy or to enhance proteostasis, which can shield them from the detrimental consequences of chromosomal imbalances. Nonetheless, the pathways which are commonly stressed in aneuploid cells represent attractive targets for the development of chemotherapeutics that could potentially be useful in treating a broad spectrum of aneuploid cancers.

Thesis Outline

This thesis will discuss several disparate projects concerning aneuploid cell biology. In Chapter 2, I will present evidence concerning one potential mechanism by which whole-chromosome aneuploidy itself could indirectly contribute to improved growth: it has previously been hypothesized that the protein imbalances caused by aneuploidy could interfere with a cell's basal mechanisms for ensuring genomic integrity (41,96,97). While this hypothesis has never been directly tested, it has been observed that the chromosomal instability of chemically-transformed Chinese hamster embryo cells varies with their degree of aneuploidy, suggesting that gross deviations from the euploid state decrease a cell's ability to ensure accurate

chromosome segregation (98). Aneuploid cells lacking p53 are also karyotypically unstable (99). However, whether this genomic instability is due to the aneuploid state or the genetic alterations necessary to grow aneuploid cells (i.e. the inactivation of p53 or the mutations that caused transformation of the hamster cells) is not clear. In Chapter 2, I tested several measures of genomic instability in genetically-identical euploid and aneuploid strains of *S. cerevisiae*. I demonstrated that single chromosome gains increase the rates of forward mutation, mitotic recombination, chromosome loss, and double-strand break formation. I also established that these phenotypes are due to stoichiometric imbalances in proteins encoded on the extra chromosomes Genomic instability has been shown to confer a proliferative advantage during evolutionary competition and can cause or contribute to cellular transformation (100–104). It is therefore possible that aneuploidy-induced genomic instability could contribute to tumor initiation and evolution.

In Chapter 3, I address a debate concerning the transcriptional consequences of aneuploidy. While several studies have been published demonstrating that aneuploid chromosomes are generally transcribed proportional to their copy number (3,35,37), the effects of aneuploidy on gene expression *in trans*; that is, on euploid chromosomes in aneuploid strains, has been contested. Torres et al. (2007) reported that aneuploidy in yeast induces a stereotypic gene expression program called the environmental stress response (ESR), in which cells down-regulate ribosomal subunits and up-regulate chaperones and stress-response transcripts (2). However, Pavelka et al. (2010) did not observe this phenotype in aneuploid yeast strains generated via triploid meiosis (4). To address this discrepancy, and to further our knowledge of the transcriptional consequences of aneuploidy, I conducted a meta-analysis on published microarray data from aneuploid cells from diverse organisms as well as microarrays that had been performed in our own lab. I found that Pavelka et al.'s interpretation was likely a consequence of the fact that they only examined 5 aneuploid strains, most of which harbored small aneuploidies. Analysis of more aneuploid yeast strains generated via triploid meiosis, as

well as aneuploidies that arose "naturally" in yeast knockout mutants, revealed a clear correlation between increasing degrees of aneuploidy and induction of the ESR. Moreover, this phenotype was conserved beyond budding yeast: aneuploid strains of fission yeast and *A. thaliana*, as well as aneuploid primary mouse and human cells, shared a set of transcriptional changes that were related to stress and were largely independent of the identity of the aneuploid chromosomes. These results demonstrate that aneuploidy induces a stereotypical stress response program in cells from diverse organisms.

In Chapter 4, I extend my study of the effects of aneuploidy on gene expression from primary cells to cancer cells. I found that chromosomally-unstable cancer cell lines exhibit highly significant transcriptional similarities with aneuploid primary cells, while clonally-aneuploid cancer cells do not. Both CIN cancer cells and aneuploid primary cells down-regulated cell cycle and RNA processing transcripts while up-regulating transcripts annotated to the membrane and the extracellular region. I also demonstrated that, while the overall gene expression patterns of CIN cells are associated with slow proliferation *in vitro*, many of the transcripts upregulated by CIN are associated with metastases and death from cancer *in vivo*. These results argue that there are two types of aneuploidy in cancer: clonal aneuploidy and aneuploidy resulting from CIN, which have distinct effects on gene expression and disease phenotypes.

In Chapter 5, I utilize our system of generating genetically-matched euploid and trisomic MEFs, as well as the Storchova lab's set of genetically-matched euploid and trisomic colon cancer cells, to test whether aneuploidy can contribute to cellular transformation. Surprisingly, I found that most aneuploid cell lines divided slowly *in vitro*, formed few colonies in soft agar, and grew poorly as xenografts, relative to their matched euploid strains. I was unable to detect a single condition in which aneuploidy contributed to improved tumorigenic capabilities. These results suggest that, rather than promoting cancer development, aneuploidy can very often function as a tumor suppressor.

In Chapter 6, I summarize several key results from this body of work, and suggest methods to address the various open questions which remain.

In the Appendix, I discuss an independent project completed in collaboration with Joan Smith. We set out to explore the under-representation of women at the faculty level in biomedical research. While 52% of doctorates in the life sciences are awarded to women, only about 36% of new professors are women. To understand some of the causes of the "leaky pipeline" in biomedical research, we collected publicly-available data on the graduate students, postdocs, and faculty employed at 39 biology departments at academic institutions in the United States. We found that high-achieving male faculty members – those who receive funding from HHMI, have been elected to the National Academy of Sciences, or who have won a major career award – employ significantly fewer women than other male faculty. For instance, while 40% of all postdocs in our dataset are women, male Nobel Laureates employ on average only 25% female postdocs. In contrast, "elite" female faculty who have achieved similar career milestones employ just as many women as other female faculty members do. We further demonstrate that most new faculty members complete postdoctoral training periods in the labs of these high-achieving faculty members. Therefore, the paucity of women employed by these PI's likely limits the number of women who are most competitive during faculty job searches.

References

1. Hardy PA, Zacharias H. Reappraisal of the Hansemann-Boveri hypothesis on the origin of tumors. Cell Biol Int. 2005;29:983–92.

2. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, et al. Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. Science. 2007;317:916–24.

3. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, et al. Aneuploidy Affects Proliferation and Spontaneous Immortalization in Mammalian Cells. Science. 2008;322:703–9.

4. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, et al. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature. 2010;468:321–5.

5. Thompson SL, Compton DA. Examining the link between chromosomal instability and aneuploidy in human cells. J Cell Biol. 2008;180:665–72.

6. Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol Syst Biol. 2012;8.

7. Storchova Z, Kuffer C. The consequences of tetraploidy and aneuploidy. J Cell Sci. 2008;121:3859–66.

8. Lee HO, Davidson JM, Duronio RJ. Endoreplication: polyploidy with purpose. Genes Dev. 2009;23:2461–77.

9. Ganem NJ, Storchova Z, Pellman D. Tetraploidy, aneuploidy and cancer. Curr Opin Genet Dev. 2007;17:157–62.

10. Hassold T, Abruzzo M, Adkins K, Griffin D, Merrill M, Millie E, et al. Human aneuploidy: Incidence, origin, and etiology. Environ Mol Mutagen. 1996;28:167–75.

11. Brown S. Miscarriage and Its Associations. Semin Reprod Med. 2008;26:391–400.

12. Pai GS, Jr RCL, Borgaonkar DS. Handbook of Chromosomal Syndromes. 1st ed. Wiley-Liss; 2002.

13. Epstein CJ. Mouse monosomies and trisomies as experimental systems for studying mammalian aneuploidy. Trends Genet. 1985;1:129–34.

14. McClintock B. A Cytological and Genetical Study of Triploid Maize. Genetics. 1929;14:180–222.

15. Lindsley DL, Sandler L, Baker BS, Carpenter ATC, Denell RE, Hall JC, et al. SEGMENTAL ANEUPLOIDY AND THE GENETIC GROSS STRUCTURE OF THE DROSOPHILA GENOME. Genetics. 1972;71:157–84.

16. Hodgkin J. X chromosome dosage and gene expression in Caenorhabditis elegans: Two unusual dumpy genes. MGG Mol Gen Genet. 1983;192:452–8.

17. Hodgkin J, Horvitz HR, Brenner S. Nondisjunction Mutants of the Nematode Caenorhabditis Elegans. Genetics. 1979;91:67–94.

18. Sigurdson DC, Herman RK, Horton CA, Kari CK, Pratt SE. An X-autosome fusion chromosome of Caenorhabditis elegans. MGG Mol Gen Genet. 1986;202:212–8.

19. Kingsbury MA, Friedman B, McConnell MJ, Rehen SK, Yang AH, Kaushal D, et al. Aneuploid neurons are functionally active and integrated into brain circuitry. Proc Natl Acad Sci U S A. 2005;102:6143–7.

20. Duncan AW, Taylor MH, Hickey RD, Hanlon Newell AE, Lenzi ML, Olson SB, et al. The ploidy conveyor of mature hepatocytes as a source of genetic variation. Nature. 2010;467:707–10.

21. Knouse KA, Wu J, Whittaker CA, Amon A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. Proc Natl Acad Sci U S A. 2014;111:13409–14.

22. Hanks S, Coleman K, Reid S, Plaja A, Firth H, FitzPatrick D, et al. Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. Nat Genet. 2004;36:1159–61.

23. Snape K, Hanks S, Ruark E, Barros-Núñez P, Elliott A, Murray A, et al. Mutations in CEP57 cause mosaic variegated aneuploidy syndrome. Nat Genet. 2011;43:527–9.

24. Segal DJ, McCoy EE. Studies on Down's syndrome in tissue culture. I. Growth rates protein contents of fibroblast cultures. J Cell Physiol. 1974;83:85–90.

25. Kops GJPL, Foltz DR, Cleveland DW. Lethality to human cancer cells through massive chromosome loss by inhibition of the mitotic checkpoint. Proc Natl Acad Sci U S A. 2004;101:8699–704.

26. Babu JR, Jeganathan KB, Baker DJ, Wu X, Kang-Decker N, van Deursen JM. Rae1 is an essential mitotic checkpoint regulator that cooperates with Bub3 to prevent chromosome missegregation. J Cell Biol. 2003;160:341–53.

27. Li M, Fang X, Baker DJ, Guo L, Gao X, Wei Z, et al. The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. Proc Natl Acad Sci USA. 2010;107:14188–93.

28. Thompson SL, Compton DA. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. J Cell Biol. 2010;188:369–81.

29. Niwa O, Yanagida M. Triploid meiosis and aneuploidy in Schizosaccharomyces pombe: an unstable aneuploid disomic for chromosome III. Curr Genet. 1985;9:463–70.

30. Niwa O, Tange Y, Kurabayashi A. Growth arrest and chromosome instability in aneuploid yeast. Yeast. 2006;23:937–50.

31. Thorburn RR, Gonzalez C, Brar GA, Christen S, Carlile TM, Ingolia NT, et al. Aneuploid yeast strains exhibit defects in cell growth and passage through START. Mol Biol Cell. 2013;24:1274–89.

32. Futcher B, Carbon J. Toxic effects of excess cloned centromeres. Mol Cell Biol. 1986;6:2213–22.

33. Runge KW, Zakian VA. Introduction of extra telomeric DNA sequences into Saccharomyces cerevisiae results in telomere elongation. Mol Cell Biol. 1989;9:1488–97.

34. Jiang J, Jing Y, Cost GJ, Chiang J-C, Kolpa HJ, Cotton AM, et al. Translating dosage compensation to trisomy 21. Nature. 2013;500:296–300.

35. Mao R, Zielke CL, Ronald Zielke H, Pevsner J. Global up-regulation of chromosome 21 gene expression in the developing down syndrome brain. Genomics. 2003;81:457–67.

36. FitzPatrick DR, Ramsay J, McGill NI, Shade M, Carothers AD, Hastie ND. Transcriptome analysis of human autosomal trisomy. Hum Mol Genet. 2002;11:3249–56.

37. Huettel B, Kreil DP, Matzke M, Matzke AJM. Effects of Aneuploidy on Genome Structure, Expression, and Interphase Organization in Arabidopsis thaliana. PLoS Genet. 2008;4.

38. Rancati G, Pavelka N, Fleharty B, Noll A, Trimble R, Walton K, et al. Aneuploidy Underlies Rapid Adaptive Evolution of Yeast Cells Deprived of a Conserved Cytokinesis Motor. Cell. 2008;135:879–93.

39. Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, et al. Identification of aneuploidy-tolerating mutations. Cell. 2010;143:71–83.

40. Dephoure N, Hwang S, O'Sullivan C, Dodgson SE, Gygi SP, Amon A, et al. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. eLife. 2014;3.

41. Torres EM, Williams BR, Amon A. Aneuploidy: Cells Losing Their Balance. Genetics. 2008;179:737–46.

42. Hasle H, Clemmensen IH, Mikkelsen M. Risks of leukaemia and solid tumours in individuals with Down's syndrome. The Lancet. 2000;355:165–9.

43. Baek K-H, Zaslavsky A, Lynch RC, Britt C, Okada Y, Siarey RJ, et al. Down's syndrome suppression of tumour growth and the role of the calcineurin inhibitor DSCR1. Nature. 2009;459:1126–30.

44. Sussan TE, Yang A, Li F, Ostrowski MC, Reeves RH. Trisomy represses ApcMin-mediated tumours in mouse models of Down's syndrome. Nature. 2008;451:73–5.

45. Kaeberlein M, McVey M, Guarente L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev. 1999;13:2570–80.

46. Ibáñez P, Bonnet A-M, Débarges B, Lohmann E, Tison F, Pollak P, et al. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. Lancet. 2004;364:1169–71.

47. Devlin RH, Holm DG, Grigliatti TA. Autosomal dosage compensation Drosophila melanogaster strains trisomic for the left arm of chromosome 2. Proc Natl Acad Sci U S A. 1982;79:1200–4.

48. McAnally AA, Yampolsky LY. Widespread transcriptional autosomal dosage compensation in Drosophila correlates with gene expression level. Genome Biol Evol. 2010;2:44–52.

49. Papp B, Pal C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. Nature. 2003;424:194–7.

50. Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, et al. Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast. Genetics. 2005;169:1915–25.

51. Kaizu K, Moriya H, Kitano H. Fragilities Caused by Dosage Imbalance in Regulation of the Budding Yeast Cell Cycle. PLoS Genet. 2010;6:e1000919.

52. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. Cell. 2009;138:198–208.

53. Bean J, Brennan C, Shih J-Y, Riely G, Viale A, Wang L, et al. MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. Proc Natl Acad Sci. 2007;104:20932–7.

54. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, et al. MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. Science. 2007;316:1039–43.

55. Oromendia AB, Dodgson SE, Amon A. Aneuploidy causes proteotoxic stress in yeast. Genes Dev. 2012;26:2696–708.

56. Tang Y-C, Williams BR, Siegel JJ, Amon A. Identification of Aneuploidy-Selective Antiproliferation Compounds. Cell. 2011;144:499–512.

57. Siegers K, Bolter B, Schwarz JP, Bottcher UMK, Guha S, Hartl FU. TRiC/CCT cooperates with different upstream chaperones in the folding of distinct protein classes. EMBO J. 2003;22:5230–40.

58. Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc Natl Acad Sci USA. 2011;108:680–5.

59. Gunjan A, Verreault A. A Rad53 Kinase-Dependent Surveillance Mechanism that Regulates Histone Protein Levels in S. cerevisiae. Cell. 2003;115:537–49.

60. Abovich N, Gritz L, Tung L, Rosbash M. Effect of RP51 gene dosage alterations on ribosome synthesis in Saccharomyces cerevisiae. Mol Cell Biol. 1985;5:3429–35.

61. Warner JR, Mitra G, Schwindinger WF, Studeny M, Fried HM. Saccharomyces cerevisiae coordinates accumulation of yeast ribosomal proteins by modulating mRNA splicing, translational initiation, and protein turnover. Mol Cell Biol. 1985;5:1512–21.

62. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science. 2009;324:1029–33.

63. Bonney ME, Moriya H, Amon A. Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. Genes Dev. 2015;29:898–903.

64. Makanae K, Kintaka R, Makino T, Kitano H, Moriya H. Identification of dosage-sensitive genes in Saccharomyces cerevisiae using the genetic tug-of-war method. Genome Res. 2013;23:300–11.

65. Delabar JM, Theophile D, Rahmani Z, Chettouh Z, Blouin JL, Prieur M, et al. Molecular mapping of twenty-four features of Down syndrome on chromosome 21. Eur J Hum Genet EJHG. 1993;1:114–24.

66. Korenberg JR, Chen XN, Schipper R, Sun Z, Gonsky R, Gerwehr S, et al. Down syndrome phenotypes: the consequences of chromosomal imbalance. Proc Natl Acad Sci USA. 1994;91:4997–5001.

67. Olson LE, Richtsmeier JT, Leszl J, Reeves RH. A Chromosome 21 Critical Region Does Not Cause Specific Down Syndrome Phenotypes. Science. 2004;306:687–90.

68. Olson LE, Roper RJ, Sengstaken CL, Peterson EA, Aquino V, Galdzicki Z, et al. Trisomy for the Down syndrome "critical region" is necessary but not sufficient for brain phenotypes of trisomic mice. Hum Mol Genet. 2007;16:774–82.

69. Korbel JO, Tirosh-Wagner T, Urban AE, Chen X-N, Kasowski M, Dai L, et al. The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. Proc Natl Acad Sci U S A. 2009;106:12031–6.

70. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70.

71. Weaver BA, Cleveland DW. Does aneuploidy cause cancer? Curr Opin Cell Biol. 2006;18:658–67.

72. Balaban GB, Herlyn M, Clark Jr. WH, Nowell PC. Karyotypic evolution in human malignant melanoma. Cancer Genet Cytogenet. 1986;19:113–22.

73. Heselmeyer-Haddad K, Sommerfeld K, White NM, Chaudhri N, Morrison LE, Palanisamy N, et al. Genomic Amplification of the Human Telomerase Gene (TERC) in Pap Smears Predicts the Development of Cervical Cancer. Am J Pathol. 2005;166:1229–38.

74. Rubin CE, Haggitt RC, Burmer GC, Brentnall TA, Stevens AC, Levine DS, et al. DNA aneuploidy in colonic biopsies predicts future development of dysplasia in ulcerative colitis. Gastroenterology. 1992;103:1611–20.

75. Conti CJ, Aldaz CM, O'Connell J, Klein-Szanto AJP, Slaga TJ. Aneuploidy, an early event in mouse skin tumor development. Carcinogenesis. 1986;7:1845–8.

76. Danielsen HE, Brøgger A, Reith A. Specific gain of chromosome 19 in preneoplastic mouse liver cells after diethylnitrosamine treatment. Carcinogenesis. 1991;12:1777–80.

77. Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, et al. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 2002;99:16144–9.

78. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, et al. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 2008;4:e1000303.

79. Selmecki A, Forche A, Berman J. Aneuploidy and Isochromosome Formation in Drug-Resistant Candida albicans. Science. 2006;313:367–70.

80. Selmecki A, Gerami-Nejad M, Paulson C, Forche A, Berman J. An isochromosome confers drug resistance in vivo by amplification of two genes, ERG11 and TAC1. Mol Microbiol. 2008;68:624–41.

81. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, et al. Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet. 2000;25:333–7.

82. Cao Y, Bryan TM, Reddel RR. Increased copy number of the TERT and TERC telomerase subunit genes in cancer cells. Cancer Sci. 2008;99:1092–9.

83. Kibbelaar RE, van Kamp H, Dreef EJ, Wessels JW, Beverstock GC, Raap AK, et al. Detection of trisomy 8 in hematological disorders by in situ hybridization. Cytogenet Genome Res. 1991;56:132–6.

84. Oudat R, Khan Z, Glassman AB. Detection of Trisomy 8 in Philadelphia Chromosome-Positive CML Patients Using Conventional Cytogenetic and Interphase Fluorescence in situ Hybridization Techniques and its Relation to c-myc Involvement. Ann Clin Lab Sci. 2001;31:68–74.

85. Peres EM, Savasan S, Cushing B, Abella S, Mohamed AN. Chromosome analyses of 16 cases of Wilms tumor: different pattern in unfavorable histology. Cancer Genet Cytogenet. 2004;148:66–70.

86. Maurici D, Perez-Atayde A, Grier HE, Baldini N, Serra M, Fletcher JA. Frequency and Implications of Chromosome 8 and 12 Gains in Ewing Sarcoma. Cancer Genet Cytogenet. 1998;100:106–10.

87. Soder AI, Hoare SF, Muir S, Going JJ, Parkinson EK, Keith WN. Amplification, increased dosage and in situ expression of the telomerase RNA gene in human cancer. Oncogene. 1997;14:1013–21.

88. Yokoi S, Yasui K, Iizasa T, Imoto I, Fujisawa T, Inazawa J. TERC Identified as a Probable Target within the 3q26 Amplicon That Is Detected Frequently in Non-Small Cell Lung Cancers. Clin Cancer Res. 2003;9:4705–13.

89. Jennings BA, Mills KI. c-myc locus amplification and the acquisition of trisomy 8 in the evolution of chronic myeloid leukaemia. Leuk Res. 1998;22:899–903.

90. La Farina M, Bellavia M, Tagliavia M, Eterno V, Colomba P, Scibetta A, et al. Two distinct amplification events of the c-myc locus in a colorectal tumour. J Cell Physiol. 2008;217:34–9.

91. Baker DJ, Jin F, Jeganathan KB, van Deursen JM. Whole Chromosome Instability Caused by Bub1 Insufficiency Drives Tumorigenesis through Tumor Suppressor Gene Loss of Heterozygosity. Cancer Cell. 2009;16:475–86.

92. Baker DJ, van Deursen JM. Chromosome missegregation causes colon cancer by APC loss of heterozygosity. Cell Cycle Georget Tex. 2010;9:1711–6.

93. Beroukhim R, Lin M, Park Y, Hao K, Zhao X, Garraway LA, et al. Inferring Loss-of-Heterozygosity from Unpaired Tumors Using High-Density Oligonucleotide SNP Arrays. PLoS Comput Biol. 2006;2:e41.

94. Gondek LP, Tiu R, O'Keefe CL, Sekeres MA, Theil KS, Maciejewski JP. Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. Blood. 2008;111:1534–42.

95. Yona AH, Manor YS, Herbst RH, Romano GH, Mitchell A, Kupiec M, et al. Chromosomal duplication is a transient evolutionary solution to stress. Proc Natl Acad Sci U S A. 2012;109:21010–5.

96. Holliday R. Chromosome error propagation and cancer. Trends Genet TIG. 1989;5:42–5.

97. Matzke M. Does the intrinsic instability of aneuploid genomes have a causal role in cancer? Trends Genet. 2003;19:253–6.

98. Duesberg P, Rausch C, Rasnick D, Hehlmann R. Genetic instability of cancer cells is proportional to their degree of aneuploidy. PNAS. 1998;95:13692–7.

99. Thompson SL, Compton DA. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. J Cell Biol. 2010;188:369–81.

100. Cahill DP, Kinzler KW, Vogelstein B, Lengauer C. Genetic instability and darwinian selection in tumours. Trends Cell Biol. 1999;9:M57–60.

101. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability - an evolving hallmark of cancer. Nat Rev Mol Cell Biol. 2010;11:220–8.

102. Thompson DA, Desai MM, Murray AW. Ploidy Controls the Success of Mutators and Nature of Mutations during Budding Yeast Evolution. Curr Biol. 2006;16:1581–90.

103. Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of E. coli. Nature. 1997;387:703–5.

104. Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, et al. Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut. Science. 2001;291:2606–8.

Chapter 2: Aneuploidy Drives Genomic Instability

Reprinted from Science:

Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, et al. Aneuploidy drives genomic instability in yeast. Science 2011;333:1026–30.

The experiments in figures 1C, 3D, S1, S2, and S9 were performed by HB.
The experiments in figures 3E, 3F, and S11 were performed by YT, YH, and ON.
The experiments in figures S13 and S14 were performed by ET.
All other experiments were performed by JS.

Abstract

Aneuploidy decreases cellular fitness, yet it is also associated with cancer, a disease of enhanced proliferative capacity. To investigate one mechanism by which aneuploidy could contribute to tumorigenesis, we examined the effects of aneuploidy on genomic stability. We analyzed 13 budding yeast strains that carry extra copies of single chromosomes and found that all aneuploid strains exhibited one or more forms of genomic instability. Most strains displayed increased chromosome loss and mitotic recombination as well as defective DNA damage repair. Aneuploid fission yeast strains also exhibited defects in mitotic recombination. Aneuploidy-induced genomic instability could facilitate the development of genetic alterations that drive malignant growth in cancer.

Results

Aneuploid yeast exhibit elevated levels of chromosome missegregation

Whole-chromosome aneuploidy, or a karyotype that is not a multiple of the haploid complement, is found in greater than 90% of human tumors and may contribute to cancer development (1). It has been suggested that aneuploidy increases genomic instability, which could accelerate the acquisition of growth-promoting genetic alterations (1,2). However, while aneuploidy is a result of genomic instability, there is at present limited evidence as to whether genomic instability can be a consequence of aneuploidy itself. To test this possibility directly, we assayed chromosome segregation fidelity in 13 haploid strains of *Saccharomyces cerevisiae* that carry additional copies of single yeast chromosomes (3). These aneuploid strains (henceforth disomes) display impaired proliferation and sensitivity to conditions that interfere with protein homeostasis (3,4). We measured segregation fidelity of a yeast artificial chromosome (YAC) containing human DNA and found the rate of chromosome missegregation to be increased in 9 out of 13 disomic strains relative to a euploid control (Fig.1A). The increase ranged from 1.7-fold to 3.3-fold, comparable to the fold increase observed in strains lacking the kinetochore components Chl4 or Mcm21. Consistent with chromosome segregation defects, 8 out of 13 disomic strains displayed impaired proliferation on plates containing the microtubule poison benomyl, including a majority of the strains that had increased rates of YAC loss (Fig.1B).

Chromosome missegregation can result from defects in chromosome attachment to the mitotic spindle or from problems in DNA replication or repair. Defects in any of these processes delay mitosis by stabilizing the anaphase inhibitor Pds1/securin (5). 5 out of 5 disomes (Disomes V, VIII, XI, XV, and XVI) exhibited delayed degradation of Pds1 relative to wild-type after release from a pheromone-induced G1 arrest (Fig.1C, S1). Defective chromosome bi-

orientation delays anaphase through the mitotic checkpoint component Mad2 (5). Deletion of *MAD2* had no effect on Pds1 persistence in 4 disomes, but eliminated this persistence in disome V cells (Fig.S1). Disome V also delayed Pds1 degradation after release from a nocodazole-induced mitotic arrest, demonstrating that this strain exhibits a bi-orientation defect, whereas disome XVI recovered from nocodazole with wild-type kinetics (Fig.S2). Thus, Pds1 persistence results predominantly from Mad2-independent defects in genome replication and/or repair (see below).

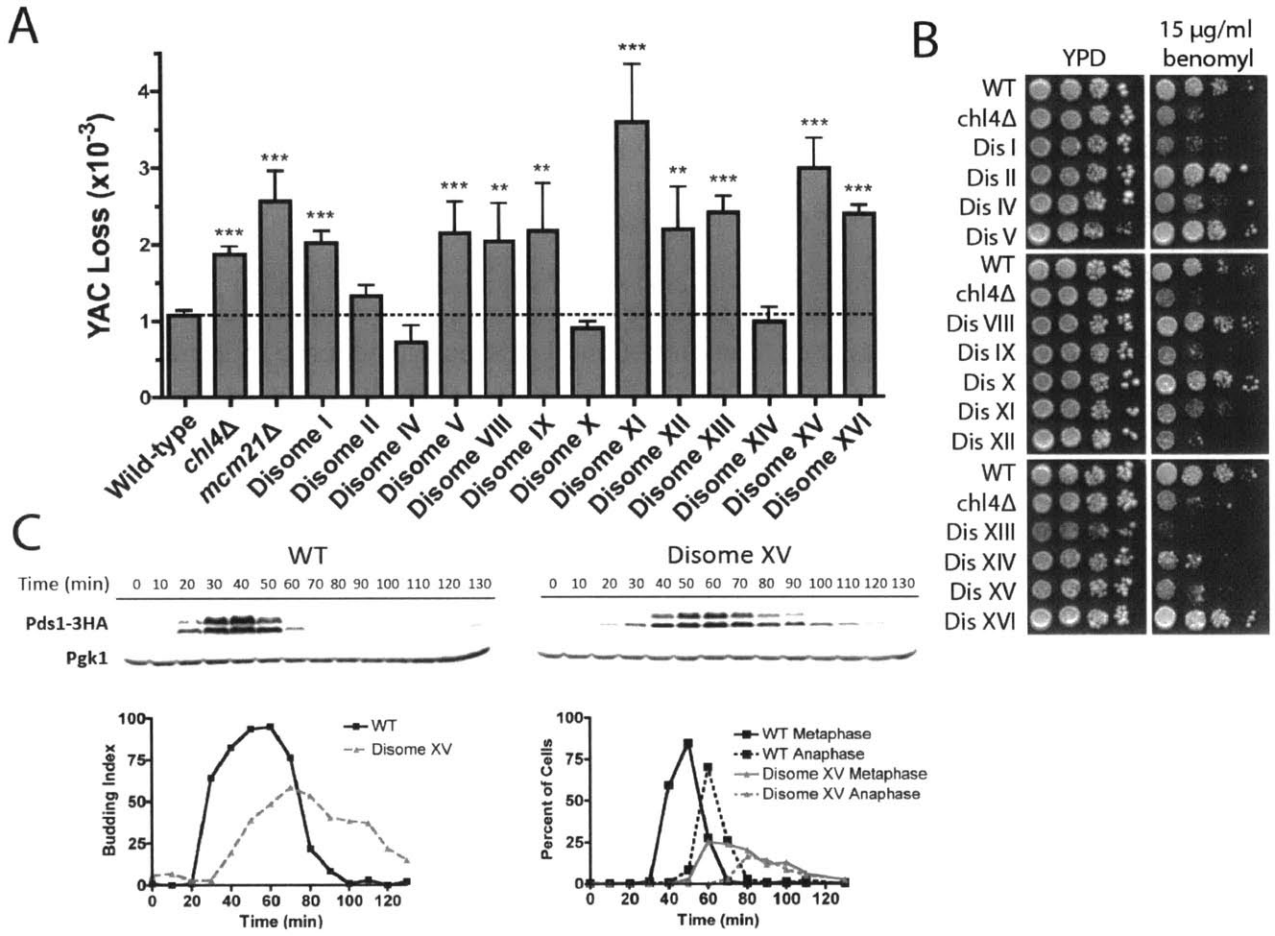


Fig.1. Aneuploidy induces chromosome missegregation. (A) YAC loss is increased in disomes and kinetochore mutants. The mean+/-SD of at least 12 cultures are displayed. ** p<.005; *** p<.0005 (Student's t-test). (B) Proliferation of disomes is decreased in the presence of benomyl. 10-fold serial dilutions of the indicated strains are shown. (C) Pds1 levels and cell cycle progression in wild-type and disome XV cells following release from a G1 arrest (6).

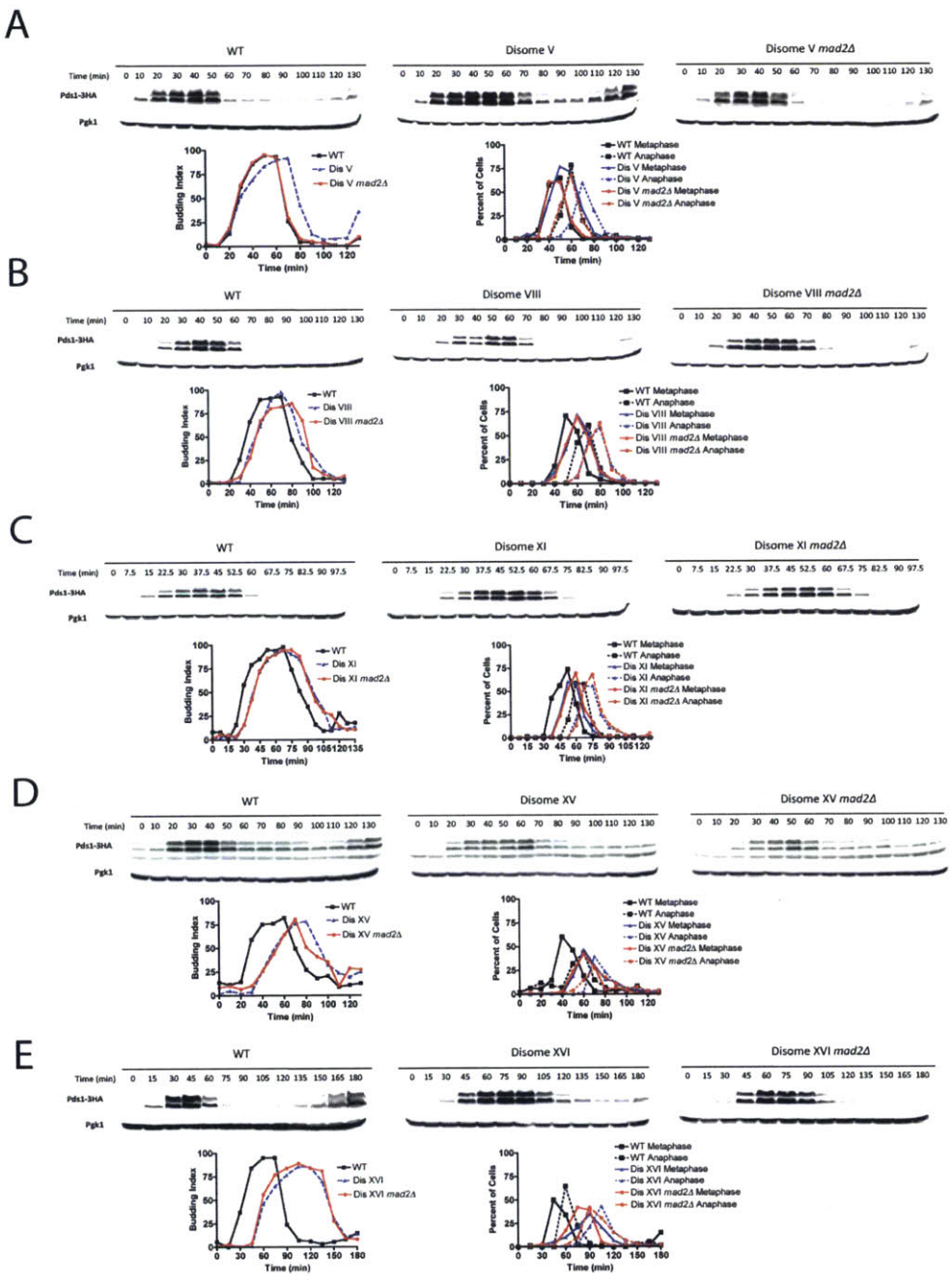


Figure S1. Pds1 accumulates in disomic strains. Wild-type (squares), disomic strains (triangles), and disomic strains lacking *MAD2* (circles) were arrested in G1 with pheromone then released into YPD medium. At the indicated time points, samples were removed and processed for Western blot analysis of Pds1, to determine the percentage of cells with buds (left graph), and metaphase and anaphase spindles (right graph). In five out of five disomes, Pds1 persisted

for a longer period of time than it did in a euploidstrain. In disome V cells this delay was due to activation of the spindle assembly checkpoint because it was eliminated by the deletion of *MAD2*. Strains are as follows: (A) A26703, A25413, and AA26708. (B) A26703, A25414, and A26709. (C) A26703, A25415, and A26704. (D) A26703, A26628, and A27925. (E) A26703, A25416, and A26705.

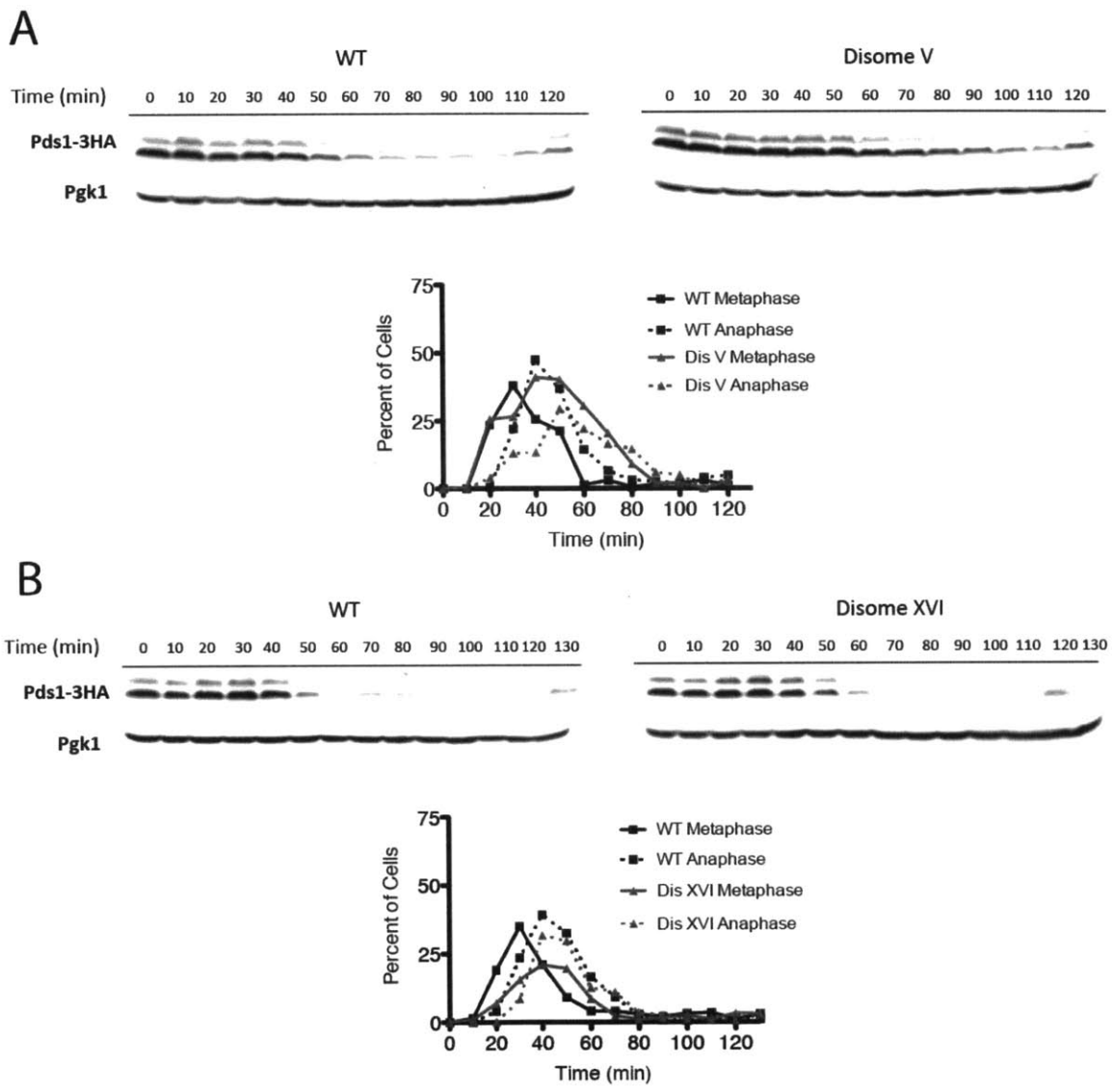


Figure S2. Disome V cells exhibit a chromosome bi-orientation defect. Wild-type and disomic strains were arrested at metaphase with 15 µg/ml nocodazole, then washed and released into YPD medium containing 5 µg/ml α-factor. At the indicated time points, samples were removed and processed for Western blot analysis of Pds1 and to determine the percentage of metaphase (squares) and anaphase (triangles) spindles. Strains are as follows: (A) A26703 and A25413. (B) A26703 and A25416.

Aneuploidy increases the mutation rate

We next investigated whether aneuploidy could affect the rate of forward mutation. Disomes V, VIII, X and XIV displayed an increased mutation rate at two independent loci, while disome IV displayed an increased mutation rate at *CAN1* but not at *URA3* (Fig.2A). The fold increase ranged from 2.2-fold to 7.1-fold, less than the 9.5-fold and 12-fold increases observed in a recombination deficient *rad51*Δ mutant and a mismatch repair-deficient *msh2*Δ mutant, respectively. Additionally, in an assay for microsatellite instability, we found that disomes VIII and XVI displayed increased instability in a poly(GT) tract (Fig.S3), demonstrating that aneuploidy can enhance both simple sequence instability and forward mutagenesis.

To define the mechanism underlying the increased mutation rate in aneuploid cells, we sequenced *CAN1* alleles from 133 wild-type and 404 disomic isolates (6). The overall spectrum of spontaneous mutation was similar, with euploid and aneuploid cells displaying equivalent frequencies of basepair substitutions, frameshifts, transitions, and transversions (Table S1). However, two significant differences were noted. First, the identity of basepairs gained and lost in the disomes differed relative to those seen in wild-type in a largely strand-specific manner (Table S2-S4, (6). Secondly, disomes exhibited a 2-fold increase in the frequency of complex events relative to wild-type ($p<0.002$, chi-square test; Fig.2A). Complex events, i.e. multiple substitutions and/or frameshifts within a 5 to 10 basepair window, are caused by the translesion polymerase Polζ (7). The frequency of complex events was increased when sequences from all mutator strains (Dis. IV, VIII, X, and XIV) were combined, but not when only non-mutator strains were examined. Deletion of *REV3*, which encodes the catalytic subunit of Polζ, abolished the increased mutation rate in the disomes (Fig.2A), demonstrating that aneuploidy-induced mutagenesis is due to translesion polymerase activity.

A

Strain	*CAN1* mutation rate x10^{-7} (95% CI)	Fold Change	*URA3* mutation rate x10^{-3} (95% CI)	Fold Change	Complex Events at *CAN1*
Wild-type	2.8 (2.3-3.4)	1	0.8 (0.5-1.9)	1	3.0% (4/133)
All Aneuploids	-	-	-	-	5.7% (23/404)††
Aneuploid Mutators	-	-	-	-	7.7% (12/156)††
msh2Δ	33.6 (28.7-45.5)	12.0***	-	-	-
rad51Δ	27.0 (21.8-48.0)	9.5***	-	-	-
Disome I	3.8 (2.2-4.1)	1.3	-	-	12.5% (3/24)
Disome II	2.4 (1.9-4.9)	0.9	-	-	8.3% (2/24)
Disome IV	8.2 (5.6-12.4)	2.9***	0.5 (0.2-1.2)	0.6	16.7% (4/24)
Disome VIII	6.6 (4.4-12.1)	2.3***	5.8 (2.8-7.0)	7.1**	6.3% (2/32)
Disome IX	4.4 (2.2-6.1)	1.6	-	-	2.5% (2/79)
Disome X	10.6 (6.0-12.7)	3.7***	2.4 (1.5-2.8)	2.9**	6.8% (5/73)
Disome XI	2.2 (1.2-2.6)	0.8	-	-	4.2% (1/24)
Disome XII	3.4 (1.5-4.9)	1.2	-	-	0% (0/25)
Disome XIII	2.6 (1.6-6.8)	0.9	-	-	0% (0/24)
Disome XIV	11.8 (8.3-18.6)	4.1***	2.4 (1.3-9.3)	3.0*	3.6% (1/28)
Disome XV	3.8 (1.2-11.4)	1.3	-	-	4.2% (1/24)
Disome XVI	2.1 (1.5-4.0)	0.7	-	-	8.3% (2/24)
Wild-type *rev3Δ*	1.4 (0.9-3.3)	0.5	-	-	-
Disome IV *rev3Δ*	1.0 (0.5-2.3)	0.3	-	-	-
Disome VIII *rev3Δ*	1.9 (1.0-3.0)	0.7	-	-	-
Disome X *rev3Δ*	0.9 (0.8-1.7)	0.3	-	-	-
Disome XIV *rev3Δ*	1.8 (0.9-2.6)	0.6	-	-	-

Strain	*LYP1* mutation rate x10^{-7} (95% CI)	Fold Change	*URA3* (Chr. VI) mutation rate x10^{-7} (95% CI)	Fold Change	
Wild-type	6.5 (4.5-10.2)	1	0.4 (0.2-0.9)	1	-
Disome V	14.1 (10.8-23.5)	2.2***	1.0 (0.6-1.2)	2.3**	-

* p< .05; ** p< .005; *** p< .0005 (Wilcoxon rank-sum test); †† p< .005 (chi-square test)
CI; Confidence Interval
Results presented are the median+/-95% CI of at least 12 cultures

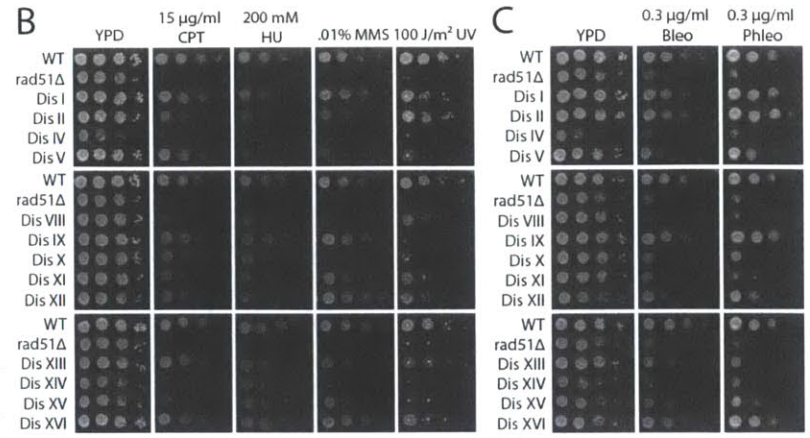


Fig.2. Aneuploidy increases the mutation rate and sensitivity to genotoxins. (A) Mutation rate in disomic strains. Note, the *CAN1* and *URA3* reporters are located on chromosome V, we therefore measured the mutation rate of disome V at *LYP1* and of *URA3* integrated on chromosome VI (6). (B) 10-fold serial dilutions of the indicated strains were spotted on medium supplemented or treated with a genotoxic agent. CPT; camptothecin. HU; hydroxyurea. MMS; methyl methanesulfonate. (C) 10-fold serial dilutions of cells on medium containing phleomycin (Phleo) or bleomycin (Bleo).

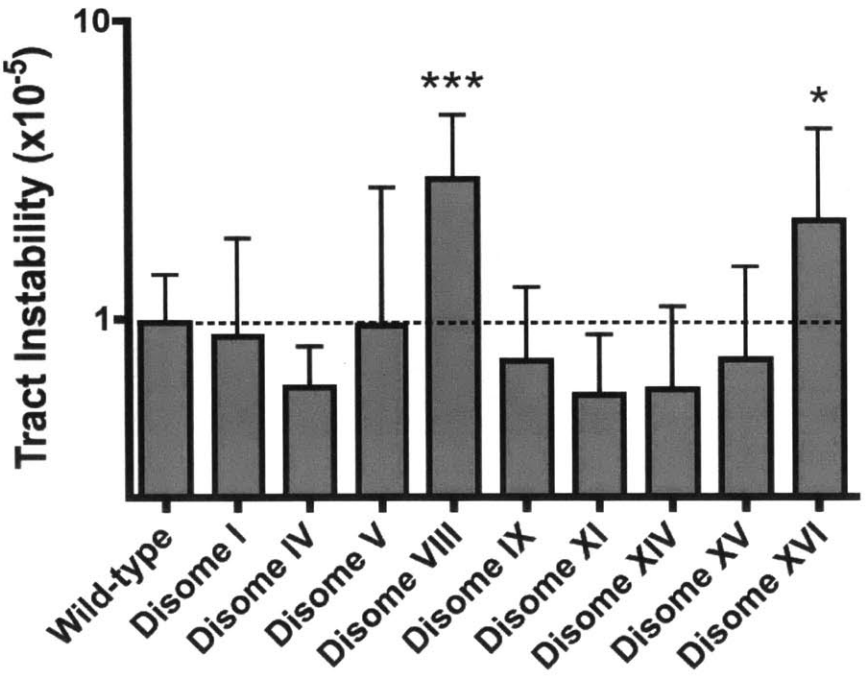


Figure S3. Disomes VIII and XVI cells display an increased rate of microsatellite instability. Disomes were assayed for maintenance of a poly(GT) tract fused in-frame with *URA3*. The median value and 95% confidence intervals of at least 12 independent cultures are shown. Strains (from the left): A24251, A24289, A26498, A24287, A24288, A26499, A26500, A24290, A24291, and A26501. * $p<.05$; *** $p<.0005$ (Wilcoxon rank-sum test).

Aneuploidy sensitizes cells to DNA damage

The mutator phenotype and frequent appearance of complex events suggested that aneuploidy interferes with the repair of genomic damage. To test this, we examined the sensitivity of the disomes to genotoxic stress (Fig.2B). A majority of disomes displayed impaired proliferation when treated with replication inhibitors (camptothecin, hydroxyurea) or DNA damaging agents (methyl methanesulfonate, UV light). Aneuploid strains derived by triploid meiosis also displayed striking sensitivities to genotoxic drugs (Fig.S4,(8). We next assessed the role of Polζ in lesion bypass. In wild-type yeast, loss of *REV3* confers only a slight increase in genotoxin sensitivity as recombinational repair is sufficient to replicate past most lesions (9). 7 out of 9 disomes displayed enhanced sensitivity to genotoxins in the absence of *REV3*, suggesting that recombinational repair is defective in the disomes (Fig.S5). We therefore assayed the sensitivity of the disomes to phleomycin and bleomycin, two double-strand break (DSB)-inducing drugs, which create lesions that are repaired by homologous recombination (10). 9 out of 13 disomes were sensitive to both drugs, and disomes IV, VIII, X, XI, and XIV displayed an approximately 100 to 1000-fold increase in sensitivity relative to wild-type (Fig.2B).

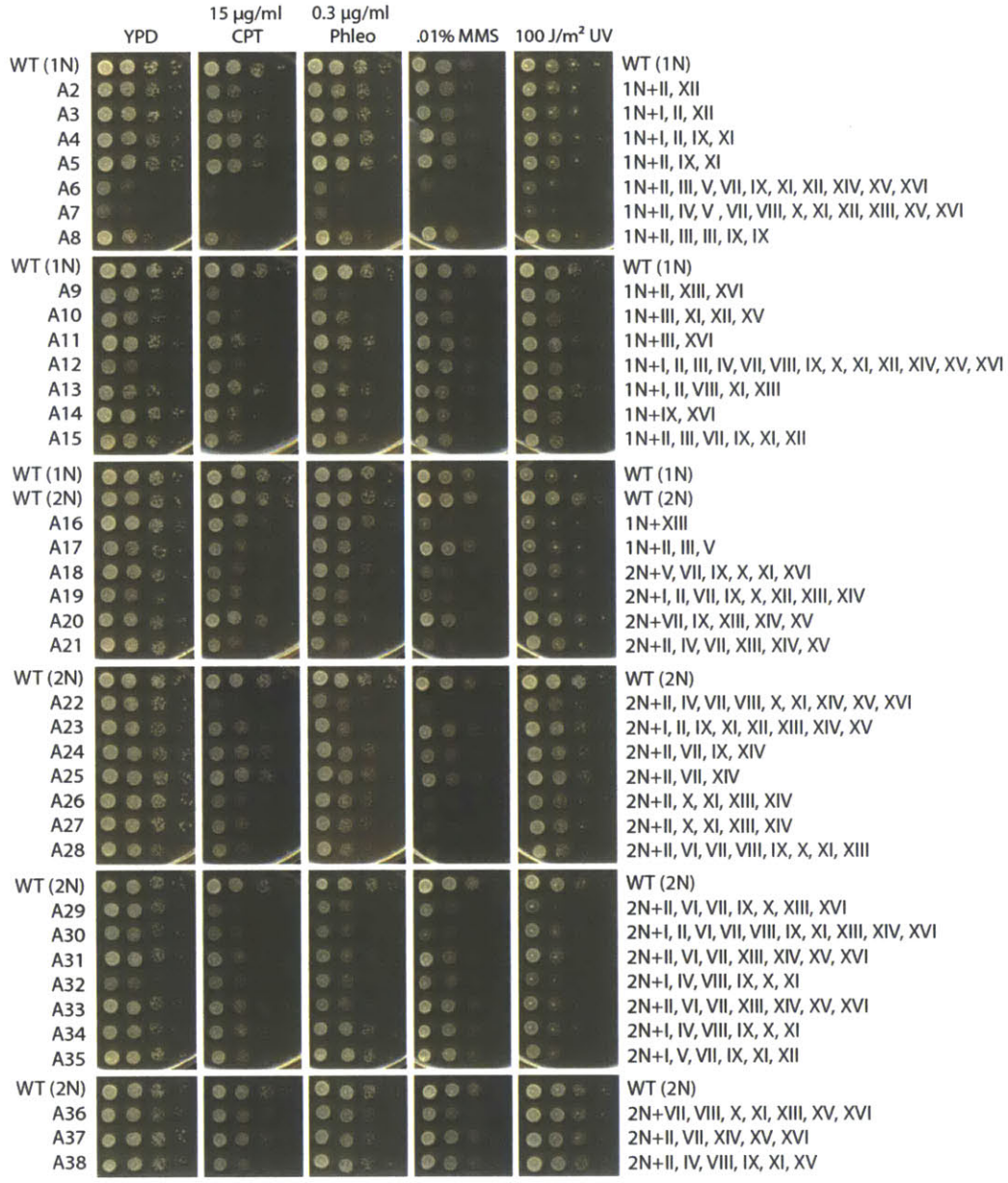


Figure S4. Aneuploid strains derived from triploid meioses are sensitive to genotoxic damage. 10-fold serial dilutions of the indicated strains were spotted on either YPD medium or medium supplemented or treated with the indicated genotoxic agent. The karyotype of these strains is indicated to the right of the figure. CPT; camptothecin. Phleo; phleomycin. MMS, methyl methanesulfonate. The construction and nomenclature of these strains are described in *(26)*. Note that in many strains the reported karyotype is present in only a subset of the cells, and some strains have euploid or near-euploid subpopulations (fig S13 and S14).

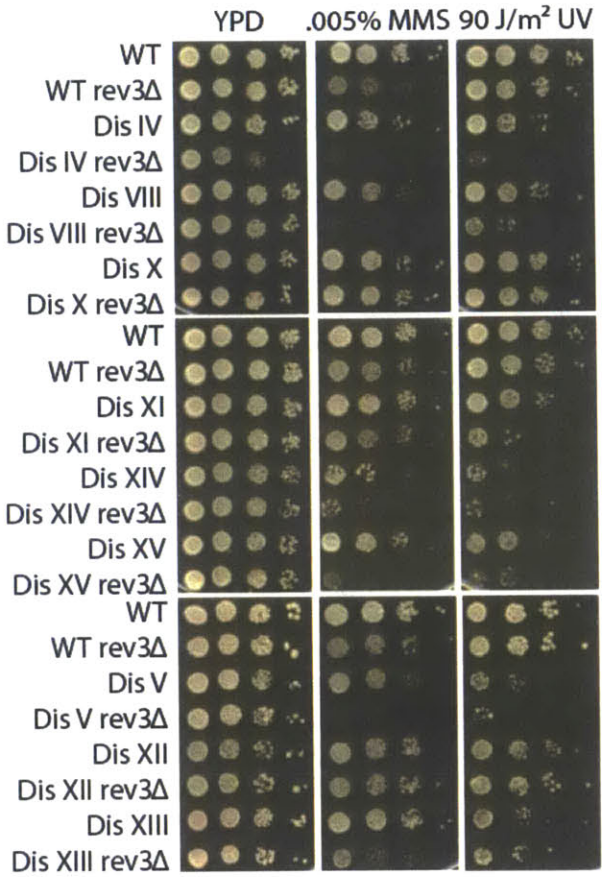


Figure S5. Several disomes require Polζ to survive DNA damage. 10-fold serial dilutions of the indicated *REV3* and *rev3*Δ strains were spotted on either YPD medium or medium supplemented or treated with the indicated genotoxic agent. Most disomes that lack *REV3* are hypersensitive to low doses of DNA damaging agents. MMS, methyl methanesulfonate. Strains (from the top): A11311, A25391, A12687, A25392, A13628, A25393, A12689, A25394, A13771, A25395, A13979, A25396, A12697, A25397, A14479, A26629, A12693, A26630, A21987, and A26631.

Aneuploidy interferes with double-strand break repair

To further investigate the effects of aneuploidy on recombination, we quantified the fraction of cells that contained double-strand breaks in 7 phleomycin-sensitive disomes by monitoring Rad52-GFP foci, which localize to sites of recombinational repair (11). Following release from a pheromone-induced G1 arrest, all 7 disomes displayed an increased frequency of Rad52 foci in large-budded cells (corresponding to late S-phase or G2). Disomes arrested with nocodazole also exhibited increased numbers of Rad52 foci (Fig.3A). The aneuploid meiotic progeny of a triploid strain displayed Rad52 foci more frequently than euploid spores did, demonstrating that the appearance of recombination foci is a common consequence of aneuploidy in yeast (Fig.3B). Consistent with aneuploidy-induced defects in DSB formation and/or repair, 7 out of 11 disomes also displayed an increased rate of spontaneous mitotic recombination between direct tandem repeats (Fig.3C).

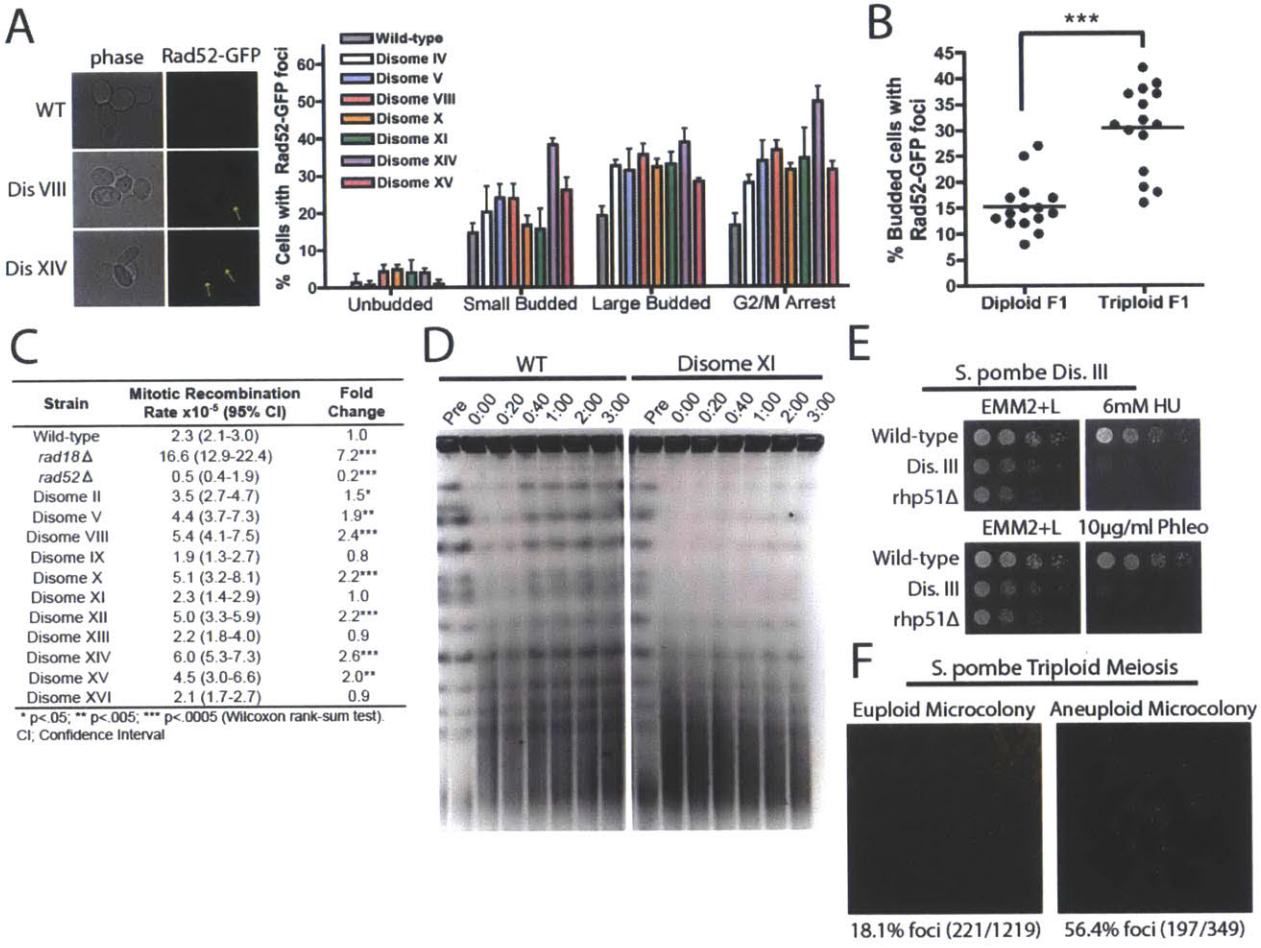


Fig.3. Aneuploidy induces recombination defects. (A) The fraction of wild-type and disomic cells displaying Rad52-GFP foci after release from a G1 arrest or arrested with nocodazole. Images display wild-type, disome VIII, and disome XIV cells arrested with nocodazole. Mean+/- SD of three experiments is shown. (B) Rad52 foci were scored in spores from triploid or diploid strains (6). The mean (black bar) of 15 spore-derived colonies (dots) are displayed. *** p<.0005 (Student's t-test). (C) Mitotic recombination between truncated alleles of *ade2* (6). (D) Wild-type and disome XI cells treated with phleomycin were released into medium containing nocodazole. Chromosome integrity was analyzed by pulse-field gel electrophoresis (6). (E) 10-fold serial dilutions of fission yeast cells on medium supplemented with hydroxyurea or phleomycin. Rhp51 is fission yeast Rad51. (F) The fraction of cells displaying SpRad22-GFP

foci in aneuploid and euploid microcolonies resulting from sporulation of a triploid strain. Images are representative euploid and aneuploid microcolonies.

To test whether disomes form more DSBs during DNA replication, we created *rad52Δ* strains, in which a single DSB is sufficient to block cell division (12). Small-budded *RAD52* and *rad52Δ* cells were isolated via micromanipulation and their proliferation monitored (6). 6% of *rad52Δ* cells arrested with large buds, while in 4 out of 6 *rad52Δ* disomes this percentage was significantly increased (Fig.S6). Thus, some aneuploid strains accumulate an increased number of DSBs during DNA replication, although the large-budded arrest in disome V may be due to defective chromosome bi-orientation, as frequent arrest was also observed in *RAD52* disome V cells (Fig.S6).

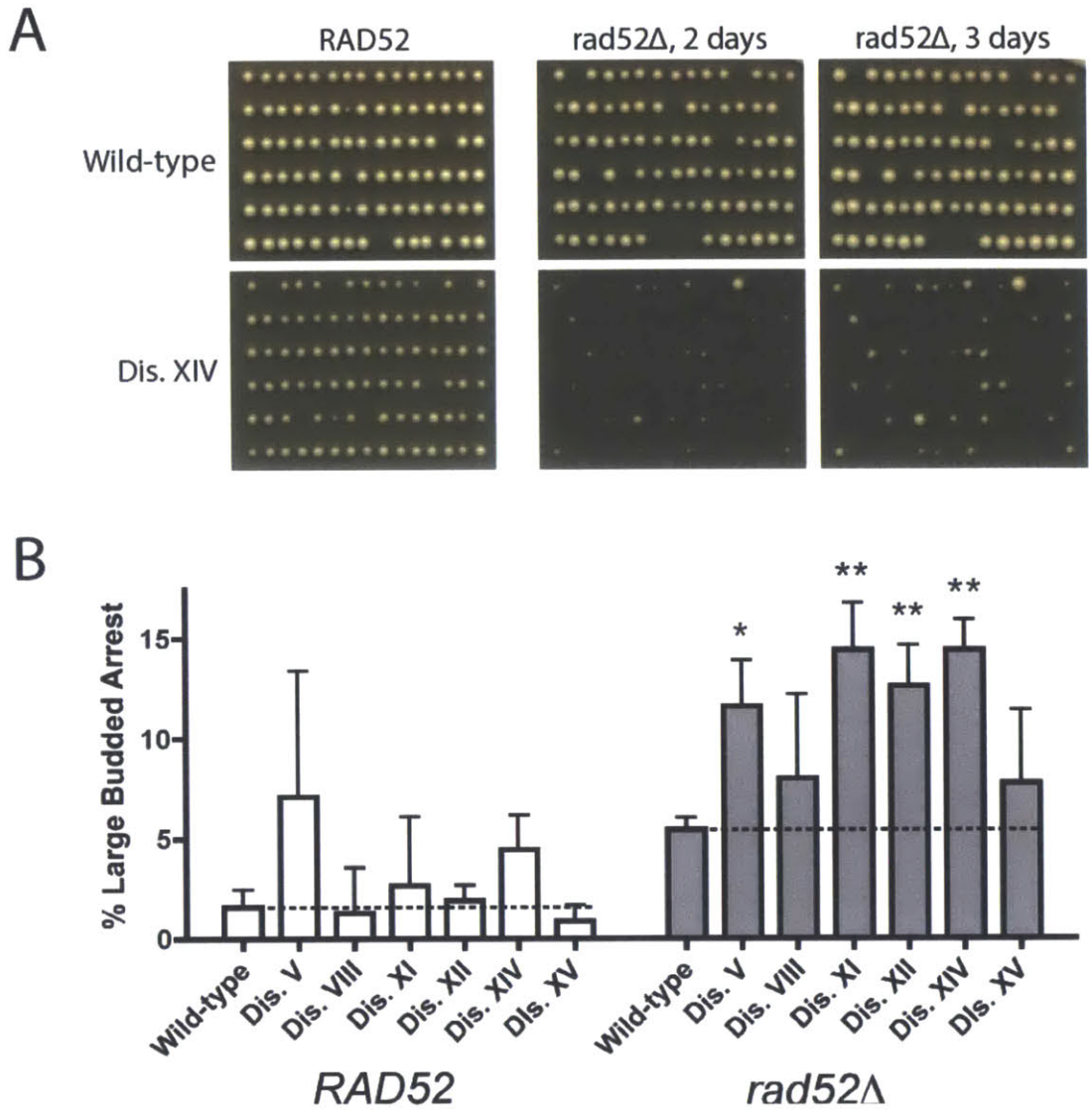


Figure S6. Some disomes acquire double-strand breaks more frequently during DNA replication. Small budded cells were isolated by micromanipulation on YPD plates and incubated for 5 hours at 30°C. Subsequently, the cells were examined under a microscope, and the number of cells that were arrested as large buds and the number of cells that had successfully divided were scored. Cells which had not progressed beyond the small-budded stage were excluded from analysis. (A) Representative plates of euploid *RAD52*, euploid *rad52*Δ, disome XIV *RAD52*, and disome XIV *rad52*Δ are displayed. Plates containing *RAD52* cells were scanned after 2 days of growth, while plates containing *rad52*Δ cells were scanned

after 2 or 3 days of growth. (B) The percentage of cells arrested as large buds is displayed. The mean and standard deviation of 3 independent experiments are shown. Strains (from the left): A15546, A27223, A15533, A15537, A15538, A15540, A15542, A26504, A28064, A19616, A27091, A19618, A19619, and A19620.

Is DNA repair also compromised in aneuploid cells? To test this, we examined Rad52 foci dynamics in disomes treated with phleomycin. In the presence of phleomycin, euploid and aneuploid strains arrested as large-budded cells and formed Rad52 foci. Following phleomycin removal, euploid cells resolved their Rad52 foci and resumed budding, while 7 out of 7 disomic strains remained arrested and displayed persistent Rad52 foci (Fig.S7). The sensitivity to phleomycin was not caused by DNA damage checkpoint defects, as exposure to phleomycin induced a prolonged cell-cycle arrest (Fig.S7) and caused hyperphosphorylation of Rad53, a marker of checkpoint activation (Fig.S8). Instead, disomes appear to be defective in DNA repair. When chromosomes were visualized by pulse-field gel electrophoresis, phleomycin treatment resulted in chromosome fragmentation in both aneuploid and euploid cells (Fig.3C, S9). After phleomycin removal, intact chromosomes quickly reappeared in a wild-type strain (Fig.3C, S9). In contrast, a significant delay in chromosome recovery was apparent in disomes V, VIII, XI and XIII (Fig.3C, S9). Disome II, which does not lose viability on plates containing phleomycin (Fig.2B), exhibited chromosome repair kinetics similar to wild-type cells (Fig.S9). Low doses of ionizing radiation (IR) had a similar though less severe effect on the disomes as phleomycin. Disomes lost viability upon treatment with IR, though several strains were able to resolve a subset of IR-induced Rad52 foci (Fig.S10). The different effects of phleomycin and IR may indicate that these treatments cause distinct forms of damage, or that disomic chromatin is particularly vulnerable to phleomycin-induced lesions. Taken together, our results indicate that multiple aneuploids exhibit wide-ranging defects in recombination and DNA repair.

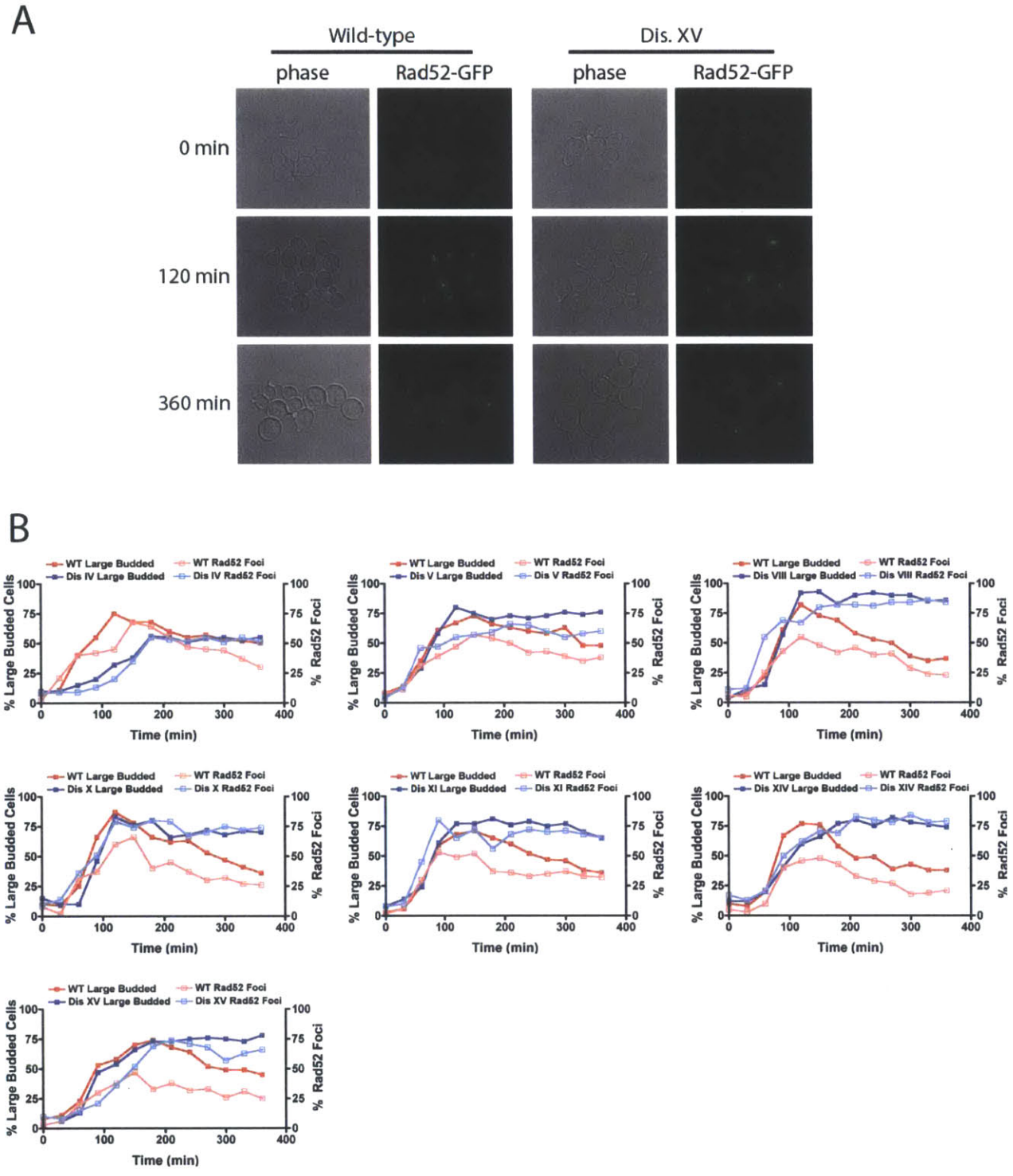


Figure S7. Disomes are unable to repair phleomycin-induced lesions. Strains were arrested in G1 with pheromone, then released into medium containing 0.15µg/ml phleomycin. After 135 minutes, the cells were washed and resuspended in medium lacking phleomycin. Cells were

removed every 30 minutes to determine the percentage of large-budded cells (closed symbols) and the percentage of cells with Rad52-GFP foci (open symbols). (A) Representative images of wild-type and Dis. XV cells after G1 arrest (top), after growth in the presence of phleomycin (middle), and after 4 hours of recovery in phleomycin-free medium (bottom). (B) Phleomycin arrest/release results for Dis. IV (A26532), V (A26533), VIII (A25342), X (A25343), XI (A25421), XIV (A25344), and XV (A25355). Note that Dis. IV cells display a severe G1 delay *(1)*, and many cells failed to bud during the course of the experiment. The Dis. IV cells that did bud were arrested as large-budded cells by phleomycin and almost all remained arrested for the duration of the time course.

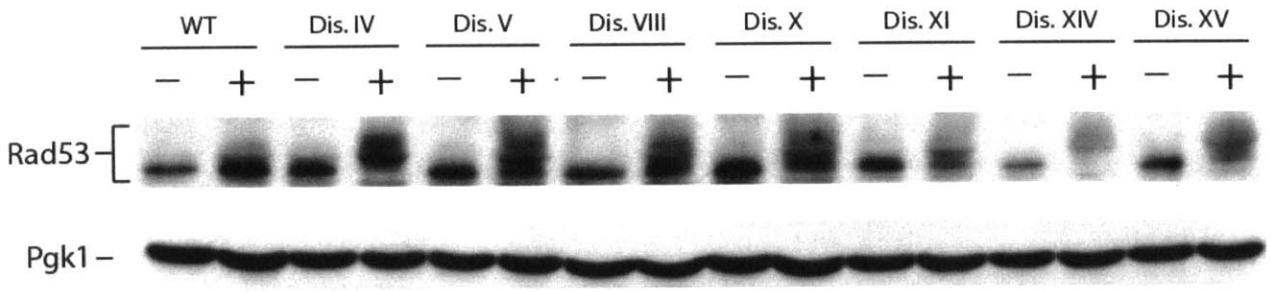


Figure S8. Disomic yeast strains phosphorylate Rad53 in response to phleomycin treatment. Rad53 mobility on SDS-PAGE was examined in untreated exponentially growing cells and in exponentially growing cells that were treated with 0.3µg/ml phleomycin for 90 minutes. The slower migrating forms of Rad53 represent phosphorylated Rad53. Pgk1 was used as a loading control. Strains (from the left): A11311, A12687, A14479, A13628, A21986, A13771, A13979, and A12697.

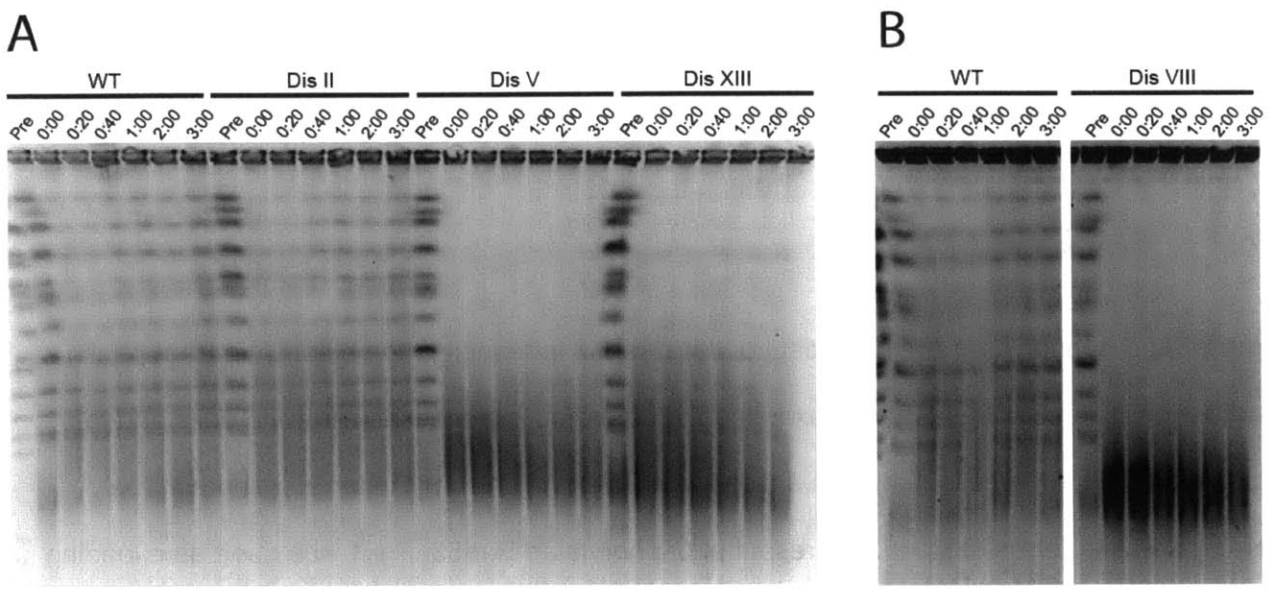


Figure S9. Disomic yeast strains display delayed chromosomal recovery following phleomycin treatment. Wild-type and disomic strains were treated with 25µg/ml phleomycin for 2 hours, then washed with 5 volumes of medium and released into YPD medium containing 15µg/ml nocodazole. Nocodazole was used to keep the number of cells in culture constant during recovery. Before treatment ("pre"), immediately following phleomycin wash-out ("0:00"), and at the indicated times after phleomycin wash-out cells were fixed with sodium azide and chromosomes were examined by pulse-field gel electrophoresis. Intact chromosomes are visible as discrete bands, while chromosome fragments run as a smear at the bottom of the gel or do not enter the gel. Strains (from the left): (A) A11311, A12685, A14479, A21987. (B) A11311 and A13628.

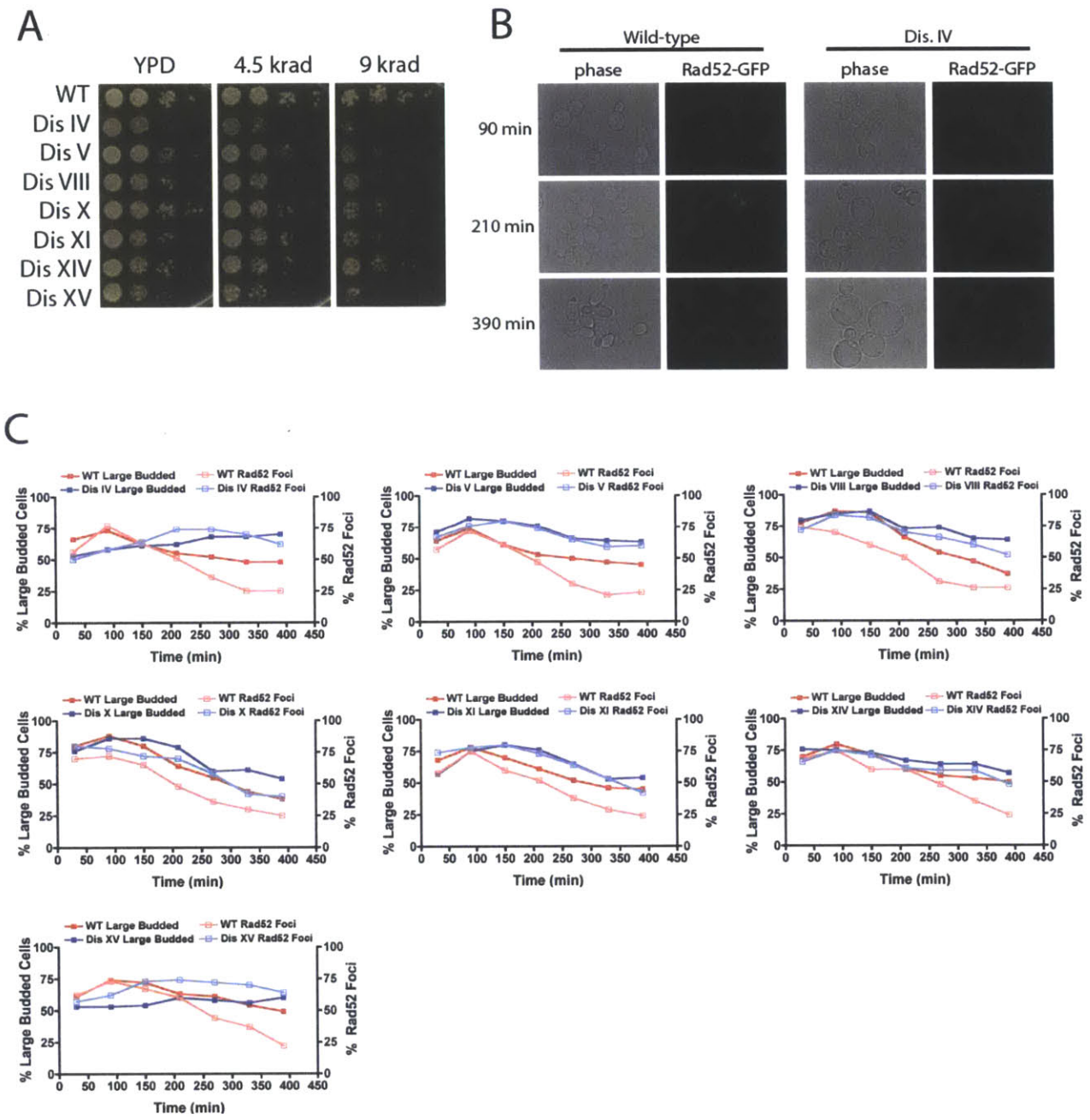
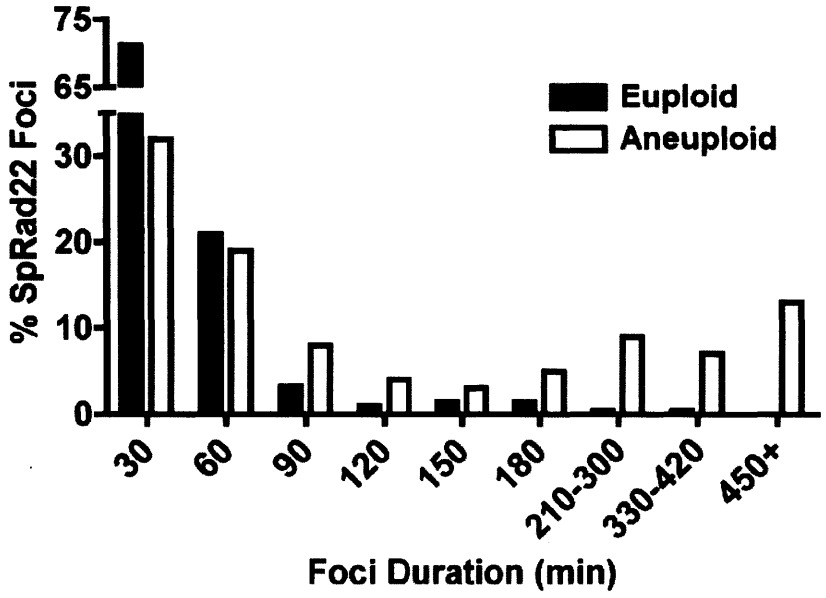


Figure S10. Disomic yeast strains are impaired in the repair of damage caused by ionizing radiation. (A) 10-fold serial dilutions of the indicated strains were spotted on YPD medium and then treated with the indicated dose of radiation. Strains (from the top): A24352, A26532, A26533, A25342, 25343, A25421, A25344, and A25355. (B) Exponentially-growing cultures were treated with 4.5krads of radiation and then allowed to recover. Cells were

removed every 60 minutes to determine the percentage of large-budded cells (closed symbols) and the percentage of cells with Rad52-GFP foci (open symbols). Representative images of wild-type and Dis. IV cells are shown 90 minutes after treatment with radiation (top), after 210 minutes of recovery (middle), and after 390 minutes of recovery (bottom). (B) Radiation recovery results for Dis. IV (A26532), V (A26533), VIII (A25342), X (25343), XI (A25421), XIV (A25344), and XV (A25355).

We also investigated the effects of aneuploidy on genomic stability in fission yeast. Fission yeast disome III, the only viable disome (13,14), displayed sensitivity to hydroxyurea and phleomycin relative to a euploid strain (Fig.3D). Additionally, Rad22 foci (fission yeast Rad52; (15,16) were present in 18% of euploid cells and in 56% of aneuploid cells resulting from sporulation of a triploid strain (Fig.3E). Time-lapse photomicroscopy revealed that approximately equal numbers of euploid and aneuploid cells formed SpRad22 foci per cell division (Fig.S11). However, Rad22 foci persisted on average 5 times longer in aneuploid than in euploid cells. We conclude that in fission yeast, aneuploidy impairs DNA damage resistance and mitotic recombination.



Microcolony Type	Mean Focus Duration in Minutes (+/- SD)	Number of Foci	% Cells which Divide without Foci Formation	Number of Cells
Euploid	45.6 (+/- 39.6)	215	53.8%	470
Aneuploid	219.8 (+/- 367.0)***	100	47.5%	141

*** p<.0005 (Wilcoxon rank-sum test)

Figure S11. Persistence of SpRad22-GFP foci duration in aneuploid strains of S. pombe. The time that SpRad22-GFP foci persisted in cells was measured in euploid and aneuploid strains generated from a triploid meiosis. Aneuploid cells were identified by their delayed germination and aberrant colony morphology. Cells were imaged every 30 minutes.

Stoichiometric imbalances in yeast proteins induce genomic instability

We next determined whether the genomic instability present in the disomic strains was caused by the presence of extra DNA or by aneuploidy-induced imbalances in protein stoichiometry. Yeast strains carrying YACs harboring human DNA were not sensitive to genotoxic agents and did not display increased mutation or Rad52-GFP foci, demonstrating that replication of an extra chromosome is not sufficient to induce genomic instability (Fig.4A-4C). If the defects in damage repair were caused by stoichiometric imbalances in yeast proteins, then the effects should be mitigated in diploids carrying single extra chromosomes (henceforth trisomes; (3). Indeed, 5 out of 5 trisomes were more resistant to genotoxic damage than their isogenic disomes, and in 3 out of 3 trisomes the fold increase in YAC loss relative to a diploid strain was less than the fold increase observed in isogenic disomes (Fig.4D-4F). Thus, excess protein but not excess DNA causes genomic instability in aneuploid cells.

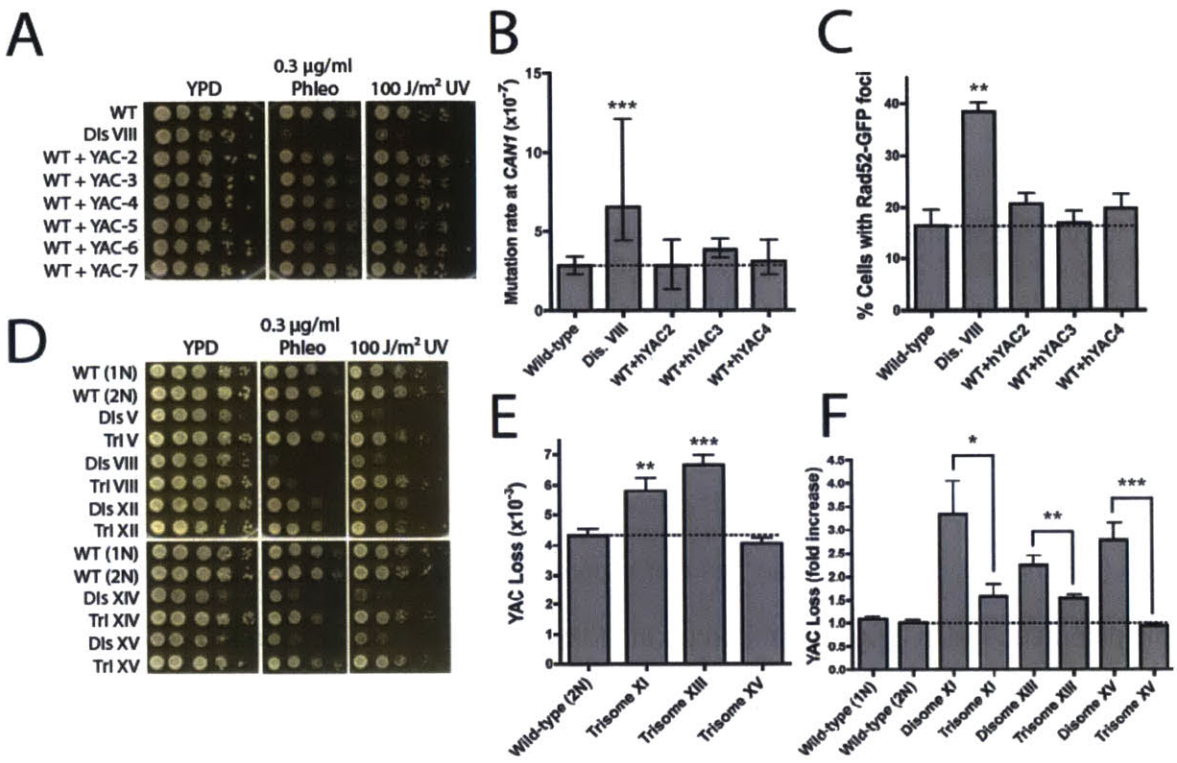


Fig. 4. Stoichiometric imbalances drive genomic instability. (A) 10-fold serial dilutions of strains harboring YACs on the indicated media. (B) The mutation rate at *CAN1*. Median and 95% confidence intervals of at least 12 independent cultures are shown. ***p<.0005 (Wilcoxon rank-sum test). (C) Fraction of nocodazole-arrested cells displaying Rad52-GFP. Mean+/-SD of three experiments is shown. **p<.005 (Student's t-test). (D) 10-fold serial dilutions of trisomic and corresponding disomic strains on the indicated medium. (E) YAC loss rates in a diploid and in three trisomic strains. Mean and standard deviation of at least 12 independent cultures are shown. (F) YAC loss rates normalized to either haploid or diploid controls. *p<.05; **p<.005; ***p<.0005 (Student's t-test).

Conclusions

This study establishes that missegregation of a single chromosome is sufficient to induce the hallmarks of genomic instability, including whole-chromosome instability, mutagenesis, and sensitivity to genotoxic stress (summarized in Table S5). Genomic instability in the disomes is not correlated with the size of the extra chromosome or the delay in cell cycle progression (Fig.S12). Thus, aneuploidy-induced genomic instability may result from imbalances in particular genes and/or from proteotoxic stress caused by aneuploidy. Aneuploid strains derived from triploid meiosis were also shown to be unstable (17) but a recent report described the construction of stable aneuploid strains using this method (8). We note that 87.5% of the spores derived from triploid meiosis in the latter study were discarded due to karyotypic instability, consistent with our finding that the vast majority (but, potentially, not all) aneuploid strains display chromosomal instability. Moreover, CGH analysis of the aneuploid strains characterized in (8) demonstrates that many have heterogenous karyotypes (figs. S13 and S14), consistent with our finding that the vast majority (but, potentially, not all) aneuploid strains display chromosomal instability. In mammals, cells derived from individuals with Down syndrome (Trisomy 21) are also sensitive to DNA damaging agents (18), and aneuploid karyotypes have been correlated with chromosomal instability in transformed Chinese hamster embryo cells (19) and in $p53^{-/-}$ colon cancer cells (20). Thus, some degree of aneuploidy-induced genomic instability may be conserved among eukaryotes.

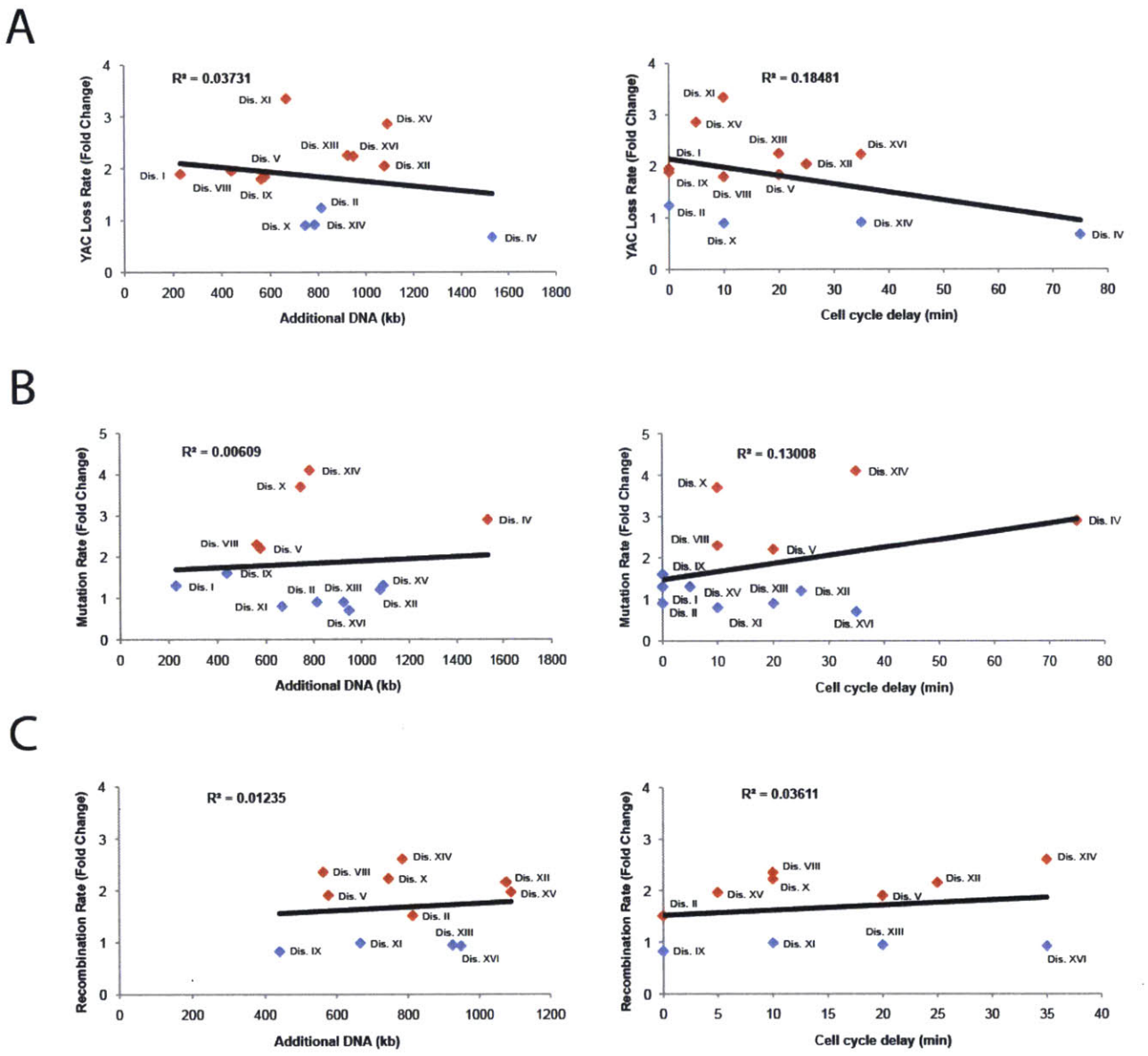


Fig. S12. **Genomic instability is not correlated with the size of the extra chromosome or the cell cycle delay in disomic strains**. Linear regression analysis (black line) reveals that the fold increase in (A) chromosome missegregation, (B) forward mutation, and (C) mitotic recombination are independent of the size of the extra chromosome and the cell cycle delay in the disomic strains. Cell cycle delay data are from *(1)*.

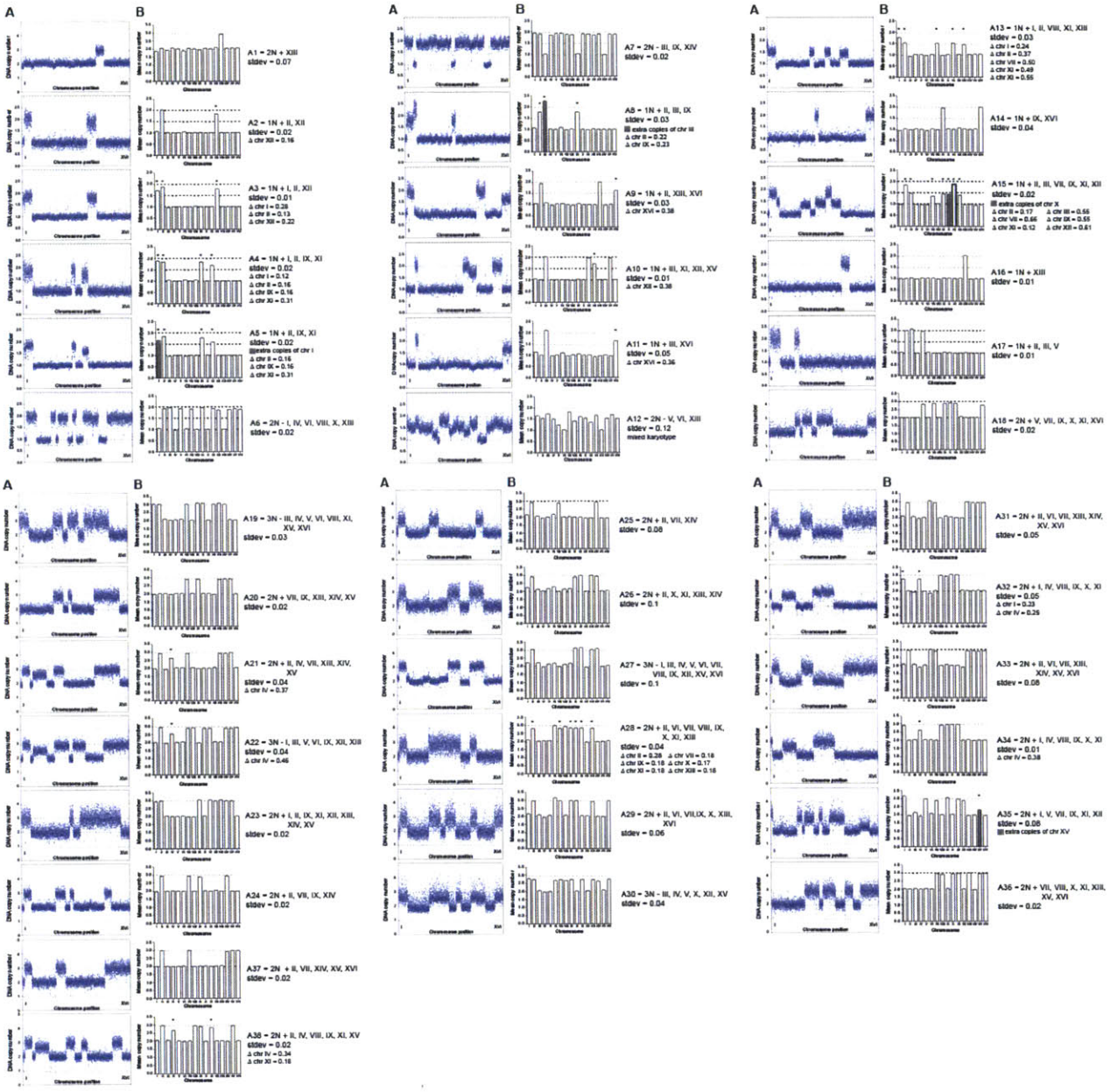


Figure S13. DNA content analysis of aneuploid strains derived from triploid meiosis.

Pavelka et al. *(26)* recently reported the analysis of aneuploid yeast strains obtained as meiotic products from triploid and pentaploid strains. 87.5% of these strains were discarded due to chromosomal instability, but 38 yeast strains were described as karyotypically stable. To determine whether these aneuploid strains displayed homogenous karyotypes, we determined their DNA content via comparative genomic hybridization (CGH) analysis as described in the

Methods. (A) CGH analysis of the 38 strains characterized in *(26)*. Each box represents the genome of an aneuploid yeast strain. Data points are ordered according to their chromosomal coordinates, starting with the gene most distally located on the left arm of chromosome I. DNA copy number was calculated from the log2 ratios of the DNA content of aneuploid strains relative to an isogenic euploid strain. 17 of the 38 aneuploid strains did not display a homogeneous karyotype. The copy number of certain chromosomes was not an integer multiple of the haploid value, indicating that different cells in the population carry a different copy number of that particular chromosome(s). (B) Average copy numbers calculated per chromosome are shown. Error bars represent the standard error of the mean per chromosome. Standard deviations (stdev) of the mean values of the euploid chromosomes are shown. The difference (Δ) between the expected value and the measured copy number is shown. Chromosomes where this difference was at least 4 times greater than the standard deviation of the euploid chromosomes are marked with an asterisk. In 4 strains, a subset of cells had gained extra copies of a chromosome; aneuploid chromosomes that were not reported in *(26)* are highlighted in grey. Note that the under-representation of the marked chromosomes cannot be explained by a dramatic decrease in signal of a few data points, because the decreased values are consistent across the chromosome. It is also not due to variability in the experimental procedure. The standard deviations of the means of the euploid chromosomes were consistently small and we only considered chromosomes as being under- or over-represented when the deviation from the expected value was at least 4-fold greater than the standard deviation of the means of the euploid chromosomes. In most cases the difference was in fact greater than 8-fold. The observation that these strains display heterogenous karyotypes, including both chromosome gains and losses, strongly suggests that they are karyotypically unstable.

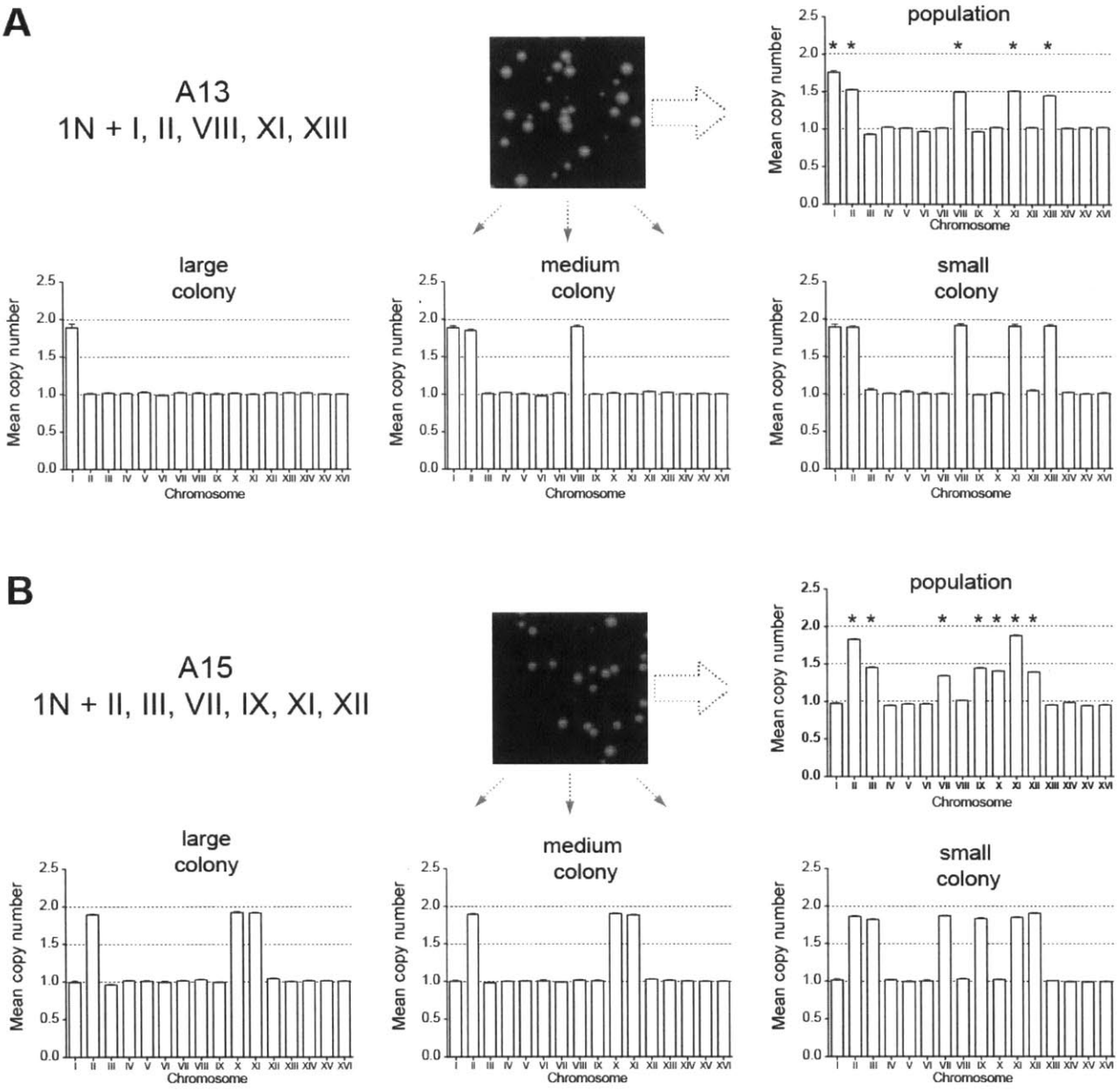


Figure S14. Strains A13 and A15 are heterogeneous. (A) We noted that single colonies derived from the strains described in *(26)* frequently displayed significant variation in colony size during routine growth on rich medium. Representative images of single colonies from strains A13 and A15 are displayed. To further test whether the aneuploid strains described in *(26)* are karyotypically unstable, we isolated single colonies of different sizes from strains A13

and A15 and analyzed their karyotype via CGH. The karyotype of strain A13 was described in *(26)* as 1N + I, II, VIII, XI, XIII. A small A13 colony exhibited the reported 1N + I, II, VIII, XI, XIII karyotype. A colony of medium size had lost the additional copies of chromosomes XI and XIII and cells of the large colony had lost all additional chromosomes except for chromosome I. (B) The analysis in (A) was repeated for strain A15. While a small colony displayed the reported karyotype of 1N + II, III, VII, IX, XI, XII, cells from the medium and large colonies had lost the extra copies of chromosomes III, VII, IX, and XII but had gained a copy of chromosome X. We conclude that many strains described in *(26)* display significant karyotypic variation.

Genomic instability provides a growth advantage during the experimental evolution of microorganisms and drives the development of tumors (21–23). While aneuploidy confers severe disadvantages to cells by stressing protein homeostasis and altering metabolism (3,4,24), our results suggest it may also benefit cells under selective pressure by increasing the likelihood that growth-promoting genetic alterations will develop. The mutagenic effects of aneuploidy that we report here may represent one mechanism by which changes in karyotype influence cancer development and evolution.

Supplemental Discussion

Analysis of genomic instability in disome IV: We noted that disome IV apparently displays lower levels of genomic instability than wild-type in three different assays (chromosome loss, forward mutation, and microsatellite instability) that use FOA-based selection. These results stand in contrast to our observations that disome IV has an increased mutation rate at *CAN1* (Fig. 2A), is sensitive to benomyl (Fig. 1B), and displays persistent Rad52-GFP foci (Fig. 3A). Disome IV has the worst proliferative capacity of the 13 disomic strains; it has a doubling time of 6-8 hours and very low plating efficiency (3). We believe that disome IV may exhibit FOA hypersensitivity which interferes with our ability to measure genomic instability in this strain using FOA-based assays. We note that disome IV exhibits elevated levels of ribonucleotide reductase (3), and RNR over-expression has been linked to FOA hypersensitivity (25). Nevertheless, results obtained with other assays confirm that disome IV displays heightened genomic instability.

Chromosome segregation analysis and experimental rationale: The increased rate of YAC loss (Fig. 1A) and benomyl sensitivity (Fig. 1B) in disomic strains demonstrates that aneuploidy

interferes with accurate chromosome segregation. We note, however, that not all strains displaying YAC loss are benomyl sensitive, and vice-versa. Strains which are sensitive to benomyl but do not display heightened YAC loss may be defective in chromosome segregation in the presence of spindle damage but not during an unperturbed cell cycle. Alternately, strains which are benomyl resistant but frequently missegregate a YAC may predominantly have defects in microtubule-independent DNA replication and repair processes.

In 5 out of 5 disomic strains, we observed prolonged Pds1 persistence, demonstrating a cell cycle delay prior to anaphase (Fig. 1C and S1). Deletion of *MAD2* suppressed this persistence in disome V, but not in four other disomes, suggesting that disome V has a chromosome alignment defect which triggers a Mad2-dependent checkpoint (Fig. 1C and S1). In order to confirm this hypothesis, we arrested disomes V and XVI in mitosis with nocodazole and then examined the kinetics of Pds1 degradation following release from the block (Fig. S2). This assay is especially sensitive in detecting bi-orientation defects, because cells have to reestablish bi-orientation of all sister chromatids following drug removal. This analysis confirmed a bi-orientation defect in disome V cells. Following the removal of nocodazole, disome XVI and wild-type strains rapidly degraded Pds1, while a degradation delay was apparent in disome V cells (Fig. S2). We conclude that in most disomes the stabilization of Pds1 and increase in chromosome loss result predominantly from Mad2-independent defects in genome replication and repair, while in disome V cells problems in sister kinetochore bi-orientation elicit a Mad2-dependent cell cycle delay.

Mutation spectra in aneuploid and wild-type strains: In order to define the molecular mechanism underlying the increased mutation rate in aneuploid cells, we sequenced the *CAN1* allele from 133 wild-type and 404 disomic canavanine-resistant isolates (Table S1). Disomic sequences included 24 to 32 *CAN1* alleles from 10 different disomes, 73 alleles from disome X

(a mutator), and 79 alleles from disome IX (a non-mutator). In order to make statistically-significant comparisons with the wild-type strain, the pooled total of all aneuploid *CAN1* alleles was initially considered. Significant differences were determined using the chi-square goodness-of-fit test. In addition to the increase in complex events discussed in the main text, we also observed that the identity of basepairs gained and lost in the disomes differed relative to wild-type in a largely strand-specific manner. For example, 11 out of 12 disomes displayed an increased frequency of substitution to adenine on the coding strand while 9 out of 12 disomes displayed a decreased frequency of substitution to adenine on the non-coding strand (Table S2). Mutation to adenine occurs preferentially opposite damaged bases, suggesting the presence of unrepaired lesions on the non-coding strand (26). This alteration in spectrum is also apparent when specific mutation events (i.e. CG>AT substitutions) are scored with respect to a particular strand, but there is no significant difference in CG>AT substitution frequency when events are scored independently of strand identity (Table S3 and S4). The only strand-independent alteration in substitution frequency is an increase in AT>TA transversions in the disomes (Table S4). Additionally, we analyzed sequence differences between the pooled total of all mutator strains (Dis. IV, VIII, X, and XIV) and the wild-type population. The decreased number of alleles considered diminished the statistical power of our tests, and some differences apparent among the pooled total of all aneuploid strains (i.e., increased substitution to adenine on the coding strand) were no longer significant. However, mutator disomes did exhibit a strand-independent decrease in CG>AT substitutions that was not present among all aneuploid alleles. Note that this is a potentially biased comparison, as disome X constitutes greater than 50% of the basepair substitutions in mutator strains and removing disome X abolishes the apparent difference. The cause of these basepair substitution frequency alterations is at present unknown.

Double-strand break formation and repair in the disomes: A common phenotype among the disomic strains is severe sensitivity to phleomycin and bleomycin, two double-strand break inducing drugs (Fig. 2C). Indeed, the sensitivity of disomes IV, VIII, X, XI, and XIV to these drugs was comparable to the sensitivity of a strain lacking Rad51. Corroborating this observation, we found that many disomes display an increased number of Rad52-GFP foci (Fig. 3A) and a moderate hyper-recombination phenotype (Fig. 3C). Two non-mutually exclusive models could account for these observations. First, disomes could accumulate more DSBs than euploid cells do, either during a normal cell cycle or upon treatment with phleomycin. Secondly, regardless of the number of DSBs which are formed, the disomes could repair the DSBs more slowly than wild-type and/or incorrectly. Our analysis of *rad52Δ* disomes suggests that disomes XI, XII, and XIV accumulate more DSB's during DNA replication, while disomes VIII and XV do not (Fig. S6). Pulse-field gel analysis of several disomes treated with phleomycin is consistent with a defect in DSB repair, though it could also be explained by defects in DNA damage avoidance: chromatin in disomes may be more vulnerable than euploid chromatin to phleomycin-induced lesions (Fig. 3D and S9). However, we believe that a drug efflux problem is unlikely to account for any differences in phleomycin sensitivity, as the disomes are also sensitive to UV light (Fig. 2B). We further note that a DSB-repair defect is not inconsistent with our finding that disomes are moderately hyper-recombinant, as many mutations in genes which compromise DNA repair also display increased mitotic recombination (i.e., *ELG1* (27), *RTT109* (28), and *SGS1* (29).

Causes of aneuploidy-induced genomic instability: We have demonstrated that replication of an extra chromosome *per se* is not sufficient to impair recombination or DNA damage resistance, but that it is likely that stoichiometric imbalances in yeast proteins cause genomic instability (Fig. 4). However, aneuploidy-induced genomic instability is uncorrelated with the

size of the extra chromosome (Fig. S12). This may suggest that genomic instability is caused by imbalances in a few specific genes which are randomly distributed throughout the genome. For instance, altering the stoichiometry of histones and spindle pole proteins has been shown to interfere with accurate chromosome segregation (30–32). Alternately, proteotoxic stress caused by the translation, folding, and degradation of many proteins on the extra chromosomes (3,4) could contribute to genomic instability. Aneuploidy may monopolize cellular chaperones and protein turnover complexes, thereby inhibiting the function of other proteins that normally require their use. Distinguishing between these possibilities will be an important aspect of our future studies.

Material and Methods

Strains and Plasmids: All budding yeast strains used in this study are derivates of W303 (A2587/A2588) and are listed in Tables S6 and S7. Strains were constructed using PCR-based methods for gene deletion and tagging (33). The generation of aneuploid strains has been previously described (3). All aneuploid strains were analyzed by comparative genomic hybridization (CGH) to ensure that the specified extra chromosome was present in its entirety (3). YACs used in this study are listed in Table S8 (34,35).

Analysis of genotoxin sensitivities: For each assay, fresh plates were prepared containing the concentrations of genotoxins indicated in the figures. Plates containing phleomycin and camptothecin were buffered with 50mM HEPES (pH 7.4). For each assay, strains were inoculated O/N in appropriate selective media, then 10-fold serial dilutions were prepared in sterile water and spotted onto plates. Camptothecin and UV-treated plates were incubated in the dark. In Fig. 1B, the following strains were used: A11311, A24694, A12683, A12685,

A12687, A14479, A13628, A13975, A12689, A13771, A12695, A13979, A12697, and A12700. In Fig. 2B and 2C, the following strains were used: A11311, A21598, A12683, A12685, A12687, A14479, A13628, A13975, A12689, A13771, A12695, A13979, A12697, and A12700. In Fig. 3E, the following strains were used: HM123, P219, and P523. In Fig. 4A, the following strains were used: A11311, A13628, A17392, A17393, A17394, A17395, A17396, and A23744. In Fig. 4D, the following strains were used: A11311, A18344, A14479, A18346, A13628, A18347, A12693, A19549, A13979, A18349, A12697, and A18350.

Determination of Chromosome Loss Rates: The rate of chromosome loss was calculated as previously described (36). YAC yWSS1572, which is marked with *TRP1* and *URA3* at opposing telomeres, was crossed into the strains of interest (34). Strains were first inoculated O/N in –HIS-TRP-URA+G418 medium. Subsequently, ~1000 cells were transferred to 2ml of –HIS+G418 medium, and appropriate dilutions were plated onto –HIS+G418 and –HIS+G418 supplemented with 1mg/ml 5-fluoroorotic acid (FOA) to determine the fraction of cells which had lost the YAC prior to the start of the assay. Cells were allowed to divide for 24-36 hours, then were diluted and plated on –HIS+G418 and –HIS+G418+FOA. For chromosome loss experiments in diploid and trisomic strains, cells were pre-grown in –HIS-LEU-URA-TRP+G418 medium, transferred to –HIS-LEU+G418 medium, then plated on either –HIS-LEU+G418 or –HIS-LEU+G418+FOA. The number of cells that maintained the chromosome during the assay was calculated using the formula $S = F(1-m)$, where F is the total number of cells and m is the fraction of cells which lost the chromosome during the assay. The chromosome loss rate per generation was calculated using the formula $L = 1 - \frac{(\frac{S}{C})^{\frac{1}{N}}}{2}$, where C is the number of cells initially inoculated and n is the number of generations that the culture grew. Cell numbers for

plating were determined using a Beckman Coulter Multisizer 3. For 6 cultures from each strain, colonies that grew on FOA plates were replica plated to –TRP plates to control for mutation or loss of a single chromosome arm. On average, >97% of FOA^R colonies were also Trp-, demonstrating that whole-chromosome loss was almost always the cause of FOA resistance. Chromosome loss rates are presented as the mean and standard deviation of at least 12 independent cultures. In Fig. 1A, the following strains were used: A23744, A24792, A24660, A23745, A25827, A25340, A23747, A23748, A23749, A25587, A23750, A23751, A25341, A23752, A23753, and A25627. In Fig. 4E, the following strains were used: A27889, A27890, A27891, and A27892. In Fig. 4F, the following strains were used: A23744, A27889, A23750, A27890, A25341, A27891, A23753, and A27892.

Cell cycle analyses of Pds1 levels and spindle morphology: Cells were inoculated into –HIS +G418 medium and grown to log phase. The cells were subsequently collected by filtration and resuspended at an OD_{600} of 0.175 in YPD medium containing 5 µg/ml α-factor to arrest the cells in G1. 2.5 µg/ml α-factor was re-added every 75 minutes to maintain cells in the G1 arrest. When both the WT and disomic strains were arrested, cells were washed by filtration with 8 volumes of YPD, and the cells were resuspended in fresh YPD medium lacking pheromone. Samples were taken at the indicated time points, and α-factor (5 µg/ml) was readded when more than 50 percent of the cells had formed a bud to prevent cells from entering the next cell cycle. To account for variability between experiments, a wild-type control culture was always analyzed concomitantly with the disomic strains. In Fig. 1C, strains A26703 and A26628 were used.

Western blot analyses: Proteins were extracted from aliquots taken at the indicated time points by adding an equal volume of 10% trichloroacetic acid to the cell culture. The samples were incubated on ice for at least 20 minutes, then washed with 1.5ml of acetone. The dried pellet was resuspended in 100 μL of protein breakage buffer (50mM Tris, pH 7.5, 1mM EDTA, 2.75 mM DTT, and Roche Complete protease inhibitor, used per the manufacturer's instructions). 100 μL of glass beads were added and the cells broken by beating for 2.5 minutes on a Biospec mini-bead beater. 50 μL of 3X SDS sample buffer was added, the samples boiled for 5 minutes, then centrifuged for 5 minutes. An equal volume of lysate was loaded onto 10% SDS polyacrylamide gel, electrophoresed, and transferred to a nitrocellulose membrane. Pds1-3HA was detected using a mouse anti-HA antibody (HA.11, Covance) at a 1:1,000 dilution. Pgk1 was detected using a mouse anti-Pgk1 antibody (A-6457, Molecular Probes) at a 1:5,000 dilution. The secondary antibody was a sheep anti-mouse antibody coupled to horseradish peroxidase (NA931, GE Healthcare) and used at a 1:5,000 dilution. Rad53 was detected using a goat polyclonal antibody (yC-19, Santa Cruz Biotechnology) at a 1:5,000 dilution, followed by donkey anti-goat antibody coupled to horseradish peroxidase (sc2020, Santa Cruz Biotechnology) at a 1:10,000 dilution. Bands were detected using Pierce SuperSignal West Pico Chemiluminescent Substrate according to the manufacturer's instructions.

Fixation and immunofluorescence of yeast cells: Mitotic spindles were visualized as previously described (37), with the following modifications. At each time point, one ml of cell culture was collected and fixed in 3.7% formaldehyde in 0.1 M potassium phosphate buffer. After digestion for 60 minutes with glusulase and zymolyase, cells were applied to polylysine-coated multiwall slides and fixed for 3 minutes in methanol followed by 10 seconds in acetone, both at -20°C. Rat anti-tubulin antibodies (Oxford Biotechnology) and anti-rat FITC antibodies

(Jackson Immunoresearch) were used at a 1:100 dilution for visualization of spindle morphology. At least 200 cells were counted per strain for each time point.

Fluctuation analysis: To determine the mutation rate at *CAN1*, ~1000 canavanine-sensitive cells were inoculated in 5ml of –HIS+G418 medium and grown to saturation. Appropriate dilutions were then plated on –HIS+G418 and on -HIS-ARG+G418 medium supplemented with 60µg/ml canavanine to determine the number of viable cells and canavanine-resistant cells, respectively. The rate of spontaneous mutation was determined according to the method of the median (38,39). Reported mutation rates for each strain are the median values of at least 12 independent cultures. Cell numbers for plating were determined using a Beckman Coulter Multisizer 3. In Fig. 2A, the following strains were used: A20814, A27898, A21630, A25328, A25329, A25330, A25331, A25332, A25333, A25334, A25335, A25336, A25337, A25338, and A25339. To determine the mutation rate at *URA3*, an identical protocol was followed, and cells were plated on –HIS +G418+FOA. In Fig. 2A, the following strains were used: A25589, A25590, A25591, A25592, and A25593. In Fig. 4B, the following strains were used: A20814, A25331, A26980, A26981, and A27683.

The *CAN1* gene is located on chromosome V, and mutations conferring canavanine resistance are recessive. Therefore, to calculate the rate of mutation of a strain disomic for chromosome V, fluctuation analysis was performed for mutation at *LYP1*, which is located on chromosome XIV and confers sensitivity to thialysine (40). –HIS-LYS+G418 plates supplemented with 100µg/ml thialysine were used to determine the fraction of thialysine-resistant cells, and the mutation rate was calculated as described above. For this analysis in Fig. 2A, the following strains were used: A24351and A14479.

To determine the rate of mitotic recombination, we used a construct in which a 5'-truncated and a 3'-truncated allele of *ade2* are separated by *URA3* (41). ~100 Ura+ cells were inoculated into 5ml of –HIS+G418 medium and grown to saturation. Mitotic recombination was measured by plating on –HIS+G418+FOA to determine the fraction of cells which lost the intervening *URA3* sequence. The rate of mitotic recombination was calculated using the method of the median as above. In Fig. 3C, the following strains were used: A26461, A26473, A26475, A26462, A26463, A26464, A26465, A26466, A26467, A26468, A26469, A26470, A26471, and A26472.

To determine the rate of microsatellite instability, a plasmid containing *TRP1* and a 16.5-poly(GT) sequence in-frame with *URA3* was crossed into the disomes (42). ~100 URA+ cells were inoculated into 5ml of medium and grown to saturation. The fraction of cells that maintained the plasmid was calculated by plating on –HIS-TRP+G418, and the fraction of plasmid-maintaining cells in which *URA3* was out of frame was calculated by plating on –HIS-TRP+G418+FOA. The rate of microsatellite instability was determined using the method of the median as above.

***CAN1* sequencing:** DNA from canavanine-resistant colonies was purified by phenol-chloroform extraction (43). The *CAN1* locus was amplified using primers 5'-TCTTCAGACTTCTTAACTCC-3' and 5'-ATAGTAAGCTCATTGATCCC-3'. Sequencing was performed using the following primers (44): 5'- AAAAAAGGCATAGCAATGAC-3', 5'-GACGTACAAAGTTCCACTGG-3', 5'- TCAAAGAACAAGTTGGCTCC-3', 5'-TAGATGTCTCCATGTAAGCC-3', 5'- AACTTTGATGGAAGCGACCC-3', and 5'-GAAATGGCGGGGAAATGTG-3'.

Rad52-GFP foci analysis and live cell microscopy: To follow the distribution of Rad52-GFP foci during the cell cycle, cells were arrested in G1 with 10 μg/ml α-factor for 120-180 minutes, then washed and released into –HIS+G418 medium lacking pheromone. Samples were removed every 30-45 minutes to determine the fraction of cells with Rad52-GFP foci. To arrest cells in mitosis, cells were treated with 15 μg/ml nocodazole for 90 minutes. In Fig. 3A, strains used for this assay were as follows: A24352, A26532, A26533, A25342, A25343, A25421, A25344, and A25345. In Fig. 4C, strains used for this assay were as follows: A24352, A25342, A26985, A26986, A27667.

To determine whether Rad52-GFP foci appear in other types of aneuploid yeast strains, we created diploid and triploid strains homozygous for the gene encoding the Rad52-GFP fusion protein. Cells were sporulated, then tetrads were dissected on YPD plates. Visible colonies which formed from euploid or aneuploid spores were inoculated in YPD for 3-4 hours, then the frequency of Rad52-GFP foci appearance was scored among budded cells. We note that the actual incidence of Rad52-GFP foci appearance may be even greater among all progeny of triploid meiosis, as spores which form the most foci are unlikely to grow into visible colonies. In Fig. 3B, strains used were A28502 and A27997.

To follow the repair of phleomycin-induced double-strand breaks, cells were arrested in G1 as described above and then released into medium containing 0.15μg/ml phleomycin and buffered with 50mM HEPES (pH 7.4). After 135 minutes, cells were washed twice and transferred to fresh medium. Cells were removed every 30 minutes to score the fraction of large-budded cells and cells containing Rad52-GFP foci.

IR experiments were performed with a Gammacell 40, which uses ^{137}Cesium as a radiation source. Exponentially-growing cultures were treated with 4.5krads to induce double-strand breaks, then samples were removed every 60 minutes to determine the fraction of large-budded cells and cells containing Rad52-GFP foci.

Microscopy was performed using a Zeiss Axioplan 2 microscope with a Hamamatsu OCRA-ER digital camera. Image analysis was performed with Openlab 4.0.2 software.

Pulsed Field Gel Electrophoresis: To directly visualize chromosome damage and repair following phleomycin treatment, yeast cultures were grown overnight in –HIS+G418 medium to log phase. Cells were then harvested and resuspended in fresh YPD medium buffered with 50mM HEPES (pH 7.4) at an OD_{600} of 0.3. Phleomycin was added to a final concentration of 25µg/mL. After two hours, cells were washed with 5 volumes of fresh HEPES buffered YPD and resuspended in HEPES buffered YPD containing 15 µg/ml nocodazole to arrest cells at mitosis to prevent further cell division.

50 ml cell culture aliquots were collected and fixed with sodium azide just prior to the addition of phleomycin (pre), at the end of phleomycin treatment (time 0), and at the indicated timepoints during recovery in YPD medium containing nocodazole. Cell density was determined with a Beckman Coulter Multisizer 3 before and after phleomycin treatment. This allowed us to adjust for cell number increase during drug treatment when making the DNA agarose plug for the "pre" sample, such that the number of cells loaded on the gel was equivalent for the "pre" and the subsequent time points. Yeast DNA agarose plugs were prepared and processed as described (45) by embedding cells in low melting point agarose to a final concentration of 0.625% agarose. The "pre" sample plug was embedded in a smaller volume such that the concentration of cells in the agarose plug would equal that of the others. Plugs were melted and equal volumes loaded on a 1% agarose gel. DNA was electrophoresed using a Bio-Rad CHEF-DR II system at 6 V/cm at 14°C. The switch time was 60s for 15 hours followed by 90s for 9 hours. The gel was stained with ethidium bromide and imaged using a GE ImageQuant LAS 4000. In all the images, the wild-type strain was both collected and run concurrently with the disomic strain(s) shown on the same gel. Because multiple disomic strains were collected

during an experiment, some of the final gel images were spliced in order to place the wild-type results next to the disomic strain. No modifications exist in exposure time, contrast, etc. of these spliced images so that the wild-type and disomic strain are directly comparable. In Fig. 3D, the following strains were used: A11311 and A13771.

***Schizosaccharomyces pombe* Genetics and Microscopy**: Strains of fission yeast used in this study are listed in Table S9. The construction of P219, a strain disomic for chromosome III, was previously described (14). A triploid strain for sporulation was created by crossing P321 with the mating-competent diploid P322. The products of a triploid meiosis include various combinations of disomy as well as haploid and diploid cells. Strains disomic for chromosome I and/or II arrest before forming a full-size colony, or undergo an unequal nuclear division that results in euploid progeny (13). Aneuploid cells can be easily identified during the microcolony stage due to their delayed germ tube formation and aberrant colony morphology.

To analyze the frequency of SpRad22 foci in aneuploid and euploid cells, spores from a triploid meiosis were transferred to YEA medium and incubated at 30°C. Spores that had germinated and undergone multiple rounds of cell division within 25.5 hours after transfer to YEA were identified as euploid and the frequency of SpRad22 foci was determined. Spores that showed delayed germination and aberrant colony morphology after 48 hours on YEA were identified as aneuploid and the frequency of SpRad22 foci was determined as above.

For time-lapse observations of SpRadequations22 foci formation, spores from a triploid meiosis were transferred to YEA medium and incubated at 30°C. Spores that had germinated within 8 hours were identified as euploid and were transferred to a cover slip. Images were acquired every 30 minutes to follow cell division and SpRad22 foci formation. Spores that had

germinated but not divided by 17.5 hours were identified as aneuploid and were followed via microscopy as described above.

Acknowledgements

We thank D. Koshland, T. Petes, R. Li, L. Symington, D. Page, T. Matsumoto, and J. Takeda for reagents, and I. Cheeseman, J. Haber, M. Rose, F. Solomon, and the Amon lab for comments on the manuscript. J.M.S. and S.J.P. are supported by NSF Graduate Fellowships and A. A by grant GM056800. A.A. is an HHMI investigator.

References

1. Weaver BAA, Cleveland DW. Does aneuploidy cause cancer? Curr Opin Cell Biol. 2006;18:658–67.

2. Matzke MA, Mette MF, Kanno T, Matzke AJM. Does the intrinsic instability of aneuploid genomes have a causal role in cancer? Trends Genet TIG. 2003;19:253–6.

3. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, et al. Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. Science. 2007;317:916–24.

4. Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, et al. Identification of aneuploidy-tolerating mutations. Cell. 2010;143:71–83.

5. Musacchio A, Salmon ED. The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol. 2007;8:379–93.

6. See Materials and Methods.

7. Harfe BD, Jinks-Robertson S. DNA Polymerase [zeta] Introduces Multiple Mutations when Bypassing Spontaneous DNA Damage in Saccharomyces cerevisiae. Mol Cell. 2000;6:1491–9.

8. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, et al. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature. 2010;468:321–5.

9. Rattray AJ, Shafer BK, McGill CB, Strathern JN. The roles of REV3 and RAD57 in double-strand-break-repair-induced mutagenesis of Saccharomyces cerevisiae. Genetics. 2002;162:1063–77.

10. Chen J, Stubbe J. Bleomycins: towards better therapeutics. Nat Rev Cancer. 2005;5:102–12.

11. Lisby M, Rothstein R, Mortensen UH. Rad52 forms DNA repair and recombination centers during S phase. PNAS. 2001;98:8276–82.

12. Paques F, Haber JE. Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev. 1999;63:349–404.

13. Niwa O, Tange Y, Kurabayashi A. Growth arrest and chromosome instability in aneuploid yeast. Yeast. 2006;23:937–50.

14. Niwa O, Yanagida M. Triploid meiosis and aneuploidy in Schizosaccharomyces pombe: an unstable aneuploid disomic for chromosome III. Curr Genet. 1985;9:463–70.

15. Meister P, Poidevin M, Francesconi S, Tratner I, Zarzov P, Baldacci G. Nuclear factories for signalling and repairing DNA double strand breaks in living fission yeast. Nucleic Acids Res. 2003;31:5064–73.

16. Takeda J, Uematsu N, Shiraishi S, Toyoshima M, Matsumoto T, Niwa O. Radiation induction of delayed recombination in Schizosaccharomyces pombe. DNA Repair. 2008;7:1250–61.

17. Charles JS, Hamilton ML, Petes TD. Meiotic Chromosome Segregation in Triploid Strains of Saccharomyces cerevisiae. Genetics. 2010;186:537–50.

18. Natarajan AT. Radiosensitivity of Cells Derived from Down Syndrome Patients. DNA Repair Hum Dis. 2006.

19. Duesberg P, Rausch C, Rasnick D, Hehlmann R. Genetic instability of cancer cells is proportional to their degree of aneuploidy. PNAS. 1998;95:13692–7.

20. Thompson SL, Compton DA. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. J Cell Biol. 2010;188:369–81.

21. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature. 1998;396:643–9.

22. Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of E. coli. Nature. 1997;387:703–5.

23. Shaver AC, Dombrowski PG, Sweeney JY, Treis T, Zappala RM, Sniegowski PD. Fitness evolution and the rise of mutator alleles in experimental Escherichia coli populations. Genetics. 2002;162:557–66.

24. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, et al. Aneuploidy Affects Proliferation and Spontaneous Immortalization in Mammalian Cells. Science. 2008;322:703–9.

25. Rossmann MP, Luo W, Tsaponina O, Chabes A, Stillman B. A Common Telomeric Gene Silencing Assay Is Affected by Nucleotide Metabolism. Mol Cell. 2011;42:127–36.

26. Strauss BS. The "A rule" of mutagen specificity: a consequence of DNA polymerase bypass of non-instructional lesions? BioEssays News Rev Mol Cell Dev Biol. 1991;13:79–84.

27. Bellaoui M, Chang M, Ou J, Xu H, Boone C, Brown GW. Elg1 forms an alternative RFC complex important for DNA replication and genome integrity. EMBO J. 2003;22:4304–13.

28. Driscoll R, Hudson A, Jackson SP. Yeast Rtt109 Promotes Genome Stability by Acetylating Histone H3 on Lysine 56. Science. 2007;315:649–52.

29. Mankouri HW, Ngo H-P, Hickson ID. Esc2 and Sgs1 Act in Functionally Distinct Branches of the Homologous Recombination Repair Pathway in Saccharomyces cerevisiae. Mol Biol Cell. 2009;20:1683–94.

30. Au W-C, Crisp MJ, DeLuca SZ, Rando OJ, Basrai MA. Altered Dosage and Mislocalization of Histone H3 and Cse4p Lead to Chromosome Loss in Saccharomyces cerevisiae. Genetics. 2008;179:263–75.

31. Chial HJ, Giddings TH, Siewert EA, Hoyt MA, Winey M. Altered dosage of the Saccharomyces cerevisiae spindle pole body duplication gene, NDC1, leads to aneuploidy and polyploidy. PNAS. 1999;96:10200–5.

32. Meeks-Wagner D. Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. Cell. 1986;44:43–52.

33. Longtine MS, III AM, Demarini DJ, Shah NG, Wach A, Brachat A, et al. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast. 1998;14:953–61.

34. Huang D, Koshland D. Chromosome integrity in Saccharomyces cerevisiae: the interplay of DNA replication initiation factors, elongation factors, and origins. Genes Dev. 2003;17:1741–54.

35. Foote S, Vollrath D, Hilton A, Page D. The human Y chromosome: overlapping DNA clones spanning the euchromatic region. Science. 1992;258:60–6.

36. Dani GM, Zakian VA. Mitotic and meiotic stability of linear plasmids in yeast. PNAS. 1983;80:3406–10.

37. Kilmartin JV, Adams AE. Structural rearrangements of tubulin and actin during the cell cycle of the yeast Saccharomyces. J Cell Biol. 1984;98:922–33.

38. Lea D, Coulson C. The distribution of the numbers of mutants in bacterial populations. J Genet. 1949;49:264–85.

39. Hall BM, Ma C-X, Liang P, Singh KK. Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics. 2009;25:1564–5.

40. Ito-Harashima S, Hartzog PE, Sinha H, McCusker JH. The tRNA-Tyr Gene Family of Saccharomyces cerevisiae: Agents of Phenotypic Variation and Position Effects on Mutation Frequency. Genetics. 2002;161:1395–410.

41. Fiorentini P, Huang K, Tishkoff D, Kolodner R, Symington L. Exonuclease I of Saccharomyces cerevisiae functions in mitotic recombination in vivo and in vitro. Mol Cell Biol. 1997;17:2764–73.

42. Strand M, Prolla TA, Liskay RM, Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature. 1993;365:274–6.

43. Hoffman CS, Winston F. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. Gene. 1987;57:267–72.

44. Lang GI, Murray AW. Estimating the Per-Base-Pair Mutation Rate in the Yeast Saccharomyces cerevisiae. Genetics. 2008;178:67–82.

45. Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell. 1984;37:67–75.

Chapter 3: Transcriptional Consequences of Aneuploidy

Reprinted from PNAS:

Sheltzer JM, Torres EM, Dunham MJ, Amon A. Transcriptional consequences of aneuploidy. Proc Natl Acad Sci U S A 2012;109:12644–9.

Abstract

Aneuploidy, or an aberrant karyotype, results in developmental disabilities and has been implicated in tumorigenesis. However, the causes of aneuploidy-induced phenotypes and the consequences of aneuploidy on cell physiology remain poorly understood. We have performed a meta-analysis on gene expression data from aneuploid cells in diverse organisms, including yeast, plants, mice, and humans. We found highly-related gene expression patterns that are conserved between species: genes that were involved in the response to stress were consistently upregulated, while genes associated with the cell cycle and cell proliferation were downregulated in aneuploid cells. Within species, different aneuploidies induced similar changes in gene expression, independent of the specific chromosomal aberrations. Taken together, our results demonstrate that aneuploidies of different chromosomes and in different organisms impact similar cellular pathways and cause a stereotypical anti-proliferative response that must be overcome prior to transformation.

Introduction

Eukaryotic organisms have evolved elaborate mechanisms which ensure that chromosomes are partitioned equally during cell division (1). Aberrant segregation events can result in aneuploidy, a condition in which cells acquire a karyotype that is not a whole-number multiple of the haploid complement. In humans, aneuploidy is the leading cause of spontaneous abortions and developmental disabilities, and aneuploid karyotypes are observed in greater than 90% of solid tumors (2–4). Thus, understanding the consequences of aneuploidy has broad relevance for the study of mammalian development and cancer.

The cause of aneuploidy-induced syndromes remains an open question. For instance, it has been hypothesized that the phenotypes associated with Down Syndrome (Trisomy 21) are caused by the triplication of a small set of genes which lie within a 5.4Mb "Down Syndrome Critical Region" on chromosome 21 (5,6). However, evidence from mouse models suggest that this region is not sufficient to recapitulate Down syndrome-like phenotypes, and genetic mapping of partially-trisomic individuals has revealed that numerous regions of chromosome 21 affect clinical presentation (7–9). Alternately, changes in gene dosage across an entire chromosome might have additive effects on organismal development (10). It has been observed that the three human trisomies which survive until birth (Trisomy 13, 18, and 21) have the fewest protein-coding genes on them, implying that these karyotypes can be tolerated *in utero* because they have the lowest net dosage imbalances (11). However, the consequences of copy number variation range from benign to extremely deleterious, demonstrating that different genes exhibit varying levels of dosage-sensitivity (12).

In order to examine the consequences of aneuploidy, we have previously constructed and analyzed a series of haploid budding yeast strains and mouse embryonic fibroblasts (MEFs) that carry single extra chromosomes (13–17). These aneuploid cells display defects in cell proliferation as well as sensitivity to conditions that interfere with protein folding and

turnover. Microarray expression analyses revealed that haploid yeast strains carrying an additional chromosome (henceforth disomes) exhibit a common transcriptional signature, dubbed the "environmental stress response" (ESR). The ESR consists of ~300 genes which are upregulated and ~600 genes which are downregulated by various exogenous stresses, including heat shock or oxidative stress (18). Most of these genes also vary in expression in response to growth rate; inducing slow proliferation by nutrient limitation mimics the ESR (19,20). No such common signature has been reported among aneuploid cells in any other organism. Additionally, it has been suggested (21) that the phenotypes detected in the disomic strains may be a unique consequence of the method that was employed to generate aneuploidy, in which genetic markers were used to select for rare chromosome transfer events between nuclei (13). In order to further our understanding of aneuploidy, we examined gene expression data from aneuploid cells from diverse organisms. We detected a conserved transcriptional response that was associated with stress and decreased cell proliferation that was apparent in aneuploid yeast, plants, mice and humans. These data suggest that aneuploidy in various species is detrimental to cell fitness, and that many consequences of aneuploidy are a common response to chromosome-wide dosage imbalances.

Results

Aneuploid strains of budding yeast share a chromosome-independent stress response

We previously reported that disomic yeast produced via chromosome transfer exhibit an ESR, in which genes related to RNA processing and the ribosome are downregulated, and genes involved in protein folding, detoxification of reactive oxidative species, and various other processes are upregulated (13). We sought to determine whether aneuploid cells generated in other ways also exhibited a stress and/or slow growth response similar to that observed in the

disomes. For all subsequent analyses, aneuploid chromosomes were excluded from consideration so as to avoid artifacts due to normalization. We first compared the disomes to 23 aneuploid strains generated during the construction of the yeast deletion collection (22). These aneuploidies include multiple chromosome gains and losses, and thus display more complex karyotypes than those present in the disomes. We found that the Pearson correlation coefficient (PCC) between the mean expression levels in these strains and in the disomic yeast was 0.37 ($p<10^{-187}$), while the PCC between these strains and a wild-type strain was 0.02 ($p>.05$), demonstrating highly significant transcriptional similarity between these aneuploid populations (Fig. 1A and 1B).

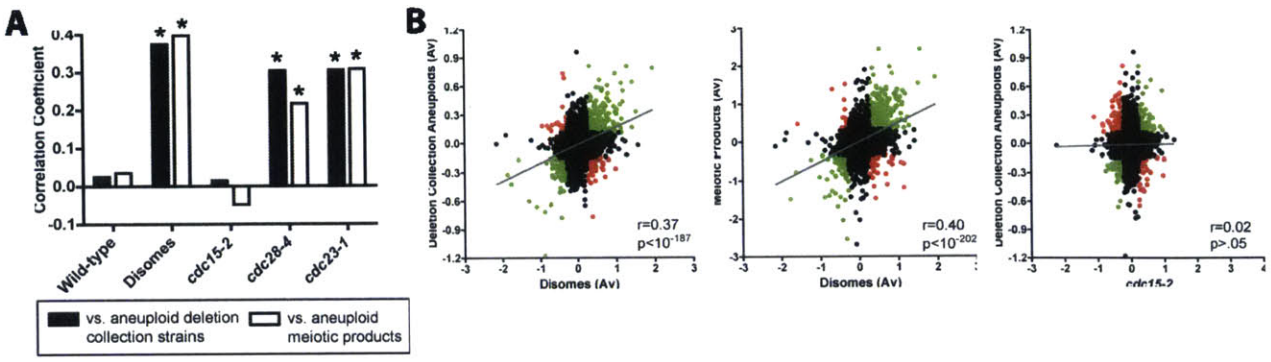


Figure 1. Transcriptional similarities among aneuploid strains of *S. cerevisiae*. (A) The correlation coefficients between either aneuploid strains from the yeast deletion collection (black bars) or aneuploid products of triploid meiosis (white bars) and the indicated strains of *S. cerevisiae* are displayed. Asterisks denote statistically significant correlations ($p<.05$). (B) Scatter plots comparing gene expression values between the indicated yeast strains. Points in green represent genes that are expressed +/- 1SD from the mean co-directionally, while points in red represent genes that are expressed +/- 1SD in opposite directions. Gray lines are linear regressions plotted against the data.

We hypothesized that the correlation between strains was due to a shared underlying stress or slow growth response. To test this, we compared the aneuploid strains obtained from the deletion collection to gene expression data from three *cdc* mutant strains: *cdc28-4* and *cdc23-1*, which divide slowly and exhibit a significant ESR at the permissive temperature, and *cdc15-2*, which proliferates at a wild-type rate at the permissive temperature and does not display an ESR (13). Both *cdc28-4* and *cdc23-1* exhibited significant correlations with aneuploid strains from the deletion collection ($r=0.30$, $p<10^{-114}$, and $r=0.31$, $p<10^{-116}$, respectively), while *cdc15-2* was uncorrelated with the aneuploid strains (Fig. 1A and 1B; $r=0.02$, $p>.05$). Next, we quantified the intensity of the stress response in each disomic strain by averaging the expression levels of genes annotated to the ESR (Fig. S1). We found that 15 out of 16 disomes exhibited significant pairwise correlations with the average expression level in the aneuploid strains obtained from the deletion collection, and the strength of the correlation increased with the intensity of the stress response in the disomes (Fig. S2A and S2B; $r=0.64$, $p<.006$). Disomes that did not display an ESR (e.g., disome I) exhibited minimal correlations with the deletion collection aneuploid strains, while disomes that displayed significant ESRs (e.g., disome IV) tended to exhibit stronger correlations. These results are consistent with our hypothesis that a shared transcriptional stress response underlies the similarity between different aneuploid populations.

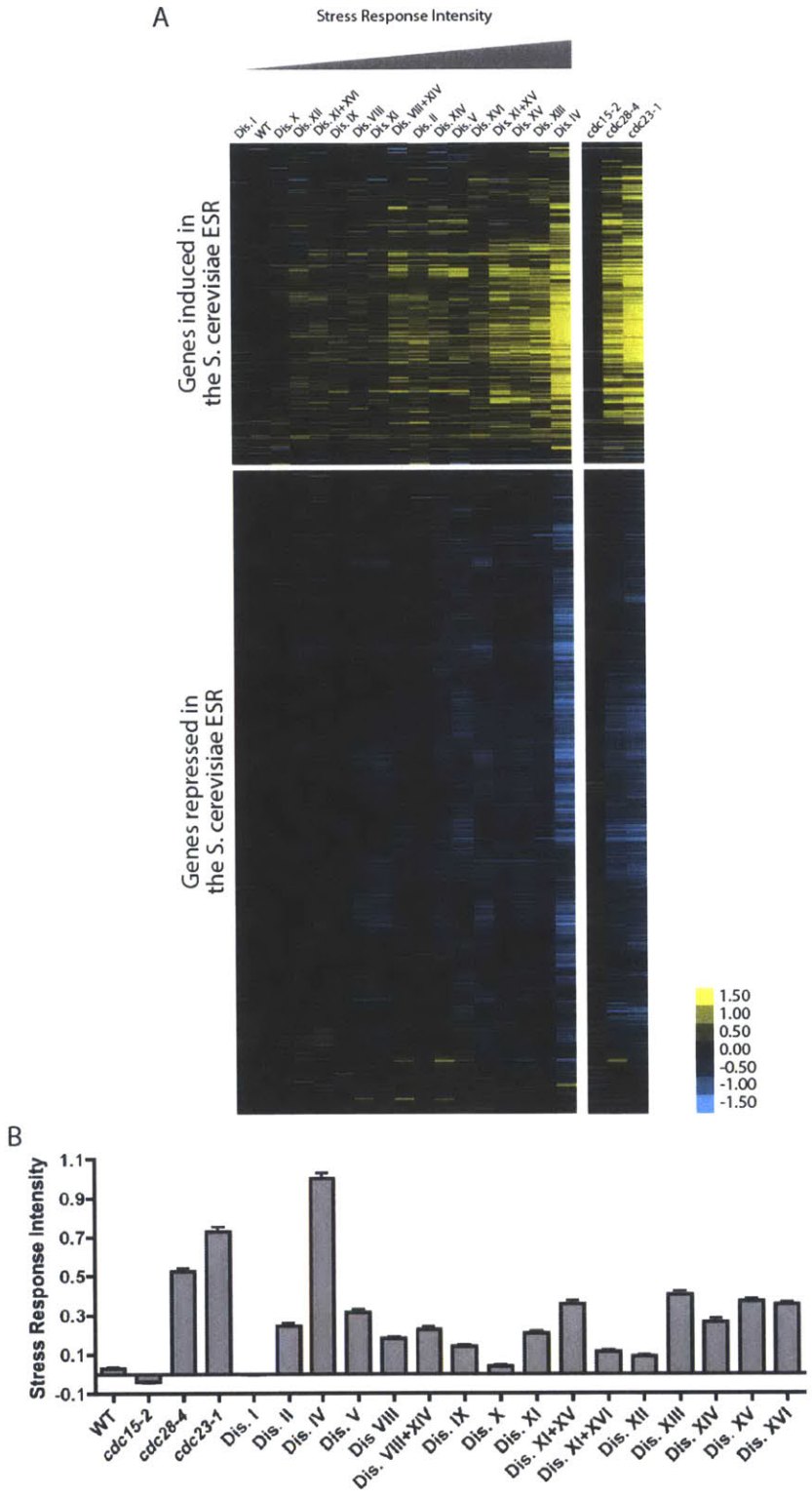


Figure S1. The environmental stress response (ESR) in disomic strains of *S. cerevisiae*.

(A) Strains were sorted according to their stress response intensity and genes annotated to the

ESR were clustered. Genes that constitute the *S. cerevisiae* ESR were downloaded from http://genome-www.stanford.edu/yeast_stress/. The stress response intensity in each disome was calculated by averaging the expression levels of genes upregulated in the ESR and then subtracting the average expression level of genes downregulated in the ESR. (B) The stress response intensities of each disomic and *cdc* mutant strain are displayed.

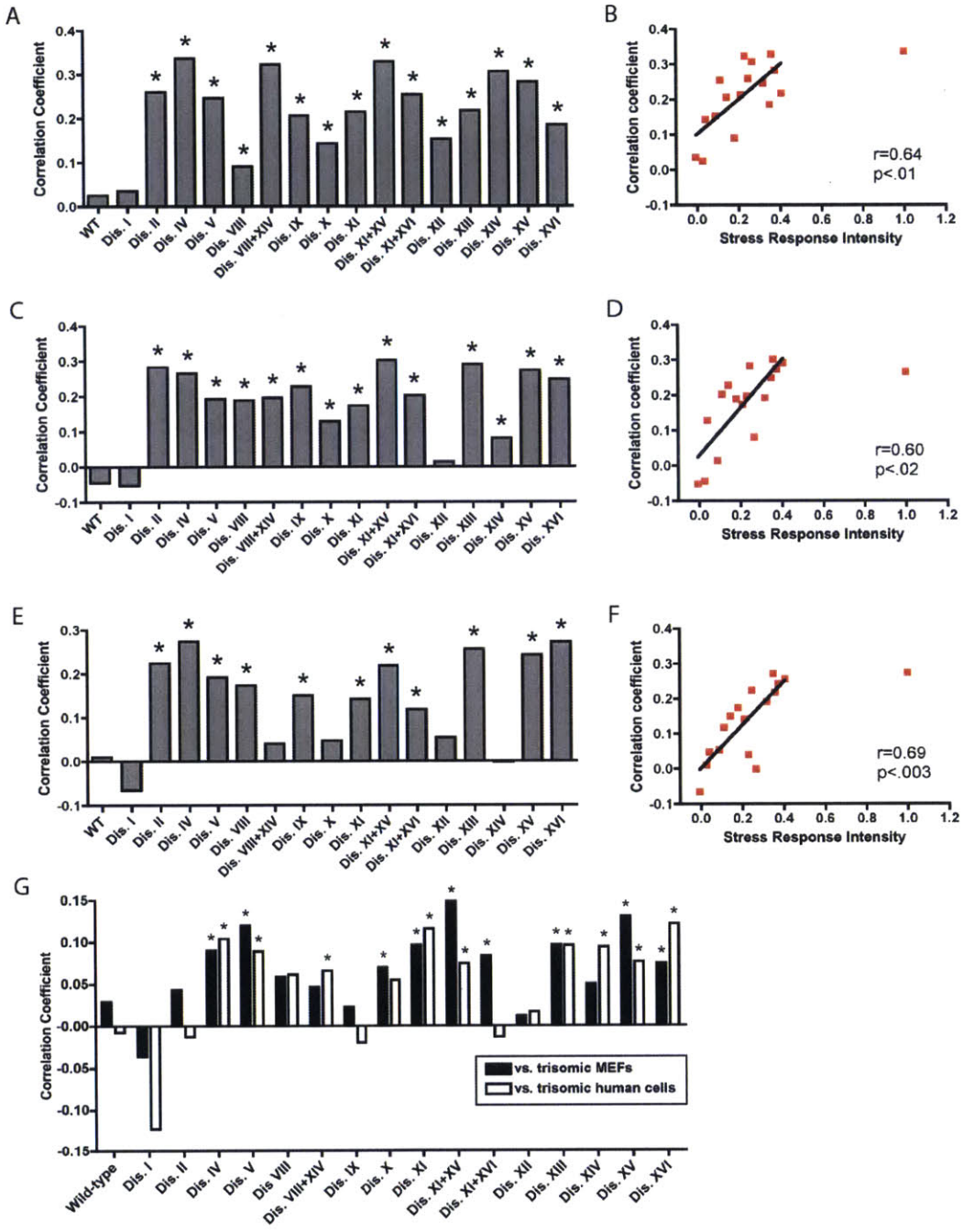


Figure S2. Stress-related transcriptional similarities between aneuploid budding yeast, fission yeast, plants, and mammalian cells. The correlation coefficients between each disomic strain and (A) aneuploid strains from the yeast deletion collection, (C) aneuploid strains of *S. pombe*, (E) trisomic *A. thaliana*, (G) and trisomic mammalian cells are displayed. The stress response intensity of each disomic strain is plotted against its correlation coefficient with

(B) aneuploid strains from the yeast deletion collection, (D) aneuploid strains of *S. pombe*, and (F) trisomic *A. thaliana*. We found that disomic strains that did not exhibit a stress response (e.g., Dis. I and XII) exhibited small or insignificant correlations with aneuploid cells in other species, while disomic strains that exhibited strong stress responses (e.g., Dis. IV and Dis. XIII) tended to exhibit high interspecies correlations. This is consistent with our hypothesis that stress-related transcription underlies the significant genome-wide similarities observed between aneuploid cells in different species. A linear regression excluding disome IV (SRI≈1) is plotted against the data in each graph.

We note, however, that the various deleted genes within the deletion collection are likely to be at least partially responsible for the stress phenotype. Two observations suggest that aneuploidy is also a relevant cause of the similarity with the disomic transcriptomes. First, among individual deletion strains, there was a significant positive correlation between the percent of the genome that was aneuploid and the PCC with the disomes (Fig. S3A; $r=0.54$ $p<.01$). Secondly, we found that the stress response gained intensity as the number of genes on aneuploid chromosomes increased (Fig. S3B; $r=0.49$, $p<.05$). This relationship was true for both aneuploid strains from the deletion collection and disomic strains that we have constructed ($r=0.58$, $p<.0005$). Thus, it is likely that aneuploidy contributes to the similarities in gene expression between these sets of strains.

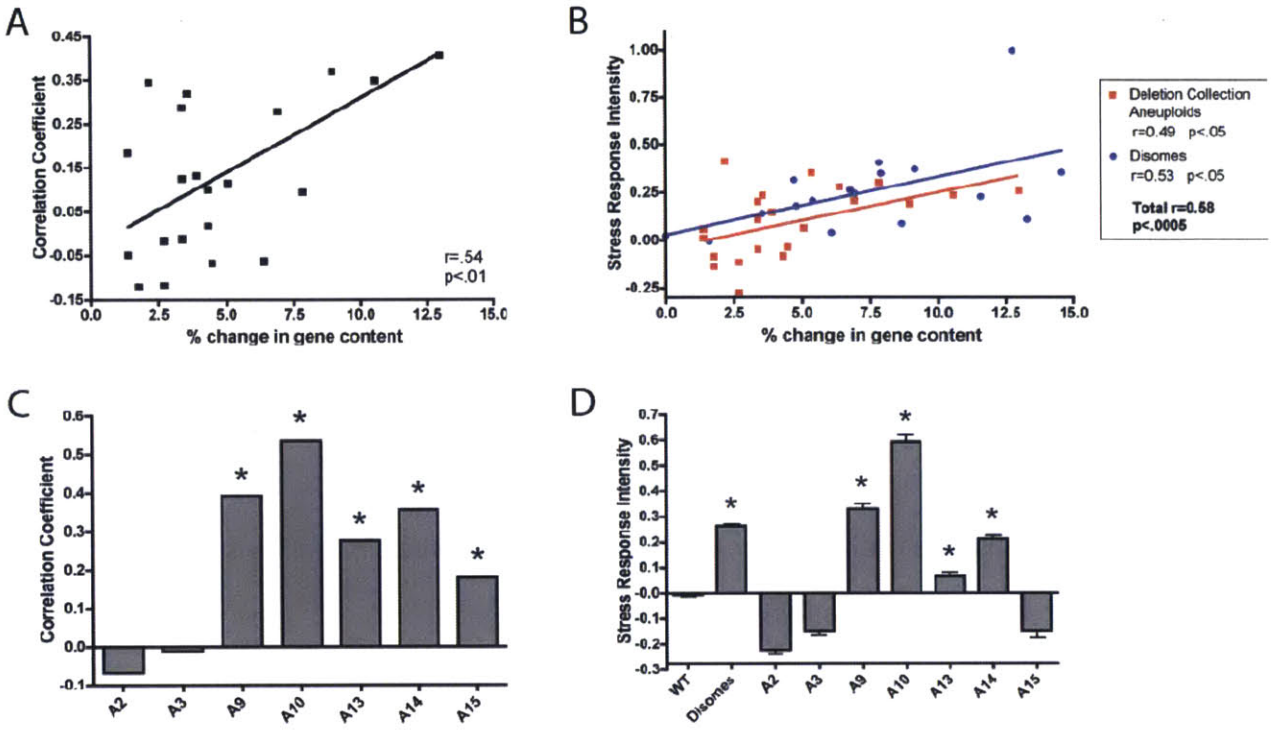


Figure S3. Effects of aneuploidy in deletion collection strains and triploid meiotic products. (A) The correlation coefficients between individual aneuploid deletion collection strains and the disomes were plotted against the percent of the genome that was aneuploid in each strain (see Materials and Methods). A linear regression is plotted against the data. (B) The stress response intensities of disomic strains and aneuploid deletion collection strains were plotted against the percent of the genome that was aneuploid in each strain. Red and blue lines represent linear regressions plotted against the data. (C) The correlation coefficients between the disomic strains and aneuploid products of triploid meiosis were plotted. Karyotypes of these strains are: A2: 1N+II,XII, A3: 1N+I,II,XII, A9: 1N+II,XIII,XVI, A10: 1N+III,XI,XII,XV, A13: 1N+1,II,VIII,XI,XIII, A14: 1N+IX,XVI, A15: 1N+II,III,VII,IX,X,XI,XII. (D) The stress response intensities of aneuploid products of triploid meiosis were calculated. An asterisk indicates a statistically significant increase in the stress response relative to wild-type ($p<10^{-5}$, Student's t test).

We next sought to identify aneuploidy-responsive genes in yeast. We sorted the disomes and aneuploid deletion strains according to the number of genes present on aneuploid chromosomes, and then calculated correlation coefficients between the expression levels of each gene and the degree of aneuploidy across the panel of strains. 446 genes were identified whose expression levels were significantly correlated or anti-correlated with the degree of aneuploidy across all strains (Fig. S4; PCC>0.5 or PCC<-0.5; p<.002). Among transcripts that were positively correlated with increasing aneuploidy, GO term analysis revealed an enrichment of genes related to the response to oxidative stress ($p<10^{-6}$) and protein refolding (Table S1; $p<10^{-3}$). Transcripts that decreased with increasing aneuploidy were enriched for non-coding RNA processing ($p<10^{-11}$) and ribosome biogenesis genes ($p<10^{-6}$). Importantly, there was highly significant overlap among aneuploidy-responsive genes and the ESR ($p<10^{-13}$, hypergeometric test). Furthermore, among the 446 genes, 414 of them exhibited a co-directional change in the slow-growing *cdc* mutants. Discordant transcripts (i.e., those that increased with aneuploidy but were expressed at less than wild-type levels in the *cdc* strains, and vice-versa) were not significantly enriched for any GO terms. Taken together, these data indicate that most (but not all) transcriptional changes caused by aneuploidy are related to stress and/or slow growth, and that increasing degrees of aneuploidy generally exert increasing degrees of stress on cell homeostasis.

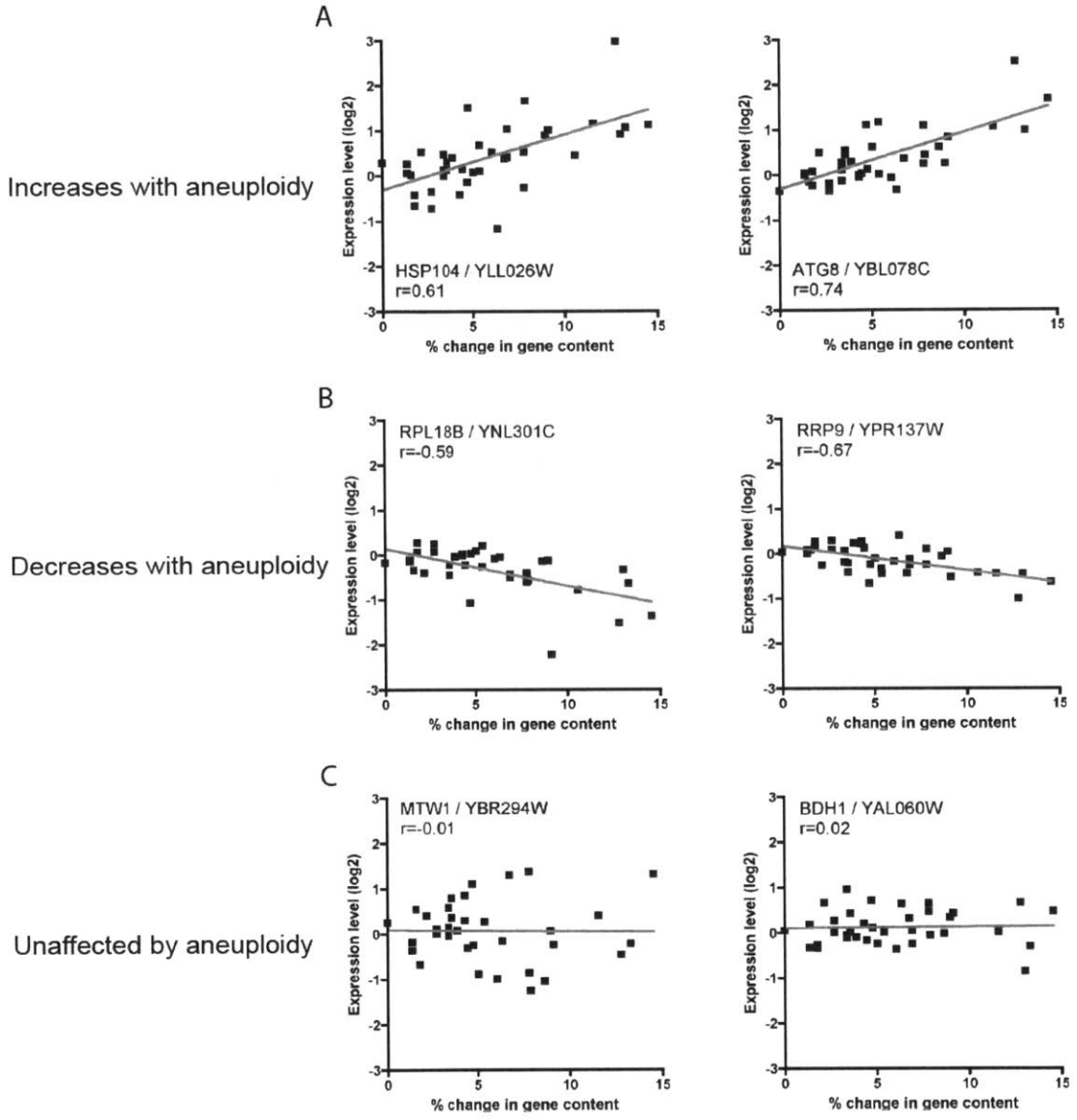


Figure S4. Aneuploidy-responsive genes in *S. cerevisiae*. Yeast disomes and aneuploid strains from the deletion collection were sorted according to the number of genes present on aneuploid chromosomes in each strain, normalized to the baseline ploidy of each strain. Then, a correlation coefficient was calculated between the degree of aneuploidy (measured as the percent change in gene content) and the transcript level of each gene. A cut-off PCC value of ±0.5, corresponding to a p value less than 0.002, was used to identify aneuploidy-responsive genes. GO term enrichment analysis of these gene sets are presented in Table S1, and examples of the different gene classes identified are displayed in this figure. (A) The expression

of *HSP104* (left panel) and *ATG8* (right panel) tend to increase as the degree of aneuploidy increases. (B) The expression of *RPL18B* (left panel) and *RRP9* (right panel) tends to decrease as the degree of aneuploidy increases. (C) The expression of *MTW1* (left panel) and *BDH1* (right panel) are not significantly affected by increasing degrees of aneuploidy.

Next, we obtained gene expression data from seven aneuploid strains that were derived via triploid meiosis (21). The average correlation between these strains and the disomes as well as five out of seven pairwise correlations with the disomes were highly significant (Fig. 1A and S3C). There was a mild anti-correlation between the aneuploid products of triploid meiosis and *cdc15-2* ($r=-0.05$, $p<.001$), but a significant positive correlation with the ESR-exhibiting mutants *cdc28-4* and *cdc23-1* ($r=0.21$, $p<10^{-55}$, and $r=0.30$, $p<10^{-113}$, respectively). Among the five aneuploid strains that were correlated with the disomes, four showed a significant stress/slow growth response relative to a euploid strain (Fig. S3D). We conclude that a shared transcriptional response is a common though not obligate consequence of aneuploidy in yeast, and this response is independent of the mechanism by which aneuploidy is generated.

Aneuploidy causes a stress response in fission yeast

We next sought to determine whether aneuploidy causes a stress response in other organisms. We averaged gene expression data from two aneuploid strains of the fission yeast *Schizosaccharomyces pombe*, then identified upregulated and downregulated genes using a +/- 1.3-fold change (FC) cutoff (23). GO term analysis of upregulated genes revealed that the most enriched functional category was the response to stress (Table S2; $p<10^{-25}$). Downregulated genes included many terms associated with the ribosome, including ribosome biogenesis ($p<10^{-13}$) and the nucleolus ($p<10^{-12}$). Similar GO term enrichments were obtained using Rank Products, a cutoff-independent method of identifying differentially-expressed genes [Table S3; (24)]. We noted that these GO terms are a hallmark of the budding yeast ESR, suggesting that aneuploidy in different yeasts causes a similar stress-related transcriptional response. Indeed, an environmental stress response has also been described in *S. pombe* (25), and out of 236 genes that constitute the fission yeast ESR, 203 genes exhibited co-directional transcriptional changes in the aneuploid *S. pombe* (Fig. S5).

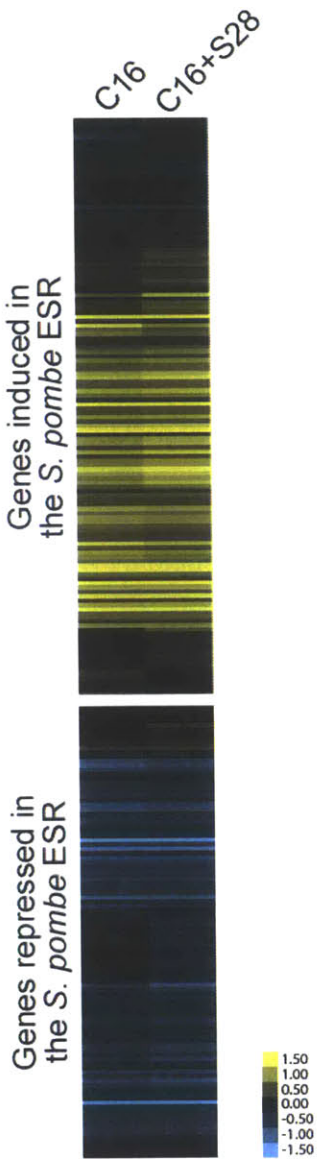


Figure S5. Aneuploid strains of *S. pombe* display a stress response. Genes that constitute the *S. pombe* ESR were downloaded from http://www.bahlerlab.info/projects/stress/ and the expression values of these genes in the two aneuploid strains were clustered. Of 132 genes that are upregulated in the ESR, 106 are upregulated when the aneuploid strains are averaged. Of 104 genes that are downregulated in the ESR, 97 are downregulated when the aneuploid strains are averaged.

To determine whether aneuploidy caused genome-wide similarities in gene expression in different species, we identified one-to-one orthologs between *S. cerevisiae* and *S. pombe* and then calculated the PCC between the averaged aneuploid strains in each organism. The correlation coefficient between disomic budding and fission yeast strains was highly significant (Fig. 2A; $r=0.30$, $p<10^{-47}$). Additionally, there was a weak correlation between aneuploid fission yeast and *cdc15-2* ($r=0.04$, $p<.05$), but stronger correlations with *cdc28-4* and *cdc23-1* ($r=0.22$, $p<10^{-26}$, and $r=0.31$, $p<10^{-52}$). 14 out of 16 individual disomes also exhibited significant pairwise correlations with *S. pombe*, and these transcriptional similarities were particularly striking when genes annotated to GO terms affected by aneuploidy were compared (Fig. 2B and S2C). Moreover, the PCC between individual disomes and *S. pombe* tended to increase based on the intensity of the stress response in each disomic strain ($r=0.60$, $p<.02$; Fig. S2D). Lastly, we sought to determine whether specific groups of genes exhibited coordinate changes in expression in both species. Orthologous genes that were upregulated in both organisms were significantly enriched for those involved in the response to oxidative stress ($p<10^{-6}$) and the response to heat ($p<10^{-3}$), while downregulated genes were enriched for ribosome biogenesis factors ($p<10^{-6}$) and those associated with the nucleolus (Table S4; $p<10^{-3}$). We conclude that aneuploidy in different fungal species induces a highly-related stress response.

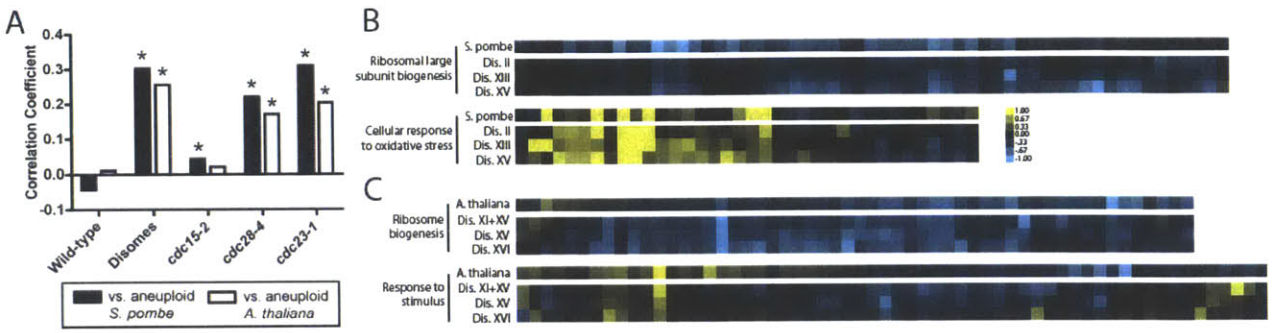


Figure 2. Aneuploidy causes a stress response in *S. pombe* and *A. thaliana*. (A) The correlation coefficients between aneuploid strains of either *S. pombe* (black bars) or *A. thaliana* (white bars) and the indicated strains of *S. cerevisiae* are displayed. (B and C) Heat maps of orthologous genes annotated to aneuploidy-related GO terms from aneuploid strains of (B) *S. pombe* or (C) *A. thaliana* and the indicated disomes.

Aneuploidy causes a stress response in *Arabidopsis thaliana*

Based on the conserved transcriptional response to aneuploidy among different fungi, we hypothesized that aneuploidy in higher organisms could also result in a stress/decreased proliferation response similar to that seen in aneuploid yeast. To test this, we analyzed gene expression data from *Arabidopsis thaliana* plants that were trisomic for chromosome 5 (26). GO term enrichment analysis revealed that many of the same pathways were perturbed by aneuploidy in plants as in yeast (Tables S5 and S6). "Response to chemical stimulus" and "response to stress" were among the most up-regulated GO terms ($p<10^{-24}$ and $p<10^{-19}$, respectively), while the cytosolic ribosome and ribosome biogenesis were highly enriched among downregulated genes ($p<10^{-11}$ and $p<10^{-8}$, respectively). Furthermore, we identified one-to-one orthologs between budding yeast and *A. thaliana*, and found that trisomic plants and disomic yeast exhibited a significant genome-wide transcriptional correlation (Fig. 2A; r=0.26, $p<10^{-5}$). There was no correlation between trisomic plants and *cdc15-2* (r=0.02, p>.05), but significant correlations with *cdc28-4* and *cdc23-1* (r=0.17, p<.002 and r=0.20, p<.0002, respectively) as well as 11 out of 16 individual disomes (Fig. 2C and S2E). As with *S. pombe*, the correlation coefficient between the yeast disomes and trisomic *A. thaliana* increased with the intensity of the stress response in the budding yeast strains (Fig. S2F; r=0.69, p<.003). We conclude that a shared stress response underlies significant transcriptional similarity between aneuploid budding yeast and *A. thaliana*.

Aneuploid mouse and human cells share slow growth-related changes in gene expression

We next analyzed expression data from mouse embryonic fibroblasts (MEFs) trisomic for one of four chromosomes (Chr. 1, 13, 16, and 19) that were normalized to MEFs obtained

from their euploid littermates (15). We first sought to determine whether different trisomies caused similar changes in gene expression. We found highly significant overlap among differentially expressed genes across the trisomies: in 12 out of 12 pairwise comparisons, a gene that was up- or downregulated in one trisomic cell line was significantly more likely to exhibit a similar change in expression in a different trisomy (Fig. 3A and 3B). For instance, ~6% of all genes on euploid chromosomes in trisomy 16 were upregulated at a 1.5-FC cutoff, but among genes that were upregulated in trisomy 19, 20% were also upregulated in trisomy 16 ($p<10^{-25}$, hypergeometric test). Significant similarities were also observed when differentially expressed genes were identified using t-tests or more stringent FC cutoffs (Fig. S6A-S6D). In order to determine whether the same genes were affected across the trisomic cell lines, we applied a permutation test, in which gene expression values were randomized within each trisomy. While 78 genes were upregulated and 168 genes were downregulated in three or more trisomies, no more than 37 and 92 genes were up- or downregulated, respectively, among 100,000 random permutations of the expression data (Fig. S7A and S7B). GO terms enriched among upregulated genes were highly variable and reflected perturbations in many aspects of cell physiology (Tables S7 and S8). Of note, we observed that many upregulated terms were related to stress and inflammation, including the response to wounding ($p<10^{-4}$), the acute inflammatory response ($p<10^{-3}$), and the response to stress ($p<10^{-3}$). The most enriched GO term among upregulated genes was the extracellular region ($p<10^{-12}$), which reflected increased transcript levels of cytokines as well as various matrix-related genes. Downregulated GO terms were more specific: the most downregulated term was cell division ($p<10^{-14}$), and nearly all affected GO terms were directly related to progression through the cell cycle, including mitosis ($p<10^{-13}$), DNA replication ($p<10^{-9}$), and chromosome condensation ($p<10^{-5}$). This is consistent with our previous finding that trisomic MEFs exhibit poor proliferative capacity relative to euploid cells (15). We conclude that trisomic MEFs display some chromosome-independent transcriptional similarities that are indicative of slow growth and cellular stress.

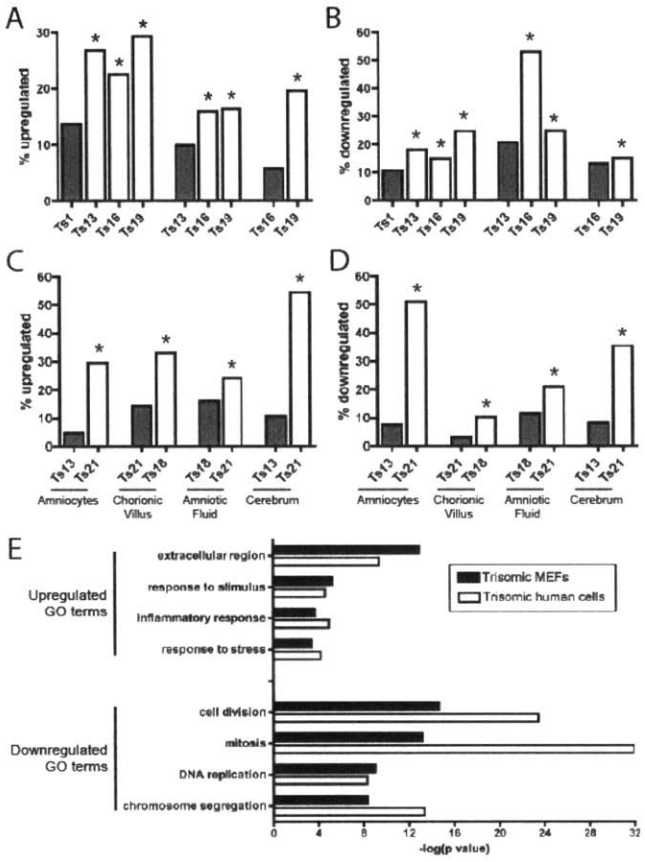


Figure 3. Aneuploidy causes similar transcriptional changes in primary mouse and human cells. (A and B) Genes up- or down-regulated in one trisomic MEF line are significantly more likely to exhibit a similar change in another trisomy. Gray bars indicate the percentage of all genes up- or down-regulated in the indicated cell line at a 1.5-FC cutoff. White bars indicate the percentage of genes up- or down-regulated in that trisomy that are also up- or down-regulated in the trisomy represented with a gray bar. Asterisks indicate statistically significant overlap ($p<.05$, hypergeometric test). (C and D) Genes up- or down-regulated in one trisomic human type are significantly more likely to exhibit a similar change in another trisomy from the same tissue of origin. Data from cultured amniocytes and chorionic villi are from (51). Data from fetal cerebra are from (52). Data from cell-free amniotic fluid are from (31) and (53). (E) GO terms that are enriched among up- and down-regulated genes in trisomic MEFs and human cells. Complete lists are presented in Table S7 and S9.

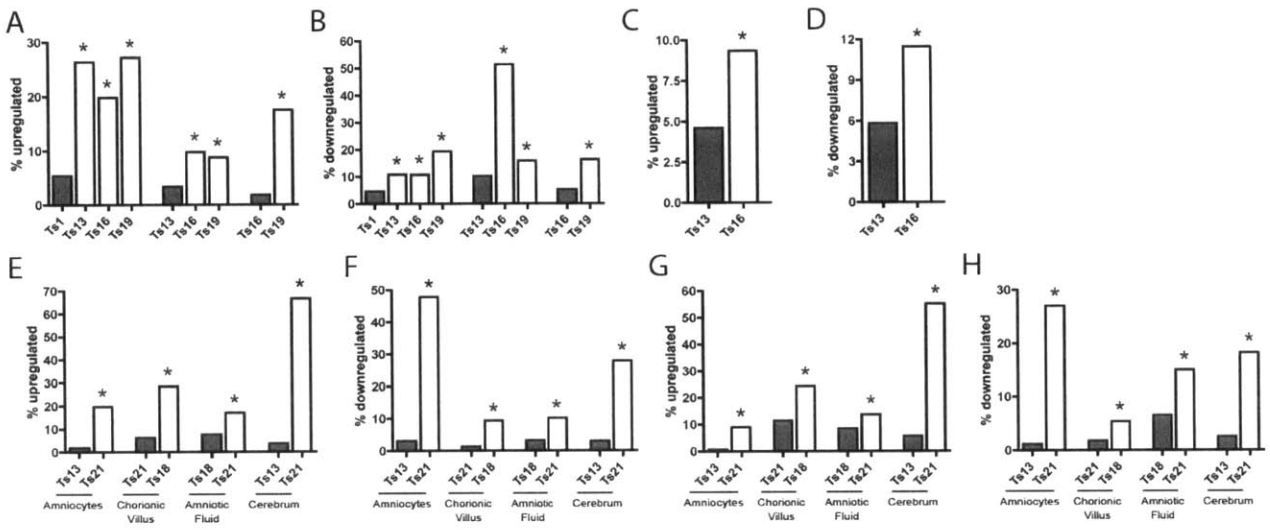


Figure S6. Expression similarity in trisomic MEFs and human cells. (A and B) Genes up- or down-regulated in one trisomic MEF line are significantly more likely to exhibit a similar change in another trisomy. Gray bars indicate the percentage of all genes up- or down-regulated at a 2-FC cutoff in the indicated cell line. White bars indicate the percentage of genes up- or down-regulated in that trisomy that are also up- or down-regulated in the trisomy represented with a gray bar. (C and D) Same analysis as above, except differentially expressed genes were identified among replicates of Ts13 and Ts16 using a 1.5-FC and $p<.05$ (Student's t-test) cutoff. Note that only single replicates of Ts1 and Ts19 were analyzed in ref. (15), precluding the use of a t-test to analyze these cell lines. (E and F) Same analysis described above using data from human trisomies. Differentially expressed genes were identified using a 2-FC cutoff. (G and H) Same analysis described above, except differentially expressed genes were identified using a 1.5-FC and $p<.05$ (Student's t-test) cutoff. Asterisks indicate statistical significance ($p<.05$, hypergeometric test).

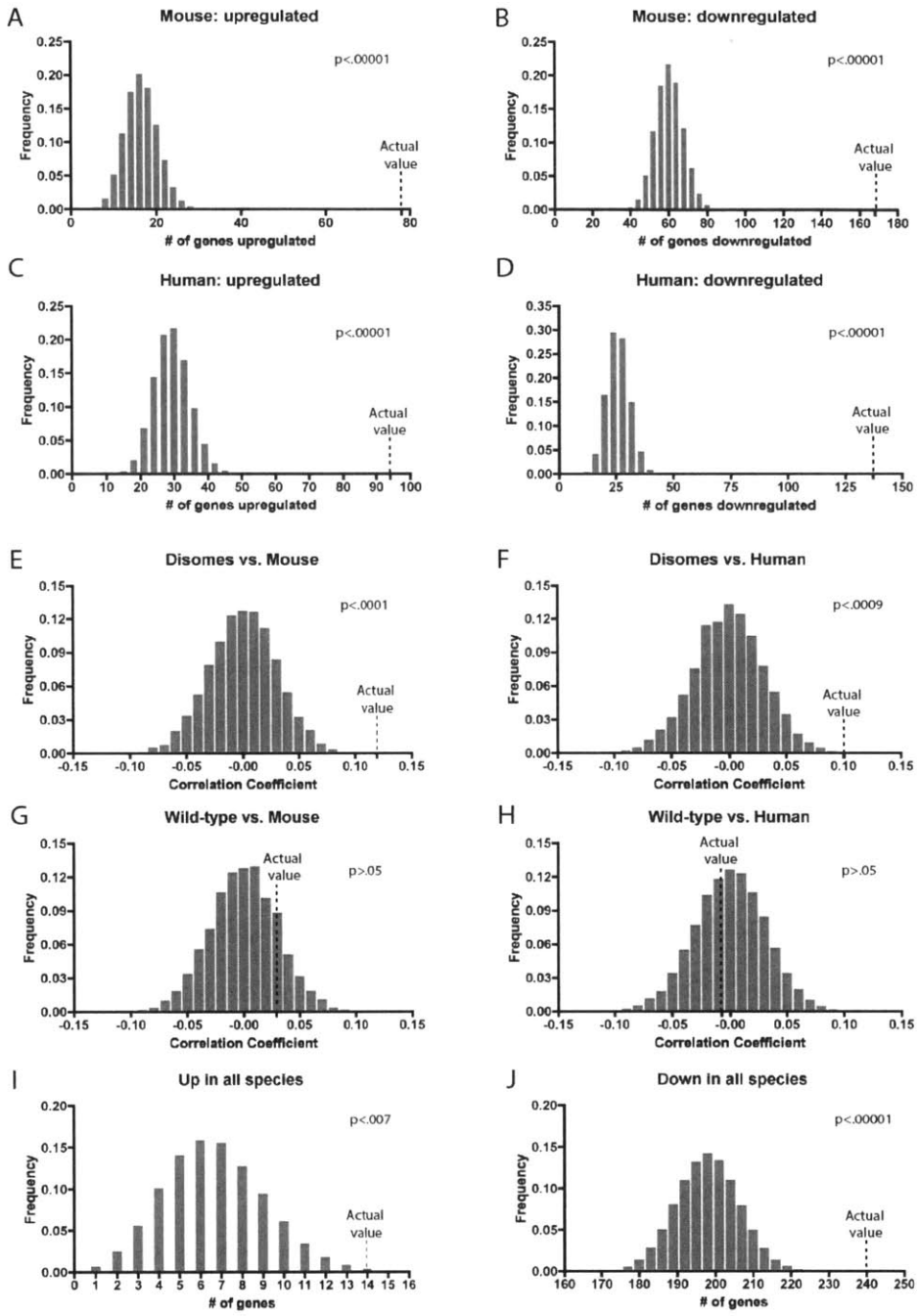


Figure S7. Permutation testing of similarity significance. (A-D) In order to confirm that the number of similarly up- or downregulated genes was significant across mammalian aneuploidies, we randomized gene expression values and then calculated the number of similarly-affected genes which occurred by chance. (A and B) 100,000 random permutations of gene expression data from trisomic MEFs. Bar graphs indicate the number of genes that are

(A) upregulated or (B) downregulated in 3 or more cell lines. (C and D) 100,000 random permutations of gene expression data from human trisomies. Bar graphs indicate the number of genes that are (C) upregulated or (D) downregulated in 4 or more sample types. (E-H) In order to confirm the significance of Pearson correlation comparisons between mammalian cells and other eukaryotes, we randomized gene expression values and then calculated the PCC for each randomized sample. (E and F) 10,000 random permutations of disomic yeast gene expression values compared to (E) mouse and (F) human values. (G and H) 10,000 random permutations of wild-type yeast gene expression values compared to (G) mouse and (H) human values. (I and J) In order to confirm the significance of the co-directional changes among 1:1 orthologs observed in aneuploid cells, we randomized gene expression values and calculated the number of co-directional changes observed in each randomized sample. The number of (I) up-regulated and (J) down-regulated genes observed among 100,000 random permutation are displayed.

Does a chromosome-independent response to aneuploidy exist in humans as well? To test this, we examined gene expression data from four datasets that included Trisomy 13, 18, and 21. Within each sample, we asked whether genes that are upregulated or downregulated in one trisomy were more likely to be up- or down-regulated in another trisomy. In 8 out of 8 pair-wise comparisons, the overlap between different trisomies was significantly more than expected by chance (Fig. 3C, 3D, and S6E-S6H). In order to determine whether similar genes were affected across datasets, we performed a permutation test. While 94 and 137 genes were up- or downregulated, respectively, in 4 or more trisomic samples, no more than 59 and 49 genes were up- or downregulated, respectively, among 100,000 random permutations of the expression data (Fig. S7C and S7D). Surprisingly, dysregulated genes in human trisomies were enriched for many of the same GO terms as were found in the trisomic MEFs (Fig. 3E; Tables S9 and S10). The extracellular region ($p<10^{-9}$), inflammatory response ($p<10^{-4}$), and response to stress ($p<10^{-4}$) were significantly enriched among upregulated genes while mitosis ($p<10^{-13}$), DNA replication ($p<10^{-8}$), and chromosome condensation ($p<10^{-6}$) were enriched among downregulated genes. Out of 97 GO terms enriched among downregulated genes in trisomic MEFs, 64 of them were also enriched among downregulated genes in trisomic human cells, including all 29 of the most enriched terms (Table S7).

Analysis of individual datasets further clarified the origins of the transcriptional changes caused by aneuploidy in humans. First, Trisomy 13 and Trisomy 18 were primarily responsible for the observed enrichment of stress- and cell cycle-related GO terms (compare Table S11 and S12). As these chromosomes are both larger than chromosome 21, this is consistent with our hypothesis that the degree of aneuploidy determines the severity of the transcriptional response. Secondly, the origin and/or culturing of trisomic tissue affected the penetrance of the stress signature. Cultured Trisomy 13, Trisomy 18, and, to a lesser extent, Trisomy 21 samples displayed enrichments of stress- and cell cycle-related GO terms, while those GO terms were

not enriched among differentially expressed genes when only data from fetal cerebra and amniotic fluid were considered (Table S13). Natural limits on cell division that exist *in utero* may partially mask the different proliferative capacities of trisomic and euploid cells, while unconstrained growth in culture highlights this disparity. For this reason, we utilized gene expression data from cultured trisomic human cells for subsequent comparisons.

The similarity between enriched GO terms in trisomic human and mouse cells suggested that aneuploidy causes a conserved transcriptional response across mammals. To test this, we identified one-to-one orthologs between humans and mice, then calculated the correlation between the average gene expression values from trisomic MEFs and cultured trisomic human cells. The PCC across all genes was moderate but highly significant ($r=0.11$, $p<10^{-20}$). Additionally, we noted significant overlap between the sets of differentially-expressed genes in trisomic mouse and human cells ($p<10^{-17}$, hypergeometric test) which was particularly evident among cell cycle transcripts (Table S14). Thus, aneuploidy induces a similar gene expression pattern indicative of slow growth and/or cellular stress in both mouse and human cells.

Stress-related transcriptional similarities across all aneuploid cell types

The common stress response in aneuploid cells of highly divergent species raised the possibility that yeast and mammalian cells share a transcriptional response to aneuploidy. To test this, we identified one-to-one orthologs between yeast, plants, mice, and humans. We found that disomic yeast exhibited a small but statistically significant correlation with the averaged expression values of trisomic MEFs and of cultured trisomic human cells (Fig. 4A; $r=0.12$, $p<10^{-4}$, and $r=0.10$, $p<.002$, respectively). The significance of these correlations was also confirmed by permutation testing (Fig. S7E-S7H). A majority of individual disomes also exhibited significant pairwise correlations with the trisomic mammalian cells (Fig. S2G).

Furthermore, we observed significant correlations between the aneuploid strains of *S. pombe* and *A. thaliana* and the trisomic MEFs as well as between *A. thaliana* and the trisomic human cells (Fig. 4A).

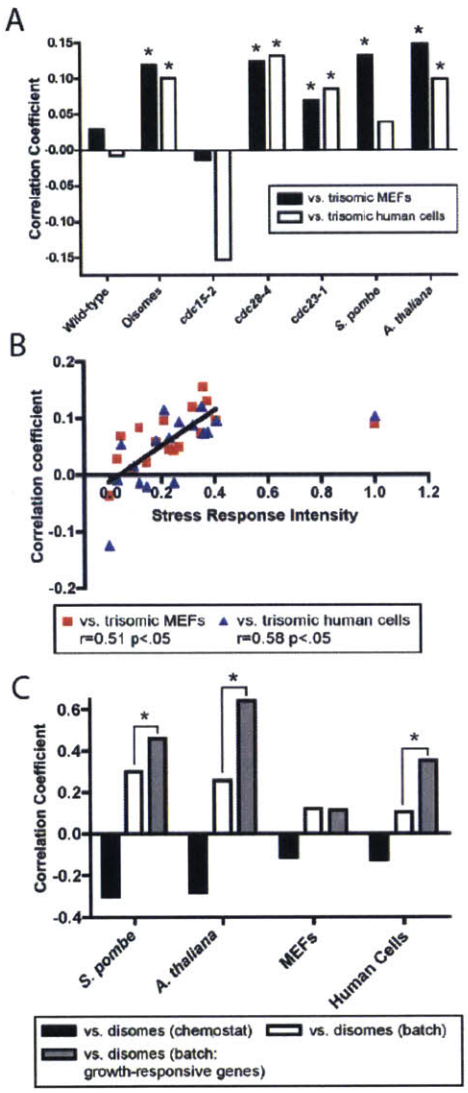


Figure 4. Transcriptional similarities among all aneuploid cell types. (A) Correlation coefficients between the indicated cell type and either trisomic MEFs (black bars) or cultured trisomic human cells (white bars) are displayed. Asterisks indicate a statistically significant correlation ($p<.05$). (B) The stress response intensity of the disomic strains is plotted against the pairwise correlations with the trisomic mouse and human cells. The black line is a linear regression plotted against the data excluding disome IV (SRI≈1). (C) Correlation coefficients between the indicated aneuploid cell types and chemostat-grown disomes (black bars), batch-grown disomes (white bars), and 500 growth responsive genes in batch grown disomes (gray bars) are displayed.

We hypothesized that the similarities in gene expression between the aneuploid transcriptomes were a consequence of the slow growth/stress response that was common among aneuploid cells of all species. Consistent with this hypothesis, there was a significant positive correlation between the stress response intensities in each disome and their PCC with the trisomic mammalian cells (Fig. 4B). Moreover, *cdc28-4* and *cdc23-1* exhibited significant positive correlations with both trisomic mammalian cell types, while *cdc15-2* did not (Fig. 4A). Taken together, these findings suggested that conserved aspects of the transcriptional response to stress and/or slow growth, rather than aneuploidy *per se*, drives the expression correlations between aneuploid cell types. To test this, we compared gene expression data from aneuploid cells to chemostat-grown yeast disomes. In chemostats, the doubling times between disomic and euploid cells were equalized by nutrient titration, thereby masking the slow growth/stress response. Indeed, for each interspecies comparison, aneuploid cells were anti-correlated with chemostat-grown disomes (Fig. 4C). Next, we examined the set of yeast genes whose expression levels directly vary according to the rate of cell division (19). In batch culture, the correlation coefficients between the disomes and other species were significantly increased in 3 out of 4 comparisons when only orthologs of the 500 strongest growth-responsive genes in yeast were compared. These results demonstrate that conserved elements of a transcriptional stress or slow growth program in eukaryotes underlie the significant transcriptional similarities between aneuploid yeast, plant, and mammalian cells. The lesser predictive value of the growth-responsive yeast genes in comparisons with mammalian cells likely reflects the fact that ribosome synthesis is strongly downregulated by slow proliferation in yeast, while the most striking transcriptional changes among aneuploid mammalian cells were the down-regulation of cell cycle transcripts (Table S14, S15 and Fig. S8).

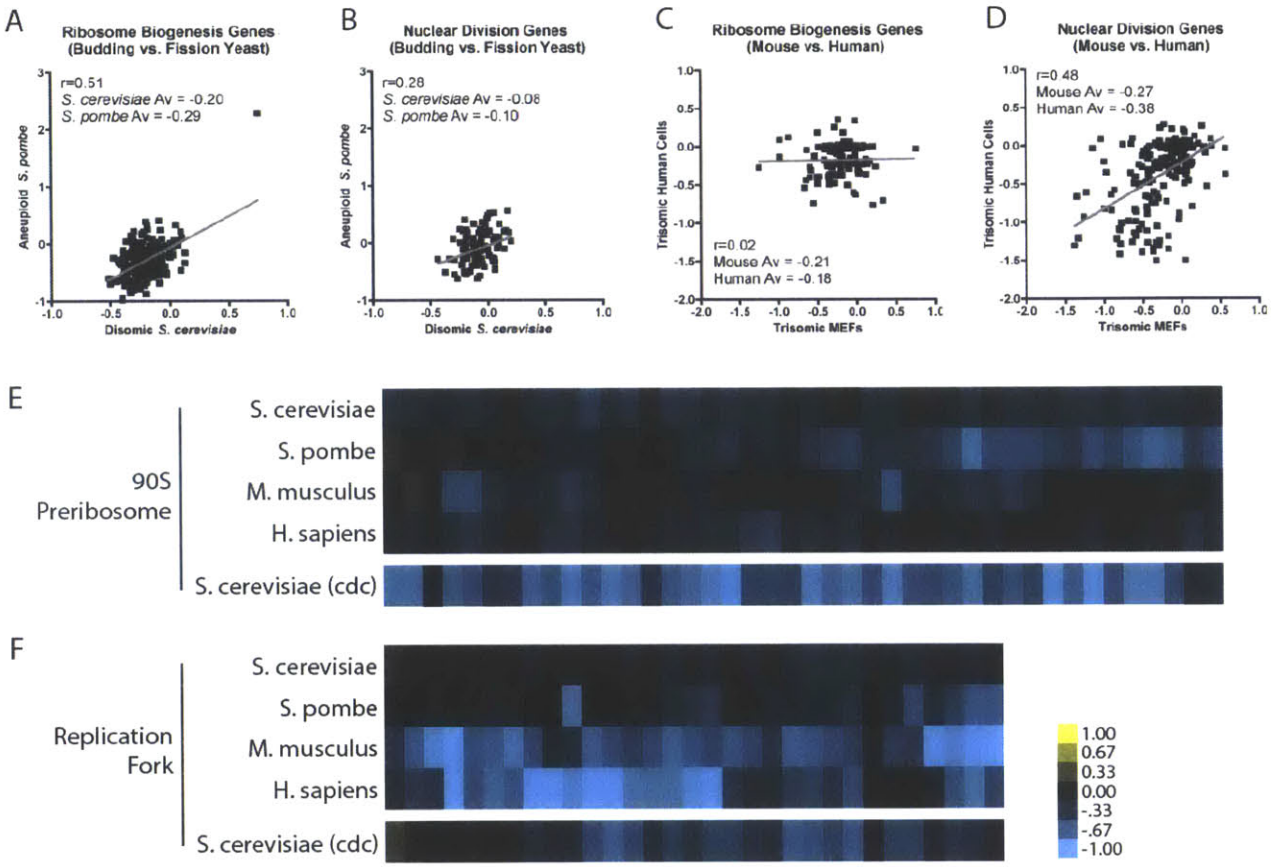


Figure S8. Cellular processes perturbed across species by stress and aneuploidy.

Scatterplots of genes annotated to the indicated GO terms comparing expression levels in either (A and B) aneuploid budding and fission yeast or (C and D) trisomic mouse and human cells. While the average transcripts in each category are expressed at less than euploid levels, cell cycle transcripts (e.g., nuclear division genes) are expressed at lower levels in mammalian cells, while ribosomal and translational transcripts (e.g., ribosome biogenesis genes) are expressed at lower levels in fungi. Additionally, ribosome biogenesis transcript levels in trisomic MEFs and human cells are uncorrelated with one another, while nuclear division genes [and cell cycle genes generally (Table S14)] show a strong correlation between species. (E and F) Heat maps of single-ortholog genes in aneuploid budding yeast, fission yeast, mouse, and human cells. Genes that are downregulated by aneuploidy are also downregulated in ESR-exhibiting yeast *cdc* mutants (the average of *cdc28-4* and *cdc23-1*, bottom row).

A second prediction derived from these results is that aneuploid cells should exhibit similar gene expression changes to euploid cells experiencing exogenous stresses. Aneuploidy in budding and fission yeast affects known stress-response genes in those organisms (Fig. S1 and S5); therefore we examined stress-induced transcription in *A. thaliana* and mammalian cells. We analyzed gene expression data from *A. thaliana* grown under salt, ROS, or drought stress (27), and from MEFs treated with salt, the protein synthesis inhibitor anisomycin, or TNFα (28). The transcriptomes of stressed plants and MEFs were significantly correlated with the corresponding trisomic tissue (Fig. S9; $r=0.40$, $p<10^{-300}$, and $r=0.11$, $p<10^{-16}$, respectively). Furthermore, the PCC values were significantly higher when we examined only genes that changed co-directionally by a certain threshold under all stress conditions. As expected, many GO terms that were enriched among up-regulated genes in both stressed and trisomic tissue were stress-related, including the response to chemical stimulus ($p<10^{-26}$) and the defense response ($p<10^{-12}$) in Arabidopsis and the inflammatory response ($p<10^{-4}$) and cytokine activity in MEFs (Table S16 and S17; $p<10^{-4}$).

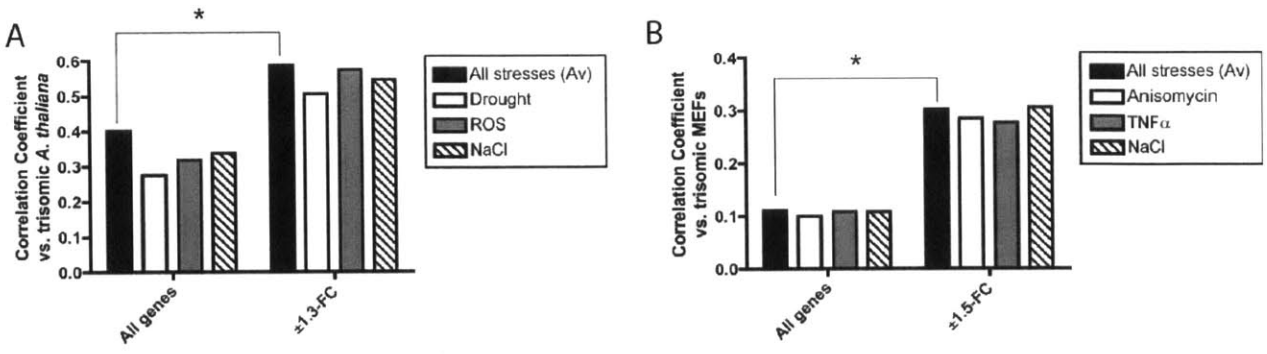


Figure S9. Similarities between stressed and aneuploid *A. thaliana* and MEFs. (A and B) The Pearson correlation coefficient between (A) trisomic and stressed *A thaliana* and (B) trisomic and stressed MEFs are displayed. When "stress response" genes were defined in these species by examining only those genes that changed +/- a certain threshold under all stress conditions, the correlation coefficients significantly increased ($p<.05$).

Lastly, we sought to identify conserved changes in expression in response to aneuploidy among single-ortholog genes across species. We used relatively permissive criteria, and asked which genes changed co-directionally in aneuploid budding yeast, fission yeast, mouse, and human cells. We found that 254 single-ortholog genes changed co-directionally in all four species (240 down, and 14 up), significantly more than expected by chance ($p<10^{-5}$ and $p<.007$, respectively, Fig. S7I and S7J). The most affected GO terms among those genes were the nucleus ($p<10^{-17}$) and nucleic acid metabolism (Table S18; $p<10^{-14}$). These terms reflected the downregulation of ribosomal genes (primarily in fungi) and cell cycle-related genes (primarily in mammalian cells; Fig. S8 and Table S18). Out of the 254 genes that exhibited co-directional changes across species, 230 of the genes exhibited a similar change in the yeast *cdc* mutants. Among discordant genes, no significant GO term enrichments were observed. We conclude that aneuploidy in different cell types induces a conserved transcriptional program that is also elicited by exogenous stress and/or slow growth.

Discussion

We have identified a stress/slow growth-related transcriptional signature that is present in aneuploid cells of diverse organisms and is largely independent of the identity of the extra chromosome(s). Previous analyses have described an oxidative stress response and the downregulation of proliferation-related genes in human samples and mouse models of Down syndrome (29–32). The data presented here suggest that these phenotypes and others found in aneuploid cells may be a common consequence of aneuploidy, as eukaryotic cells appear to exhibit a stereotypical transcriptional response to chromosome-wide gene dosage changes.

Why is aneuploidy associated with a stress response? First, aneuploidy increases a cell's energy needs. This may result from the wasteful transcription, translation, and degradation of

proteins encoded by extra chromosomes, and is evidenced by the decreased efficiency at which aneuploid cells convert nutrients into biomass (13,15). Altered metabolism may also increase the production of reactive oxygen species (33), and ROS-related GO terms were commonly upregulated in aneuploid cells. Secondly, protein folding and turnover pathways are burdened by aneuploidy. Over-expression of certain proteins may saturate key chaperones, prohibiting them from folding client proteins whose functions are required for viability. Proteins that escape proper folding or degradation may also form cytotoxic aggregates (34,35). Lastly, while many aneuploidy-induced phenotypes appear to be independent of the identity of the extra chromosome, copy number changes of a few particularly dosage-sensitive genes may have direct consequences. For instance, budding yeast cells are exquisitely sensitive to tubulin levels, and a single extra copy of beta-tubulin causes the lethality of disome VI (13,36,37). We posit that these factors contribute to limit the proliferative capacity of aneuploid cells, thereby resulting in the common downregulation of cell cycle and ribosomal genes.

It is interesting to note that euploid yeast strains that displayed a basal stress response (*cdc23-1* and *cdc28-4*) exhibited significant correlations with aneuploid cells in every organism, and the intensity of the ESR in disomic yeast predicted the strength of their transcriptional similarity with aneuploid cells in other organisms. The ESR was first described as a common transcriptional signature in yeast cells treated with multiple independent stresses, though later research demonstrated that it could also result from a slowed rate of cell division (18–20). Whether the ESR-like transcriptional changes observed in aneuploid higher eukaryotes result from stress, or whether they are also a byproduct of differences in growth rate, remains to be tested.

Not every aneuploid strain that we examined displayed a significant stress response. In many cases, this can be explained by threshold effects of dosage imbalance, as the degree of aneuploidy is proportional to the intensity of the transcriptional response (Fig. S3B). Still, some

outliers fail to follow this overall trend: while chromosome XII contains the third most ORF's of any yeast chromosome, disome XII displays the third lowest stress response of any disomic strain (Fig. S1). An extra copy of chromosome XII is also found in strains A2 and A3, the two products of triploid meioses that were uncorrelated with the gene expression pattern in the disomic yeast strains. Interestingly, strains A2 and A3 do exhibit significant PCC's with disome XII ($r=0.30$, $p<10^{-92}$, and $r=0.24$, $p<10^{-59}$, respectively), but they do not exhibit a PCC>0.2 with any other disome. Thus, a gene or genes present on chromosome XII may modulate the effects of aneuploidy on transcription. Discovering what underlies the differences between aneuploid cells that do and do not exhibit stress responses may shed further light on the interplay between gene dosage alterations and cellular phenotype.

Among higher eukaryotes, the shared transcriptional response of primary cells to aneuploidy has relevance for the study of cancer. While 90% of solid tumors display whole-chromosome aneuploidy, it is not clear what role aneuploidy plays in transformation and cancer progression. In mouse models, chromosomal instability (CIN) can both instigate and inhibit tumorigenesis (33,38–41). In human patients, CIN in tumor cells is generally associated with aggressive disease (42), although in some contexts high levels of CIN actually correlate with improved prognosis (43). We have found that single-chromosome aneuploidy causes a transcriptional response indicative of increased cellular stress and decreased proliferative capacity. These results present aneuploidy as a complex phenomenon with potentially anti-tumorigenic properties. While aneuploidy can contribute to transformation by altering the dosage of oncogenes and tumor suppressors, the stresses induced by aneuploidy on metabolism and protein folding limit growth potential. Yeast can adapt to aneuploidy by developing mutations which improve their proliferative capacity (14), and it may be the case that cancer cells must develop similar genetic changes which allow them to tolerate aneuploidy and acquire robust proliferative capacity as well.

Materials and Methods

Yeast expression data. Gene expression data from disomic and *cdc* mutant yeast strains were downloaded from (13). These data include gene expression values from yeast grown in batch culture, in which cells were allowed to reach logarithmic growth. Under these conditions, the yeast disomes display an ESR. This dataset also includes expression values from yeast grown under phosphate limitation in chemostats. Under this growth regime, the doubling time between euploid and disomic strains was equalized, thereby masking the ESR.

Gene expression data for strains derived via triploid meiosis were acquired as previously described and were deposited in the Gene Expression Omnibus (accession number GSE35853) (13). In strain A15, we detected extra copies of chromosome X (17), which was not present in the published karyotype. Gene expression data from 23 aneuploid strains from the deletion collection were downloaded from http://hugheslab.ccbr.utoronto.ca/supplementary-data/rii/. A few of these strains are believed to have developed aneuploidies because the deletion directly affects the spindle assembly checkpoint or chromosome segregation. Other strains gained chromosomes that carry a paralog of the deleted gene. However, in the majority of cases, the causes of aneuploidy in these strains are unknown (22). The set of genes that constitute the budding yeast ESR were downloaded from http://genome-www.stanford.edu/yeast_stress/. Growth rate-responsive genes in yeast were acquired from (19).

For analyses involving *S. cerevisiae*, ORFs annotated as "dubious" and ORFs that were detected in fewer than 10 disomic strains were removed from consideration. Additionally, YAR015W and YBR115C were excluded, as they were the sites of marker integration in the euploid strain, and YDR342C and YDR343C were excluded, as they were amplified at the DNA level in aneuploid strains (13). 3 wild-type vs. wild-type replicates were used as a control. The percent of the genome that was aneuploid was calculated as the sum of the number of ORFs

present on a chromosome that was gained or lost, divided by the total number of ORFs in a euploid cell of that ploidy. The intensity of the stress response in yeast was calculated as follows: for all genes *x* which are up-regulated in the ESR and all genes *y* which are down-regulated in the ESR, the stress response intensity (SRI) of a given strain was calculated as

$$SRI = \frac{\sum_{i=1}^{m} x_i + \sum_{j=1}^{n} -y_j}{m+n}$$

As the expression levels of most ESR genes change in response to growth rate, stress response intensities are not bimodal, but instead vary along a continuum. Strains were considered to exhibit an ESR if their SRI was positive and their SRI significantly differed ($p<10^{-5}$, Student's t test) from that of an isogenic wild-type strain.

Gene expression data from aneuploid strains of fission yeast were downloaded from GEO (accession number GSE8782). Strain C16, containing one additional copy of 163 ORFs, and strain Ch16+S28, containing two additional copies of the same 163 ORFs and one additional copy of 63 ORFs, were used for analysis. As these two strains contain very similar duplicated regions, they were averaged for subsequent comparisons. *S. pombe* genes involved in the fission yeast ESR were downloaded from http://www.bahlerlab.info/projects/stress/.

***A. thaliana* expression data.** Gene expression data for plants trisomic for chromosome 5 were downloaded from http://bioinf.boku.ac.at/pub/trisomy2008/. Log2-FC values were calculated between aneuploid and euploid samples for all genes not on chromosome 5. Gene expression data from stressed plants were downloaded from http://www.weigelworld.org/resources/microarray/AtGenExpress/. Log2-FC values were calculated between treated and mock-treated samples at the 24h time point.

Mammalian expression data. Gene expression data from trisomic mouse cell lines were downloaded from (15). Probes were updated using Release 32 of the NetAffx probe annotations, and the set of probes classified as "expressed" were utilized in this study. Gene expression data from stressed mouse cells were downloaded from GEO (accession number GSE18320). Gene expression data from human aneuploidies were downloaded from GEO (accession numbers GSE6283, GSE1397, GSE16176, and GSE25634). For all mammalian aneuploidies, non-specific probesets and probesets mapping to the X or Y chromosome were excluded from consideration. Probesets that mapped to the same gene were collapsed by averaging. Within each dataset, log2-FC values were calculated between aneuploid and euploid samples.

Data analysis. Gene expression data were analyzed in Excel, MATLAB, and Python using custom scripts. For all correlative studies, only gene expression values from euploid autosomes were considered, and multiple replicates of individual strains were averaged for comparisons. Aneuploid chromosomes were excluded due to the possibility that dosage compensation mechanisms decreased the transcription of a select number of genes present on extra chromosomes, which would introduce biases into subsequent analyses. Correlation values reported in the text represent the Pearson coefficient between two samples. Every comparison found to be significant via the Pearson method was also found to be significant when the Spearman coefficient was calculated instead.

GO term enrichment analysis was performed using GProfiler with a Benjamini-Hochberg corrected p value of .05 and a maximum p value of 10^{-3} (44). Enrichments were performed against the relevant background gene set, e.g., against all genes not on an aneuploid chromosome or against the set of one-to-one orthologs between species. Consistent with our previous methodology (13,14), differentially expressed genes were identified using a +/- 1.3-FC

cutoff in yeast and plants. We utilized a +/- 1.5-FC cutoff in multiple samples for mammalian expression data, a similarly low-stringency threshold that allowed us to detect expression changes in datasets that contained variable numbers of replicates. As a secondary method, we used the rank products algorithm, a FC cutoff-independent protocol that is particularly useful in identifying differentially expressed genes in datasets with small sample sizes (24,45,46). The Rank Products algorithm was implemented in TM4 using a $p<.05$ significance threshold for yeast and plant data and a $p<.005$ threshold for mammalian data (47). P values reported in the text are from GO term analysis using a FC cutoff.

For budding yeast, mice, and humans, orthologous genes were identified using GProfiler (44). For *Arabidopsis* and fission yeast, orthologous genes were identified using InParanoid (48). All interspecies comparisons were between genes in which a one-to-one orthology relationship existed. Clustering was performed in Gene Cluster 3.0 (49) and visualized in Java TreeView (50). For clustering and visualization, the expression values of genes present on aneuploid chromosomes were replaced by the average value of that gene on euploid chromosomes in other strains.

Permutation tests to confirm the significance of transcriptional similarities were performed in Python. For each strain or species, blank cells were fixed in place. Blank cells usually resulted from the exclusion of an aneuploid chromosome, and these cells were locked so as to keep the number of relevant comparisons constant. Next, non-empty cells were randomly shuffled, and after each shuffle the relevant parameter was scored. P values reported for permutation tests represent maximum probabilities based on the number of permutations performed.

Acknowledgments

We thank Joan Clara Smith and Emily Mitchell for technical assistance, Rong Li for providing strains used in this analysis, and Mike Laub, Aviv Regev, and members of the Amon and Gilbert labs for reading this manuscript. This work was supported by the National Institutes of Health (GM056800 to A.A. and GM094306 to M.J.D.) and an NSF Predoctoral Fellowship (to J.M.S.). M.J.D. is a Rita Allen Foundation Scholar. A.A. is an HHMI investigator.

References

1. Musacchio A, Salmon ED. The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol. 2007;8:379–93.

2. Weaver BA, Cleveland DW. Does aneuploidy cause cancer? Curr Opin Cell Biol. 2006;18:658–67.

3. Hassold T, Abruzzo M, Adkins K, Griffin D, Merrill M, Millie E, et al. Human aneuploidy: Incidence, origin, and etiology. Environ Mol Mutagen. 1996;28:167–75.

4. Brown S. Miscarriage and Its Associations. Semin Reprod Med. 2008;26:391–400.

5. Delabar JM, Theophile D, Rahmani Z, Chettouh Z, Blouin JL, Prieur M, et al. Molecular mapping of twenty-four features of Down syndrome on chromosome 21. Eur J Hum Genet EJHG. 1993;1:114–24.

6. Korenberg JR, Chen XN, Schipper R, Sun Z, Gonsky R, Gerwehr S, et al. Down syndrome phenotypes: the consequences of chromosomal imbalance. Proc Natl Acad Sci USA. 1994;91:4997–5001.

7. Olson LE, Richtsmeier JT, Leszl J, Reeves RH. A Chromosome 21 Critical Region Does Not Cause Specific Down Syndrome Phenotypes. Science. 2004;306:687–90.

8. Olson LE, Roper RJ, Sengstaken CL, Peterson EA, Aquino V, Galdzicki Z, et al. Trisomy for the Down syndrome "critical region" is necessary but not sufficient for brain phenotypes of trisomic mice. Hum Mol Genet. 2007;16:774–82.

9. Korbel JO, Tirosh-Wagner T, Urban AE, Chen X-N, Kasowski M, Dai L, et al. The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. Proc Natl Acad Sci U S A. 2009;106:12031–6.

10. Roper RJ, Reeves RH. Understanding the Basis for Down Syndrome Phenotypes. PLoS Genet. 2006;2.

11. Torres EM, Williams BR, Amon A. Aneuploidy: Cells Losing Their Balance. Genetics. 2008;179:737–46.

12. Zhang F, Gu W, Hurles ME, Lupski JR. Copy Number Variation in Human Health, Disease, and Evolution. Annu Rev Genomics Hum Genet. 2009;10:451–81.

13. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, et al. Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. Science. 2007;317:916–24.

14. Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, et al. Identification of aneuploidy-tolerating mutations. Cell. 2010;143:71–83.

15. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, et al. Aneuploidy Affects Proliferation and Spontaneous Immortalization in Mammalian Cells. Science. 2008;322:703–9.

16. Tang Y-C, Williams BR, Siegel JJ, Amon A. Identification of Aneuploidy-Selective Antiproliferation Compounds. Cell. 2011;144:499–512.

17. Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, et al. Aneuploidy Drives Genomic Instability in Yeast. Science. 2011;333:1026–30.

18. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Mol Biol Cell. 2000;11:4241–57.

19. Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, et al. Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast. Mol Biol Cell. 2008;19:352–67.

20. Regenberg B, Grotkjær T, Winther O, Fausbøll A, Åkesson M, Bro C, et al. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. Genome Biol. 2006;7:R107.

21. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, et al. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature. 2010;468:321–5.

22. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, et al. Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet. 2000;25:333–7.

23. Chikashige Y, Tsutsumi C, Okamasa K, Yamane M, Nakayama J, Niwa O, et al. Gene expression and distribution of Swi6 in partial aneuploids of the fission yeast Schizosaccharomyces pombe. Cell Struct Funct. 2007;32:149–61.

24. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 2004;573:83–92.

25. Chen D, Toone WM, Mata J, Lyne R, Burns G, Kivinen K, et al. Global Transcriptional Responses of Fission Yeast to Environmental Stress. Mol Biol Cell. 2003;14:214–29.

26. Huettel B, Kreil DP, Matzke M, Matzke AJM. Effects of Aneuploidy on Genome Structure, Expression, and Interphase Organization in Arabidopsis thaliana. PLoS Genet. 2008;4.

27. Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, et al. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J. 2007;50:347–63.

28. Ferreiro I, Joaquin M, Islam A, Gomez-Lopez G, Barragan M, Lombardia L, et al. Whole genome analysis of p38 SAPK-mediated gene expression upon stress. BMC Genomics. 2010;11:144.

29. Moldrich RX, Dauphinot L, Laffaire J, Vitalis T, Hérault Y, Beart PM, et al. Proliferation deficits and gene expression dysregulation in Down's syndrome (Ts1Cje) neural progenitor cells cultured from neurospheres. J Neurosci Res. 2009;87:3143–52.

30. Contestabile A, Fila T, Bartesaghi R, Ciani E. Cell Cycle Elongation Impairs Proliferation of Cerebellar Granule Cell Precursors in the Ts65Dn Mouse, an Animal Model for Down Syndrome. Brain Pathol. 2009;19:224–37.

31. Slonim DK, Koide K, Johnson KL, Tantravahi U, Cowan JM, Jarrah Z, et al. Functional genomic analysis of amniotic fluid cell-free mRNA suggests that oxidative stress is significant in Down syndrome fetuses. Proc Natl Acad Sci USA. 2009;106:9425–9.

32. Jovanovic SV, Clements D, MacLeod K. Biomarkers of oxidative stress are significantly elevated in Down syndrome. Free Radic Biol Med. 1998;25:1044–8.

33. Li M, Fang X, Baker DJ, Guo L, Gao X, Wei Z, et al. The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. Proc Natl Acad Sci USA. 2010;107:14188–93.

34. Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc Natl Acad Sci USA. 2011;108:680–5.

35. Stefani M. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. Biochim Biophys Acta. 2004;1739:5–25.

36. Katz W, Weinstein B, Solomon F. Regulation of tubulin levels and microtubule assembly in Saccharomyces cerevisiae: consequences of altered tubulin gene copy number. Mol Cell Biol. 1990;10:5286–94.

37. Anders KR, Kudrna JR, Keller KE, Kinghorn B, Miller EM, Pauw D, et al. A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. BMC Genet. 2009;10:36.

38. Weaver BAA, Silk AD, Montagna C, Verdier-Pinard P, Cleveland DW. Aneuploidy Acts Both Oncogenically and as a Tumor Suppressor. Cancer Cell. 2007;11:25–36.

39. Sotillo R, Hernando E, Díaz-Rodríguez E, Teruya-Feldstein J, Cordón-Cardo C, Lowe SW, et al. Mad2 Overexpression Promotes Aneuploidy and Tumorigenesis in Mice. Cancer Cell. 2007;11:9–23.

40. Remeseiro S, Cuadrado A, Carretero M, Martínez P, Drosopoulos WC, Cañamero M, et al. Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. EMBO J. 2012;31:2076–89.

41. Sotillo R, Schvartzman J-M, Socci ND, Benezra R. Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. Nature. 2010;464:436–40.

42. Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut. 2008;57:941–50.

43. Birkbak NJ, Eklund AC, Li Q, McClelland SE, Endesfelder D, Tan P, et al. Paradoxical Relationship between Chromosomal Instability and Survival Outcome in Cancer. Cancer Res. 2011;71:3447–52.

44. Reimand J, Arak T, Vilo J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. 2011;39:W307–15.

45. Jeffery IB, Higgins DG, Culhane AC. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics. 2006;7:359.

46. Hong F, Breitling R. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. Bioinformatics. 2008;24:374–82.

47. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. BioTechniques. 2003;34:374–8.

48. Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, et al. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res. 2009;38:D196–203.

49. De Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinforma Oxf Engl. 2004;20:1453–4.

50. Saldanha AJ. Java Treeview—extensible visualization of microarray data. Bioinformatics. 2004;20:3246–8.

51. Altug-Teber O, Bonin M, Walter M, Mau-Holzmann UA, Dufke A, Stappert H, et al. Specific transcriptional changes in human fetuses with autosomal trisomies. Cytogenet Genome Res. 2007;119:171–84.

52. Mao R, Wang X, Spitznagel EL Jr, Frelin LP, Ting JC, Ding H, et al. Primary and secondary transcriptional effects in the developing human Down syndrome brain and heart. Genome Biol. 2005;6:R107.

53. Koide K, Slonim DK, Johnson KL, Tantravahi U, Cowan JM, Bianchi DW. Transcriptomic analysis of cell-free fetal RNA suggests a specific molecular phenotype in trisomy 18. Hum Genet. 2010;129:295–305.

Chapter 4: A Transcriptional and Metabolic Signature of Primary Aneuploidy is Present in Chromosomally Unstable Cancer Cells and Informs Clinical Prognosis

Reprinted from Cancer Research:

Sheltzer JM. A Transcriptional and Metabolic Signature of Primary Aneuploidy is Present in Chromosomally Unstable Cancer Cells and Informs Clinical Prognosis. Cancer Res 2013;73:6401-12.

Abstract

Aneuploidy is invariably associated with poor proliferation of primary cells, but the specific contributions of abnormal karyotypes to cancer, a disease characterized by aneuploidy and dysregulated proliferation, remain unclear. In this study, I demonstrate that the transcriptional alterations caused by aneuploidy in primary cells are also present in chromosomally-unstable cancer cell lines, but the same alterations are not common to all aneuploid cancers. Chromosomally-unstable cancer lines displayed increased glycolytic and TCA-cycle flux, also a feature of aneuploid primary cells. The biological response to aneuploidy is associated with cellular stress and slow proliferation, and a 70-gene signature derived from primary aneuploid cells was defined as a strong predictor of increased survival in several cancers. Inversely, a transcriptional signature derived from clonal aneuploidy in tumors correlated with high mitotic activity and poor prognosis. Together these findings suggested that there are two types of aneuploidy in cancer, one of which is clonal aneuploidy selected during tumor evolution and associated with robust growth, and the second of which is subclonal aneuploidy caused by chromosomal instability (CIN). Subclonal aneuploidy more closely resembles the stressed state of primary aneuploid cells, yet CIN is not benign: a subset of genes upregulated in high-CIN cancers predict aggressive disease in human patients in a proliferation-independent manner.

Introduction

Aneuploidy and chromosomal instability are inter-related but not identical. Aneuploidy is a description of a cellular *state*; it specifically describes a cell whose karyotype is not a whole-number multiple of the haploid complement. Chromosomal instability is most accurately characterized as a *rate*; it refers to a cell that missegregates chromosomes more frequently than a wild-type cell does. CIN can be caused by mutations or drug treatments that impair the various cellular processes required for accurate chromosome segregation (1). In certain circumstances, aneuploidy itself may cause CIN (2–5). But a cell can become aneuploid without exhibiting CIN, as events like chromosome non-disjunction, micronuclei formation, and cytokinesis failure occur sporadically in normal cells.

Aneuploidy and chromosomal instability are of particular interest due to their paradoxical role in cancer development. Nearly all solid tumors are aneuploid (6), and many types of cancers display elevated levels of chromosome missegregation (7). However, aneuploid primary cells generally exhibit poor proliferative capacity, in stark contrast to the robust growth displayed by aneuploid cancer cells (8–12). Individuals with Trisomy 21/Down syndrome are at elevated risk for the development of childhood leukemias, but have significantly lower levels of solid tumor formation throughout life (13). Furthermore, in mouse models of chromosomal instability, CIN mice are typically tumor-prone (14–16), but in some contexts CIN appears to protect against cancer development (17–19). Thus, a complete understanding of the role of CIN and aneuploidy in cancer must account for both its apparently positive and negative roles in tumorigenesis.

Various strategies to detect aneuploidy and chromosomal instability, including FISH, flow cytometry, and the direct analysis of mitotic figures, have been utilized in an attempt to associate a tumor's chromosomal content with patient risk (7). Most studies have identified a link between aneuploidy and/or CIN and poor clinical outcome (7), though in some contexts,

excessive CIN may actually promote survival (20). Moreover, interpretation of some assays may be complicated by the conflation of aneuploidy (a state), CIN (a rate), and other aspects of cancer biology, including heightened proliferation (21). Nonetheless, the biological changes associated with aneuploidy and CIN could potentially be useful for stratifying patient risk.

We have recently analyzed gene expression data from primary aneuploid cells in various species, including yeast, plants, mice and humans (22). We reported that aneuploidy elicits a transcriptional response indicative of stress and slow growth, and this signature was conserved across eukaryotes. In cancers, gene expression levels generally scale with chromosome copy number (23), but the wider effects of aneuploidy and chromosomal instability on cancer transcriptomes are unknown. Here, I demonstrate that chromosomally-unstable cancer cell lines are transcriptionally and metabolically similar to aneuploid primary cells. While the gene expression changes caused by primary aneuploidy are associated with poor proliferation *in vitro* and improved patient survival *in vivo*, a subset of genes upregulated by chromosomal instability are also associated with aggressive tumors and increased patient risk. I discuss the relevance of these findings for our understanding of the role of aneuploidy in tumorigenesis.

Results

Transcriptional similarities between aneuploid primary cells and chromosomally-unstable cancer cells

I first developed a transcriptional signature of primary aneuploidy using gene expression data from 18 lines of mouse embryonic fibroblasts (MEFs) that were either euploid or trisomic for one of four chromosomes (9). I sorted these cell lines according to their degree of aneuploidy, as determined by the number of protein-coding genes that were present on the trisomic chromosomes. I then calculated the Spearman rank-order correlation coefficient (SCC) between each gene expression vector and the degree of aneuploidy across the panel of MEFs.

I extracted genes whose expression tended to either increase or decrease according to the degree of aneuploidy ($\rho>0.3$ or $\rho<-0.3$), then identified gene ontology (GO) terms enriched among these gene sets. Among up-regulated genes (those that were positively correlated with aneuploidy), GO term analysis revealed an enrichment of genes annotated to the membrane ($p<10^{-12}$), the vesicle ($p<10^{-4}$), and the extracellular region ($p<10^{-3}$), among other processes (Table S1). Downregulated genes were enriched for those annotated to nucleic acid metabolism ($p<10^{-17}$), DNA repair ($p<10^{-12}$), and RNA processing ($p<10^{-9}$). These GO terms are consistent with our hypothesis that primary aneuploidy is an anti-proliferative cellular stress (22).

Next, I analyzed gene expression data from the NCI60 panel of cancer cell lines, which comprise 59 independent cell lines derived from 9 types of cancer. Importantly, several karyotypic parameters have been determined for each of these lines, including their index of numerical heterogeneity [INH; (26)]. The INH is a measure of cell-to-cell variability in chromosome content: cancers that exhibit CIN have large INH values, while karyotypically stable cancers have INH values that are close to 0. Aneuploidies that are present in all cells do not contribute to INH (see Materials and Methods). I hypothesized that if aneuploidy is a cellular stress, then the continuous generation of new aneuploidies in cell lines that displayed CIN could cause an analogous stress response. This would be reflected in similar gene expression patterns between primary aneuploid cells and CIN cancer lines. To test this, I calculated Spearman correlation coefficients between gene expression vectors in the NCI60 lines and their respective indexes of numerical heterogeneity (36). As INH values varied depending on the tissue of origin of the cell lines, I removed genes that displayed tissue-specific expression patterns via an ANOVA. Representative genes are displayed in Fig. 1A and 1B: transcript levels of the DNA replication factor *MCM6* tend to decrease in cell lines with high CIN, while transcript levels of the ER-associated gene *CALU* are correlated with increasing CIN. Surprisingly, genes correlated or anti-correlated with CIN were enriched for many of the same

functional categories as observed in aneuploid primary cells (Fig. 1C and Table S2). Nucleic acid metabolism ($p<10^{-34}$) and the chromosome ($p<10^{-13}$) were highly enriched among genes that decreased with increasing CIN, while the membrane ($p<10^{-6}$) and the extracellular matrix ($p<10^{-6}$) were positively correlated with CIN. Additionally, gene expression ($p<10^{-21}$) and RNA splicing ($p<10^{-25}$) transcripts were strongly downregulated by CIN, but were more modestly affected by aneuploidy ($p<10^{-10}$ and $p<10^{-6}$, respectively). In total, 26% of GO terms upregulated by aneuploidy in MEFs were also upregulated by CIN in cancer cells, and 55% of GO terms downregulated by aneuploidy in MEFs were also downregulated by CIN in cancer cells. Among the 35 GO terms that were most strongly anti-correlated with primary aneuploidy, all 35 were also anti-correlated with CIN.

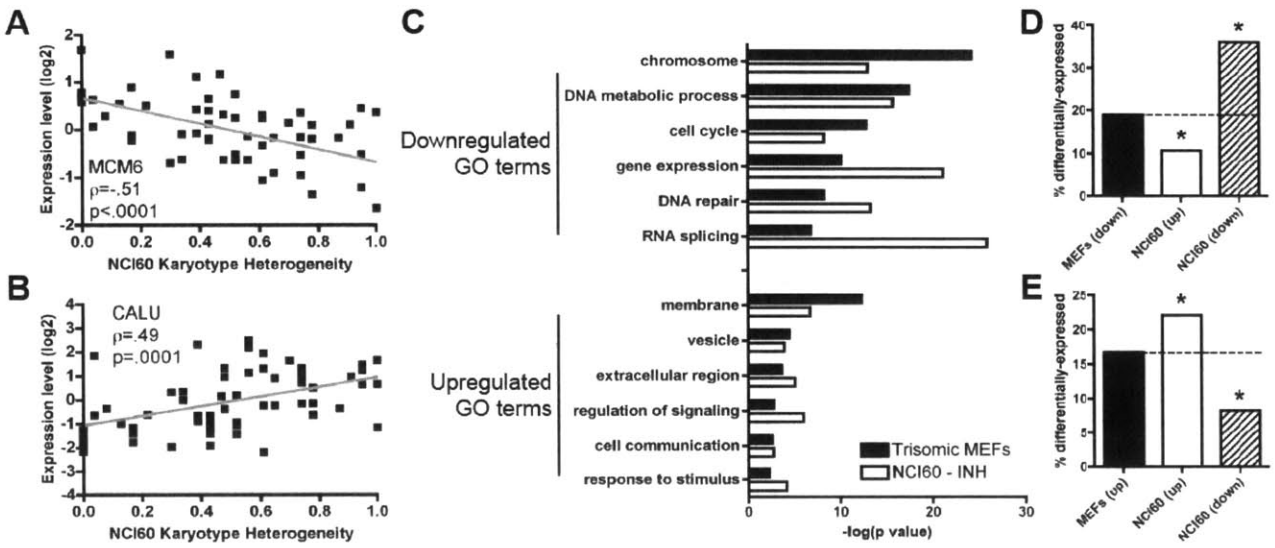


Figure 1. Transcriptional similarities between trisomic MEFs and chromosomally-unstable cancer cell lines. (A and B). Representative examples of genes that are negatively (*MCM6*) or positively (*CALU*) correlated with karyotype heterogeneity in the NCI60 panel. Each square represents the expression level of the indicated gene in one cell line, and a linear regression (gray line) is plotted against the data. (C) Gene Ontology categories that are enriched among genes negatively or positively correlated with aneuploidy in MEFs (black bars) or karyotype heterogeneity in the NCI60 panel (white bars). Complete GO term lists are listed in Tables S1 and S2. (D) Genes that are negatively correlated with aneuploidy in MEFs are significantly enriched for those that are also negatively correlated with karyotype heterogeneity in the NCI60 panel. The black bar depicts the percent of genes that are negatively correlated with aneuploidy in MEFs. White bars represent the percent of genes negatively correlated with aneuploidy in MEFs that are also positively correlated with karyotype heterogeneity in the NCI60 panel. Hashed bars represent the percent of genes negatively correlated with aneuploidy that are also negatively correlated with karyotype heterogeneity. Asterisks represent a degree of overlap that is greater than or less than the overlap expected by chance ($p<.05$; hypergeometric test). (E) Same analysis as in (D) for genes positively correlated with aneuploidy in MEFs.

To confirm that these GO term enrichments were not due to the specific correlation cutoff used, I also analyzed the primary aneuploidy and NCI60 datasets with Gene Set Enrichment Analysis (GSEA), a threshold-independent method of correlating phenotypes with gene sets (31). GSEA revealed similar GO term enrichments as described above, particularly among down-regulated genes. DNA replication, RNA splicing, and the cell cycle were significantly anti-correlated with both increasing aneuploidy and increasing CIN (Fig. S1, Tables S3-S4). I conclude that aneuploidy in primary cells and CIN in cancer cell lines affect the expression of genes involved in a limited and highly similar set of biological functions.

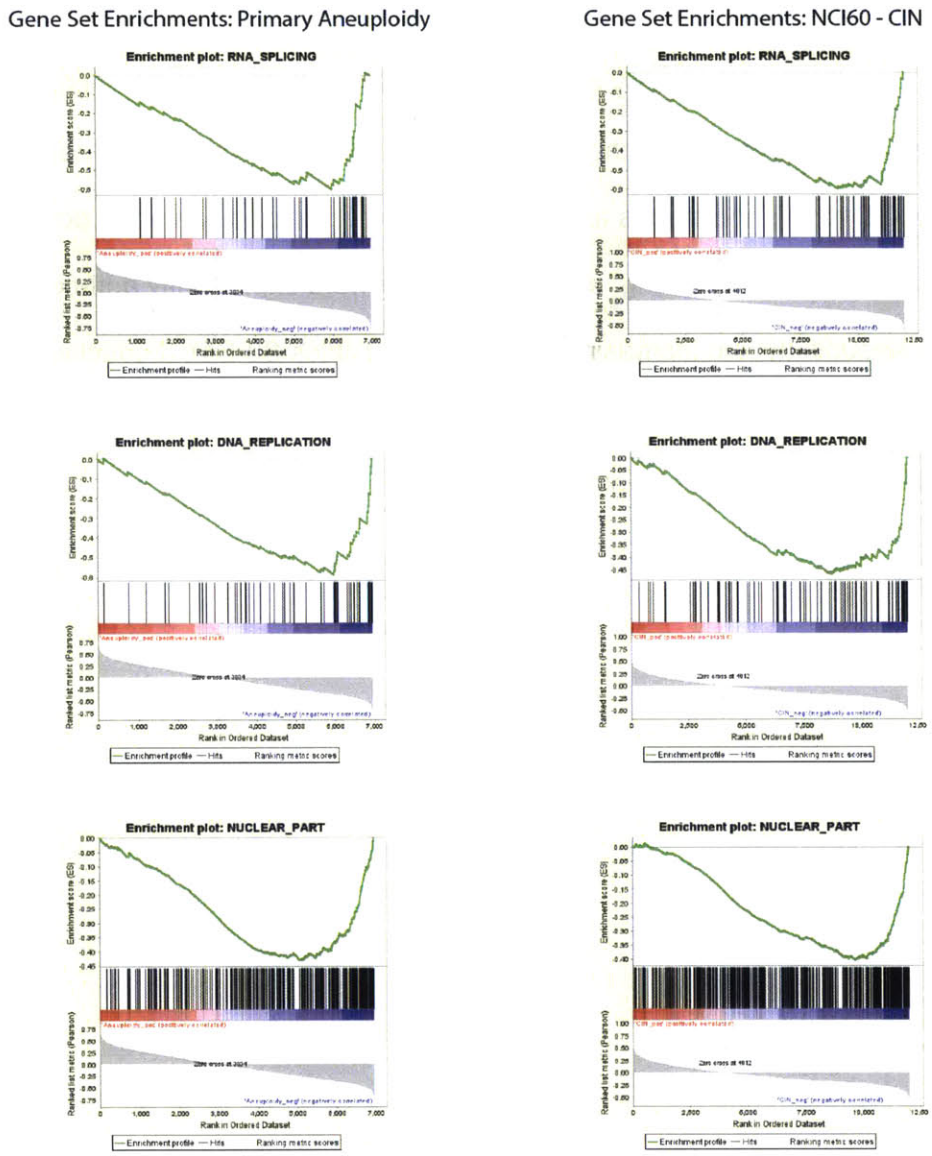


Figure S1. Gene Set Enrichment Analysis of aneuploidy in trisomic MEFs and karyotype heterogeneity in cancer cell lines. Representative enrichment plots are shown of Gene Ontology categories that are enriched among genes correlated with aneuploidy (left) and CIN (right).

I next sought to determine whether the transcriptional similarities between aneuploid and CIN cell lines extended to the level of individual genes. To test this, I identified one-to-one orthologs between mouse and human genes (see *Materials and Methods*), and then compared the SCC values for each transcript. I found highly significant overlap among the sets of orthologous genes that were upregulated or downregulated by aneuploidy and CIN (Fig. 1D and 1E). For instance, 19% of all genes were anti-correlated with aneuploidy in MEFs at a $\rho<-0.3$ cut-off. Among orthologous genes anti-correlated with CIN in cancer cells, a significantly greater percentage (36%) were also anti-correlated with aneuploidy in MEFs ($p<10^{-16}$, hypergeometric test). In contrast, among genes that were positively correlated with CIN, significantly fewer (10%) were anti-correlated with aneuploidy ($p<10^{-3}$, hypergeometric test). Correspondingly, SCC values in both MEFs and cancer cells displayed a small but highly significant genome-wide correlation ($\rho=0.14$, $p<10^{-23}$). When the analysis was restricted to only those genes affected by CIN in cancer cells ($\rho<-0.3$ and $\rho>0.3$), the strength of the correlation increased to $\rho=0.28$ ($p<10^{-13}$). I conclude that aneuploid primary cells and chromosomally-unstable cancer cells share a high degree of transcriptional similarity at the level of both functional processes and individual genes.

Aneuploid primary cells and human tumors display distinct transcriptional programs

As the aneuploid transcriptional program was apparent in chromosomally-unstable cancer cells, and as the vast majority of human tumors are aneuploid, I next sought to determine whether the transcriptomes of human tumors resemble those of primary aneuploid cells. I examined gene expression data from 9 common cancer types that were normalized to disease-free tissue, and I compared ranks between the mean fold change expression levels in the tumors and the SCC values from trisomic MEFs (Fig. S2). Interestingly, the average expression levels from tumors were anti-correlated with the aneuploidy-responsiveness of

genes in MEFs (Fig. S2A; average ρ=-0.16, $p<10^{-29}$). I then identified differentially-expressed genes between tumors and normal tissue using a combined t test ($p<.05$) and mean fold change (±1.3) threshold. Transcripts upregulated in aneuploid primary cells were significantly more likely to be downregulated in tumors, while transcripts downregulated in aneuploid primary cells were likely to be upregulated in tumors (Fig. S2B-S2C). Few GO terms were significantly enriched among the small number of genes that were similarly affected by aneuploidy and cancer (Fig. S2D-S2E and Table S5). Conversely, GO terms related to cell cycle progression were strongly enriched among genes downregulated by aneuploidy but upregulated in tumors (Fig. S2E and Table S6). To address the possibility that these results were due to comparing tumors to untransformed tissue, I next analyzed a cohort of breast tumors that had either a diploid or aneuploid basal ploidy (38). After normalizing gene expression in aneuploid tumors to the mean levels in diploid tumors, I found that gene expression ranks in aneuploid tumors were also anti-correlated with the expression levels from trisomic MEFs (Fig. S2F; ρ=-0.12, $p<10^{-13}$). Additionally, the GO categories perturbed in aneuploid tumors were dissimilar to those affected by aneuploidy in primary cells (Fig. S2G and Table S7). In particular, transcripts upregulated in aneuploid tumors were highly enriched for those involved in cell cycle progression. Thus – unlike in primary cells, where aneuploidy induces a chromosome- and species-independent gene expression response (22) – aneuploid human tumors are transcriptionally distinct from aneuploid primary cells.

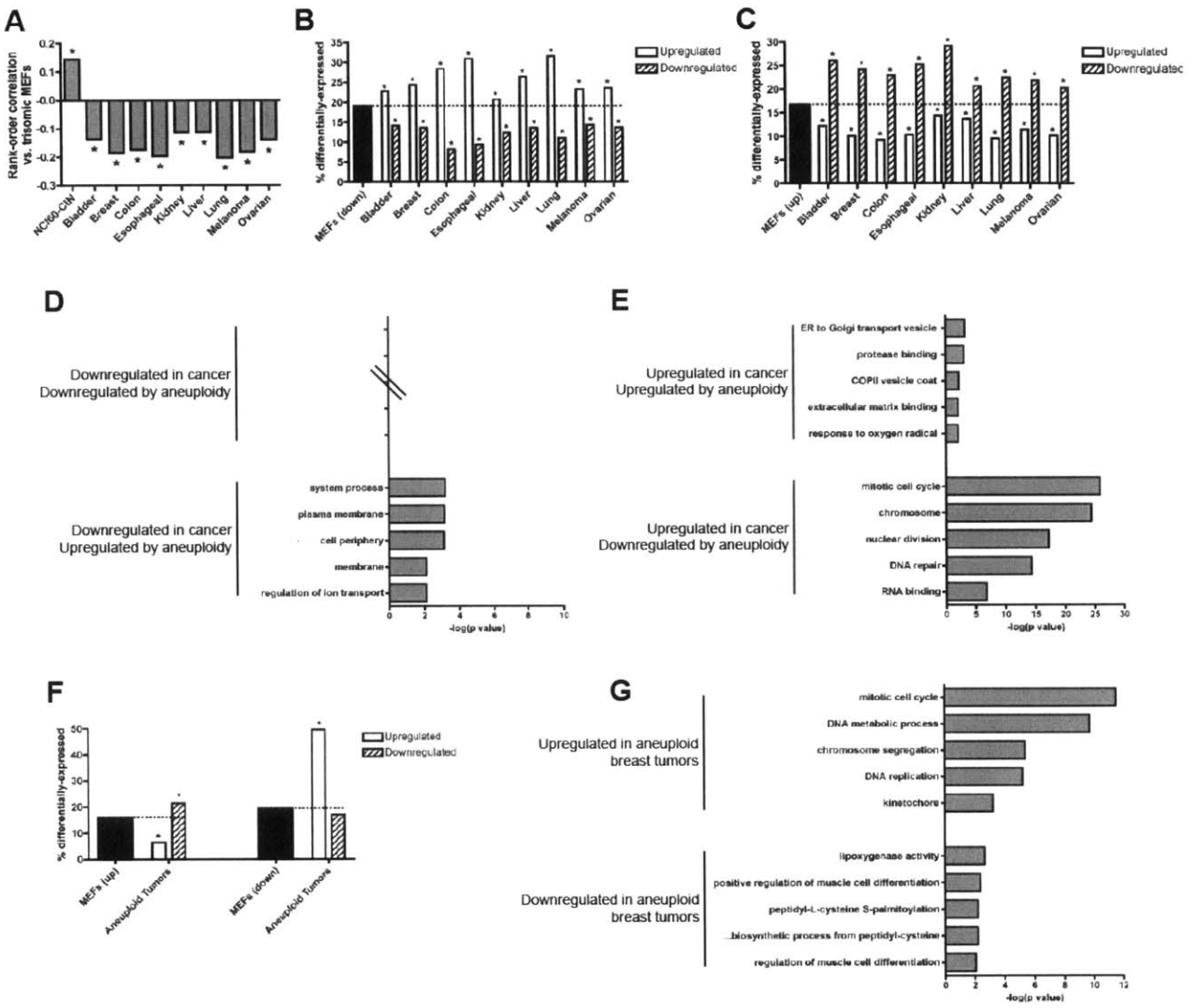


Figure S2. Trisomic MEFs and human cancers are transcriptionally distinct. (A) Spearman correlation coefficients between the SCC values from trisomic MEFs and the SCC-CIN values from the NCI60 panel or the ranked fold-change values from nine different cancer types. (B) Genes that are negatively correlated with aneuploidy are significantly enriched for those that are upregulated in cancer cells. The black bar depicts the percent of genes that are negatively correlated with aneuploidy in MEFs. White bars represent the percent of genes negatively correlated with aneuploidy in MEFs that are also upregulated in the indicated cancer type. Hashed bars represent the percent of genes negatively correlated with aneuploidy that are also downregulated in cancer cells. Asterisks represent a degree of overlap that is greater

than or less than the overlap expected by chance ($p<.05$; hypergeometric test). (C) Same analysis as in (B) for genes positively correlated with aneuploidy in MEFs. (D) GO terms enriched among genes downregulated by aneuploidy and cancer, or downregulated in cancer but upregulated by aneuploidy. (E) GO terms enriched among genes upregulated by aneuploidy and cancer, or upregulated in cancer but downregulated by aneuploidy. Complete GO term lists are presented in Tables S5 and S6. (F) Genes that are positively or negatively correlated with aneuploidy in MEFs are more likely to exhibit an opposite transcriptional change in aneuploid breast tumors ($p<.05$; hypergeometric test). (G) GO terms enriched among genes that are differentially expressed in aneuploid breast tumors, relative to diploid breast tumors. Complete GO term lists are presented in Table S7.

The CIN transcriptional signature resembles a slow proliferation response

Why would primary aneuploid cells resemble high-CIN cancer cell lines but not all aneuploid cancers? We have proposed that the transcriptional similarities between aneuploid cells result from their shared stress and/or slow-growth phenotype (22). I hypothesized that a similar effect could drive the transcriptional program observed in karyotypically-unstable cancer cells: the stresses associated with a continuously-varying karyotype could lead to the down-regulation of cell cycle genes and the upregulation of stress- and signaling-related genes. Consistent with this hypothesis, I found that the index of numerical heterogeneity in the NCI60 panel was significantly correlated with increasing doubling times: cells with high CIN typically divided more slowly than did cells with low CIN (Fig. 2A; ρ=0.40, p<.003). This effect remained true when cell lines derived from hematopoietic cancers were excluded (ρ=0.36, p<.02). Consistent with a relationship between CIN and cell cycle progression, genes downregulated by CIN were highly enriched for targets of the canonical cell cycle transcription factor E2F1 as well as other E2F family members ($p<10^{-8}$; Table S8). I recalculated SCC values according to how strongly a gene expression vector correlated with doubling time, and I observed that GO terms enriched among genes that were correlated or anti-correlated with doubling time were very similar to those enriched among CIN-correlated genes. Among both gene sets, membrane and signaling-related transcripts were upregulated and RNA processing and cell cycle-related transcripts were downregulated (Fig. 2B and Table S9). However, the transcriptional responses to CIN and slow growth were not identical: I subtracted doubling time-SCC values from CIN-SCC values and analyzed the sets of genes that were more strongly correlated or anti-correlated with CIN than with proliferation (ρ_{CIN}- ρ_{DT}>0.3 or <-0.3). Genes annotated to the mitochondrion ($p<10^{-15}$), the mitochondrial matrix ($p<10^{-6}$), and cellular metabolic processes ($p<10^{-3}$) displayed a stronger correlation with CIN than with doubling time (Fig 2C and Table S10). No GO terms were significantly enriched among genes that were more strongly correlated with doubling time than with CIN. These data are intriguing as primary aneuploid

cells display various metabolic alterations (9,12,16), and they suggest that the upregulation of certain metabolic processes may be a growth-rate independent consequence of CIN in cancer. I conclude that, aside from metabolic gene expression, the transcriptional programs observed in chromosomally-unstable and slowly-dividing cancers are highly similar.

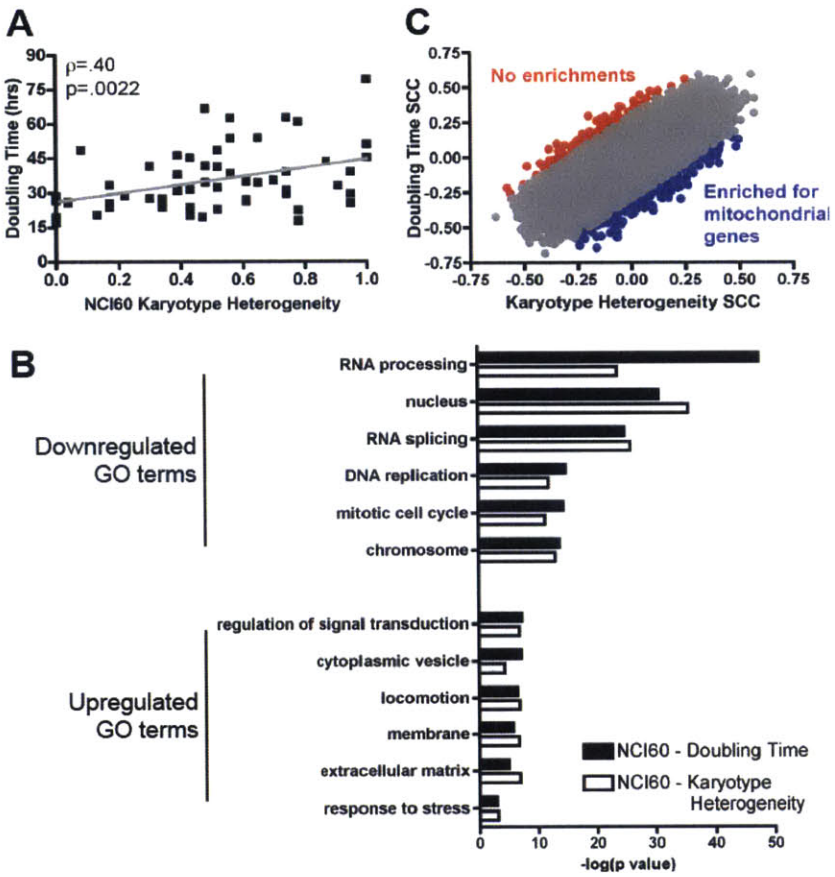


Figure 2. Transcriptional similarities between chromosomally-unstable cancer cell lines and cancer cell lines with slow doubling times. (A) Karyotype heterogeneity and doubling time are significantly correlated in the NCI60 panel. A linear regression (gray line) is plotted against the data. (B) Gene Ontology categories that are enriched among genes negatively or positively correlated with doubling time (black bars) or karyotype heterogeneity (white bars) in the NCI60 panel. Complete GO term lists are listed in Tables S2 and S9. (C) Spearman rank correlations for each gene expression vector and karyotype heterogeneity (X axis) or doubling time (Y axis) were plotted. Genes that display a stronger positive correlation with karyotype heterogeneity are displayed in blue, while genes that display a stronger positive correlation with doubling time are displayed in red.

Metabolic similarities between aneuploid primary cells and chromosomally-unstable cancer cells

Aneuploid MEFs display various metabolic alterations: in general, trisomic cells utilize more glutamine, and produce more lactate, glutamate, and ammonium than euploid cells do. Additionally, some aneuploid cells utilize more glucose (9,16). Do chromosomally-unstable cancer cells exhibit the same metabolic patterns? To address this question, I calculated SCC values between numerical heterogeneity and the consumption/release rates of 140 different metabolites from a recently-published analysis of the NCI60 panel (28). High CIN cells tended to utilize more glutamine and glucose, and produce more lactate and glutamate, than low CIN cells did (Fig. 3A; ammonium was not measured). Indeed, the released metabolite that was most strongly correlated with increasing CIN was glutamate, while the second-strongest correlation among consumed metabolites was with glutamine (Fig. 3B). More broadly, metabolites whose production tended to increase with CIN were associated with the citric acid cycle (e.g., fumarate, malate; $p<10^{-3}$) and glycolysis (e.g., pyruvate, 2-phosphoglycerate; $p<10^{-3}$; Table S11). Consistent with the microarray data (Fig. 2C), some of these metabolites were also correlated with doubling time, though generally less strongly than they were correlated with CIN (Fig. 3B). Moreover, the citric acid cycle and glycolysis were not enriched among metabolites that were correlated with doubling time (Table S12). Thus, in addition to their similarities at the transcriptional level, aneuploid primary cells and chromosomally-unstable cancers share several metabolic alterations. The cause of these alterations is not currently clear (see Discussion).

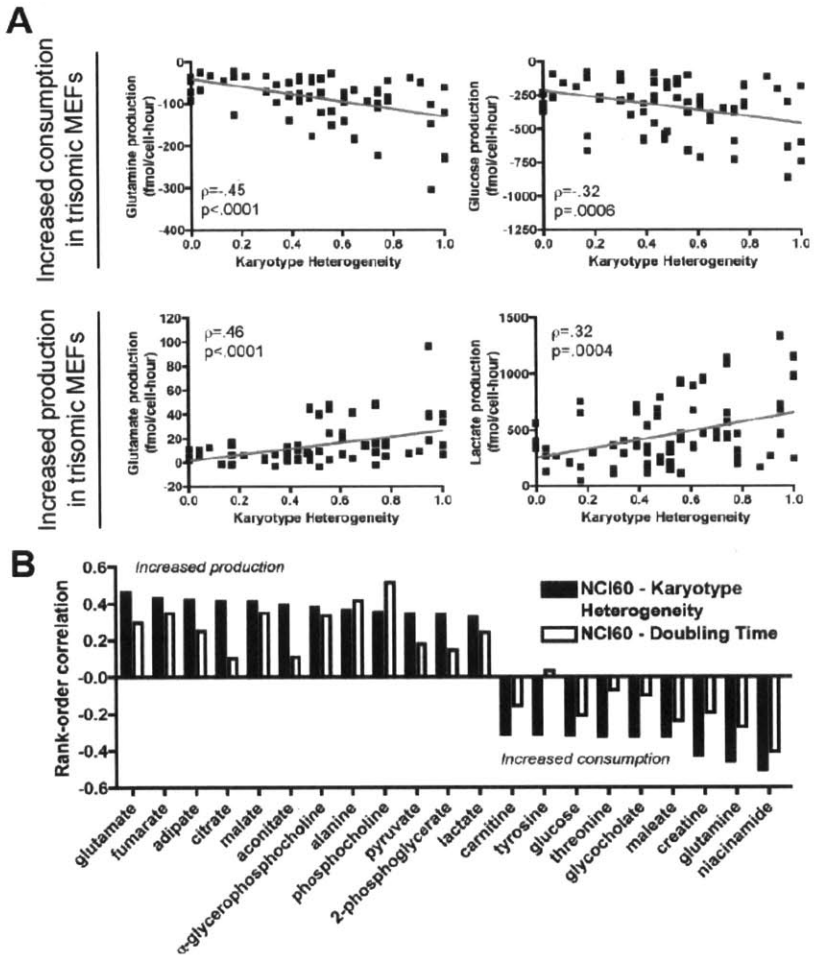


Figure 3. Metabolic similarities between trisomic MEFs and chromosomally-unstable cancer cell lines. (A) Glucose and glutamine consumption tend to increase, and glutamate and lactate production tend to increase, with karyotype heterogeneity in the NCI60 panel. Gray lines represent linear regressions plotted against the data. (B) Comparison between the metabolic profiles of chromosomally-unstable and slow-growing cancer lines. Metabolites whose consumption or production increase with doubling time (white bars) tend to show similar but smaller trends than when those metabolites are correlated with karyotype heterogeneity (black bars; $p<.01$ for both produced and consumed metabolities, paired t-test).

The CIN transcriptional signature is apparent in spindle checkpoint-defective MEFs and after chromosome transfer into karyotypically-stable cancer cells

Thus far, I have established a correlative relationship between the metabolic and transcriptional programs of aneuploid primary cells and chromosomally-unstable cancer lines. However, it remains possible that other aspects of the cancer lines cause the similarity with aneuploid primary cells, and CIN is an unrelated covariant. To demonstrate a direct link between chromosome missegregation and the CIN transcriptional signature, I examined the transcriptomes in chromosomally-unstable primary cells and in chromosomally-stable cancer cell lines in which aneuploidy was ectopically induced.

Chromosomal instability can be genetically encoded by introducing mutations that affect the function of the Spindle Assembly Checkpoint (SAC). Hypomorphic alleles of two crucial SAC components, Cdc20 and BubR1, have been constructed, and MEFs containing these mutations exhibit significant CIN (39,40). I derived MEFs from embryos that were homozygous for either hypomorphic allele ($Cdc20^{AAA/AAA}$ or $BubR1^{H/H}$) as well as from their wild-type littermates ($Cdc20^{+/+}$ and $BubR1^{+/+}$). I then performed microarray analysis on two early-passage BubR1 sibling pairs and one early-passage Cdc20 sibling pair, and in each case normalized gene expression in the mutant MEF line to its wild-type littermate. Transcript levels were similar between the two CIN alleles (data not shown), and therefore the following comparisons were made using the average expression levels in the CIN MEFs. As expected, gene expression levels in the chromosomally-unstable MEFs were moderately but significantly correlated with SCC-CIN values from the NCI60 panel and with expression levels in the trisomic MEFs ($\rho=0.15$, $p<10^{-26}$, and $\rho=0.26$, $p<10^{-106}$, respectively). Genes upregulated in CIN MEFs tended to be upregulated in high-CIN cancer cells and in trisomic MEFs, and vice-versa (Fig. S3A-S3D). GO term enrichment analysis revealed the upregulation of extracellular and cell motility genes, and the down-regulation of cell cycle transcripts, as is also observed in the trisomic MEFs and NCI-

60 panel (Table S13 and Fig. S3E). Similar results were obtained when either $Cdc20^{AAA/AAA}$ or $BubR1^{H/H}$ MEFs were analyzed independently (data not shown). I conclude that the transcriptional patterns present in primary cells that exhibit chromosomal instability are highly similar to both primary aneuploid cells and to cancer cells that are chromosomally-unstable.

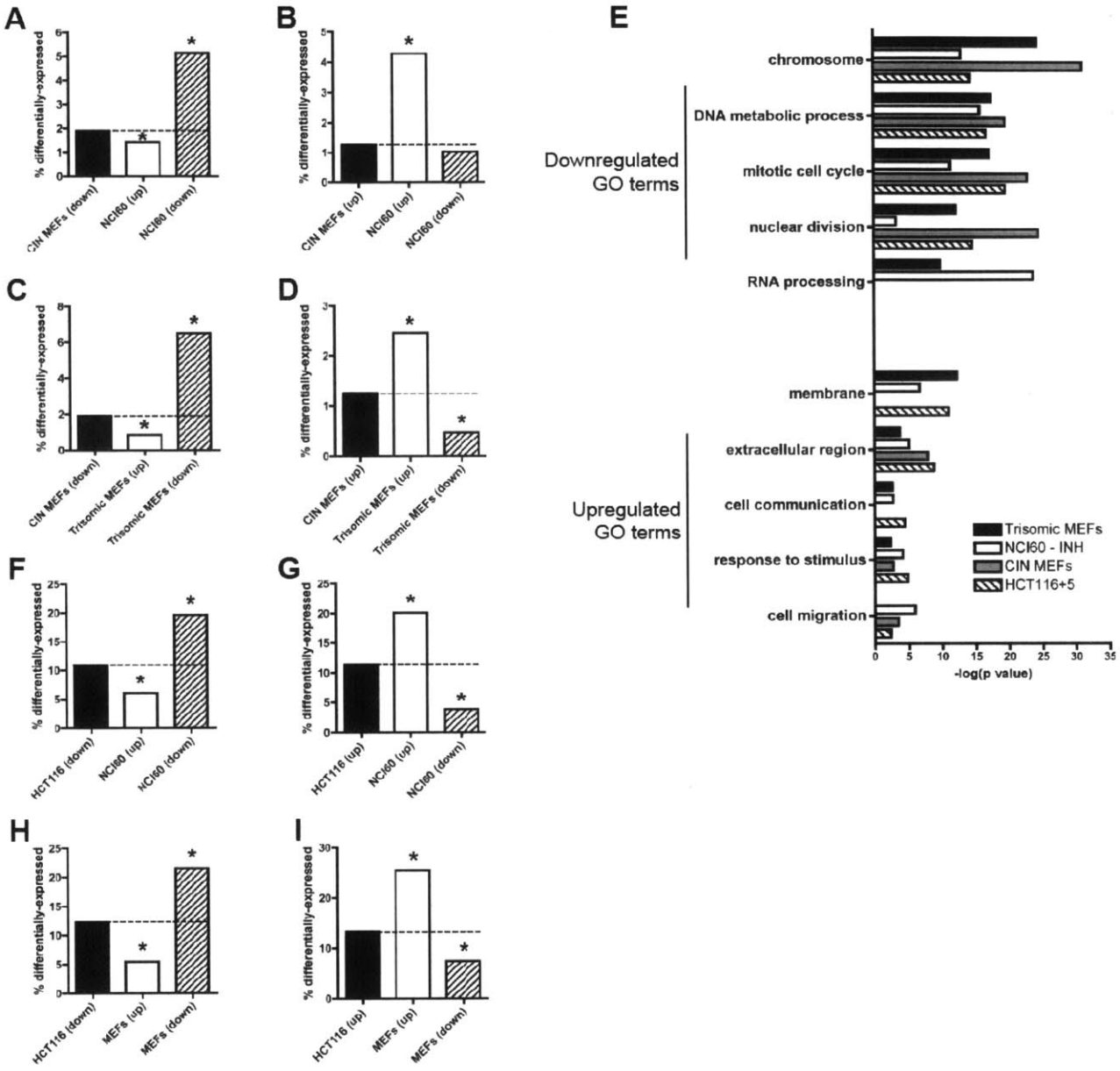


Figure S3. Transcriptional similarities between chromosomally-unstable cancer cell lines, trisomic MEFs, chromosomally-unstable MEFs, and cancer cells that have received extra chromosomes via MMCT. (A and B). Differentially expressed genes in $Cdc20^{AAA/AAA}$ and $BubR1^{H/H}$ MEFs (CIN MEFs) are significantly more likely to display a co-directional correlation with karyotype heterogeneity in the NCI60 panel. Black bars depict the percent of genes that are (A) downregulated or (B) upregulated in CIN MEFs, relative to wild-type lines. White bars represent the percent of genes that are differentially expressed in CIN MEFs that are also

positively correlated with karyotype heterogeneity in the NCI60 panel. Hashed bars represent the percent of differentially expressed genes that are negatively correlated with karyotype heterogeneity in the NCI60 panel. Asterisks represent a degree of overlap that is greater than or less than the overlap expected by chance ($p<.05$; hypergeometric test). (C and D). Same analysis as in (A) and (B) but comparing CIN MEFs with trisomic MEFs. (E) Gene Ontology categories that are enriched among genes negatively or positively correlated with primary aneuploidy (black bars), that are correlated with karyotype heterogeneity in the NCI60 panel (white bars), that are differentially-expressed in CIN MEFs (gray bars), or that change in expression after chromosome transfer into HCT116 cells (hashed bars). Complete GO term lists are presented in Tables S1, S2, S13, and S14. (F and G). Same analysis as in (A) and (B) but comparing tetrasomic HCT116 cells with karyotype heterogeneity in the NCI60 panel. (H and I). Same analysis as in (A) and (B) but comparing tetrasomic HCT116 cells with trisomic MEFs.

To further assess the link between the CIN transcriptional signature and chromosome missegregation, I examined the effects of ectopic aneuploidy in cancer cells. A recent report has analyzed the consequences of transferring additional chromosomes into HCT116 cells, a karyotypically stable (INH=0) colon cancer line (29). I acquired microarray data from HCT116 cells that had two extra copies of chromosome 5 introduced via microcell-mediated chromosome transfer (MMCT). I then repeated the above analysis comparing gene expression in the newly-tetrasomic cancer cells to the SCC-CIN values from the NCI60 panel and to the trisomic MEFs. The transcriptional changes induced by aneuploidy in the HCT116 cells were moderately similar to those correlated with CIN across the NCI60 panel and with aneuploidy in primary MEFs (Fig. S3F-S3I; $\rho=0.26$, $p<10^{-47}$, and $\rho=0.22$, $p<10^{-43}$, respectively). The induction of aneuploidy led to the downregulation of cell cycle transcripts and the upregulation of stress and membrane-related transcripts (Table S14 and Fig. S3E). Thus, inducing aneuploidy in an otherwise stable cancer cell line mimics the CIN transcriptional signature. Amongst all four microarray datasets examined, there was significant overlap among GO terms, particularly among downregulated transcripts and those associated with the extracellular matrix. However, transcripts involved in RNA processing were enriched among downregulated genes in the NCI60 panel and in aneuploid MEFs, but not in CIN MEFs or HCT116+5 cells (Fig. S3E). The cause of the minor transcriptional differences that are apparent between these cell types is at present unknown.

A previous transcriptional signature of CIN is anti-correlated with karyotype heterogeneity and doubling time

Carter et al. (41) have described a transcriptional signature for chromosomally-unstable cancers that has been widely used as a marker for genomic instability (20,42–44). They determined the degree of aneuploidy across several cancer datasets by summing the total

number of chromosomal regions that showed consistent alterations in expression levels. Genes were then ranked according to how strongly they correlated with the total aneuploidy in each tumor, and the top 70 genes were used to create the CIN70 gene signature. Finally, they showed that in various cancer types, a high CIN70 score was associated with poor clinical prognosis. It is important to note that their expression signature most likely detects clonal aneuploidy, i.e. aneuploidy that is present in the bulk of a tumor, as rare aneuploidies would not cause detectable differences in gene expression levels *in cis*. Furthermore, a genome with many chromosomal aneuploidies is not necessarily indicative of a current state of chromosomal instability: genetic alterations may accumulate at a low rate over a long period of time, with evolutionary pressure selecting the spread of cells with the most growth-advantageous mutations (45). As my analysis of CIN suggested that it was associated with poor proliferation, I decided to further explore the relationship between the CIN70 gene set and cancer.

Surprisingly, I found that the CIN70 score was anti-correlated with karyotype heterogeneity in the NCI60 panel (Fig. S4A; ρ=-0.35, $p<.01$). Chromosomally stable cancer cell lines (HCT116, CCRF-CEM) generally had higher CIN70 scores than unstable cell lines (HOP-92, TK-10). Upon inspection of the CIN70 gene list, I noted that many transcripts were well-characterized markers of proliferation (e.g., *PCNA*, *CDK1*, *MCM2*). It has recently been suggested that many prognostic gene signatures, including CIN70, function in part by capturing information about cell proliferation rates (46). I hypothesized that clonal aneuploidy in a tumor was not a strong indicator of CIN, but instead reflected a highly-evolved cancer state, in which a tumor had acquired numerous growth-promoting genetic alterations. Indeed, the CIN70 score was strongly correlated with decreased doubling times in the NCI60 panel (Fig. S4B; ρ=-0.46, $p<.001$). To further test the link between CIN70 and proliferation rates, I examined a microarray dataset of 81 non-diseased human tissues and cell populations. As a proxy for their proliferative index, I sorted the tissues by their level of PCNA expression. While 30% of all

genes were positively correlated with PCNA in this dataset (average ρ=0.12, p>.05), I found that 100% of CIN70 genes were positively correlated with PCNA (Fig. S4C; average ρ=0.45, $p<10^{-7}$). Similar correlations were observed between CIN70 and other common proliferative markers (e.g., ρ_{MKI67}=0.57, $p<10^{-13}$, ρ_{TOP2A}=0.50, $p<10^{-9}$). It has been argued that CIN70's prognostic utility is not derived from its ability to detect proliferation rates, as all cell cycle-regulated genes can be removed from CIN70 without abolishing its power to stratify tumors (41). However, the remaining genes [referred to as the "CIN27wp" signature (42)] were still highly correlated with PCNA (average ρ=0.40, $p<10^{-6}$), demonstrating that this method is insufficient to control for the widespread effects of proliferation on gene expression. I conclude that the CIN70 score is not an accurate predictor of chromosomal instability in the NCI60 panel, but instead captures information about cell proliferation rates in both the NCI60 lines and in chromosomally-stable normal tissue.

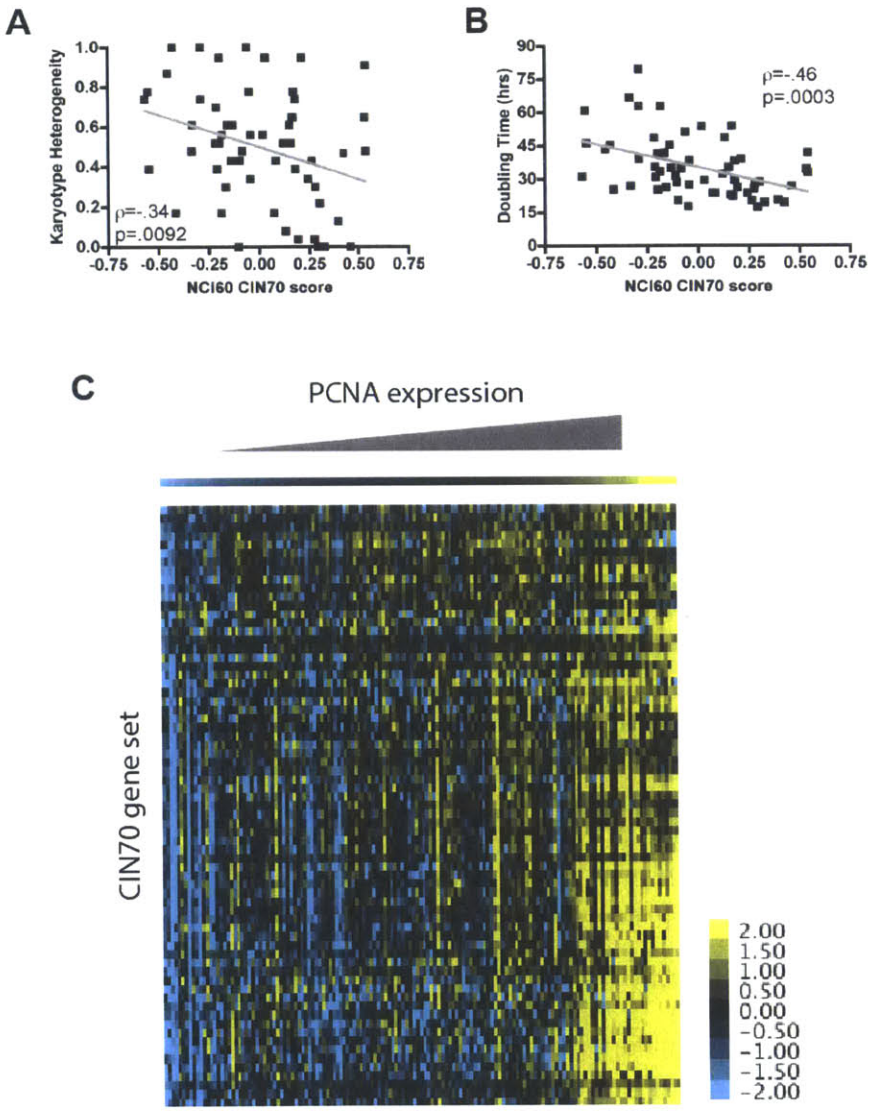


Figure S4. The CIN70 gene signature is correlated with proliferation rates and anti-correlated with karyotype heterogeneity. (A) The CIN70 scores of cell lines from the NCI60 panel are negatively correlated with karyotype heterogeneity. (B) The CIN70 scores of cell lines from the NCI60 panel are correlated with decreasing doubling times. (C) Gene expression data from 81 non-diseased human tissues and cell populations were sorted according to their level of PCNA expression. Each gene in the CIN70 gene set is positively correlated with increasing PCNA. A clustergram of CIN70 gene expression levels, sorted according to the level of PCNA expression in each tissue, is displayed.

Construction of gene signatures associated with aneuploidy and chromosomal instability

As aneuploidy and chromosomal instability have been hypothesized to be integral drivers of cancer evolution, I sought to determine whether the aneuploidy and CIN-associated gene expression patterns identified herein could serve as useful *in vivo* markers for cancer progression. Because our analysis of the CIN70 gene set suggested that it largely reflected proliferative capacity, I also sought to clarify whether the aneuploidy-induced transcriptional programs were identical to or distinct from the transcriptional response to slow growth.

To explore the prognostic relevance of the gene expression changes associated with aneuploidy and CIN, I constructed three additional univariate gene signatures that are analogous to CIN70. First, I defined PCNA25, which consists of the 25 genes that displayed the strongest correlation with PCNA expression in normal human tissue, and which I used as a non-specific marker of cell proliferation [(46), Table S15]. Secondly, I defined TRI70, which consists of the 70 genes that displayed the strongest absolute correlation with aneuploidy in trisomic MEFs (Table S16). Thirdly, I defined HET70, which consists of the 70 genes that displayed the strongest correlation with karyotype heterogeneity in the NCI60 panel (Table S17). I hypothesized that the most prominent gene expression changes associated with clonal aneuploidy in cancer and primary aneuploidy in MEFs were related to proliferation, and that therefore CIN70 could be used to identify rapidly dividing cells while TRI70 would identify slowly dividing cells. Indeed, CIN70 and TRI70 were able to sort human tissues by their level of PCNA expression, demonstrating that, in the absence of aneuploidy or chromosomal instability, these gene signatures reflect cell proliferation rates (Fig. S5A and S5B; $\rho=0.61$, $p<10^{-15}$, and $\rho=-0.29$, $p<10^{-3}$, respectively). Interestingly, the HET70 score was neither correlated nor anti-correlated with PCNA in human tissue (Fig. S5C and S5D; $\rho=0.11$, $p>.05$). While the overall gene expression pattern of cells with heterogeneous karyotypes resembled that of a slow growth response (Fig. 2), the 70 genes most associated with karyotype heterogeneity were not

enriched for proliferation markers and only 30% of the transcripts were individually correlated with PCNA. I conclude that, unlike the CIN70 and TRI70 gene sets, HET70 scores are uncorrelated with proliferative index.

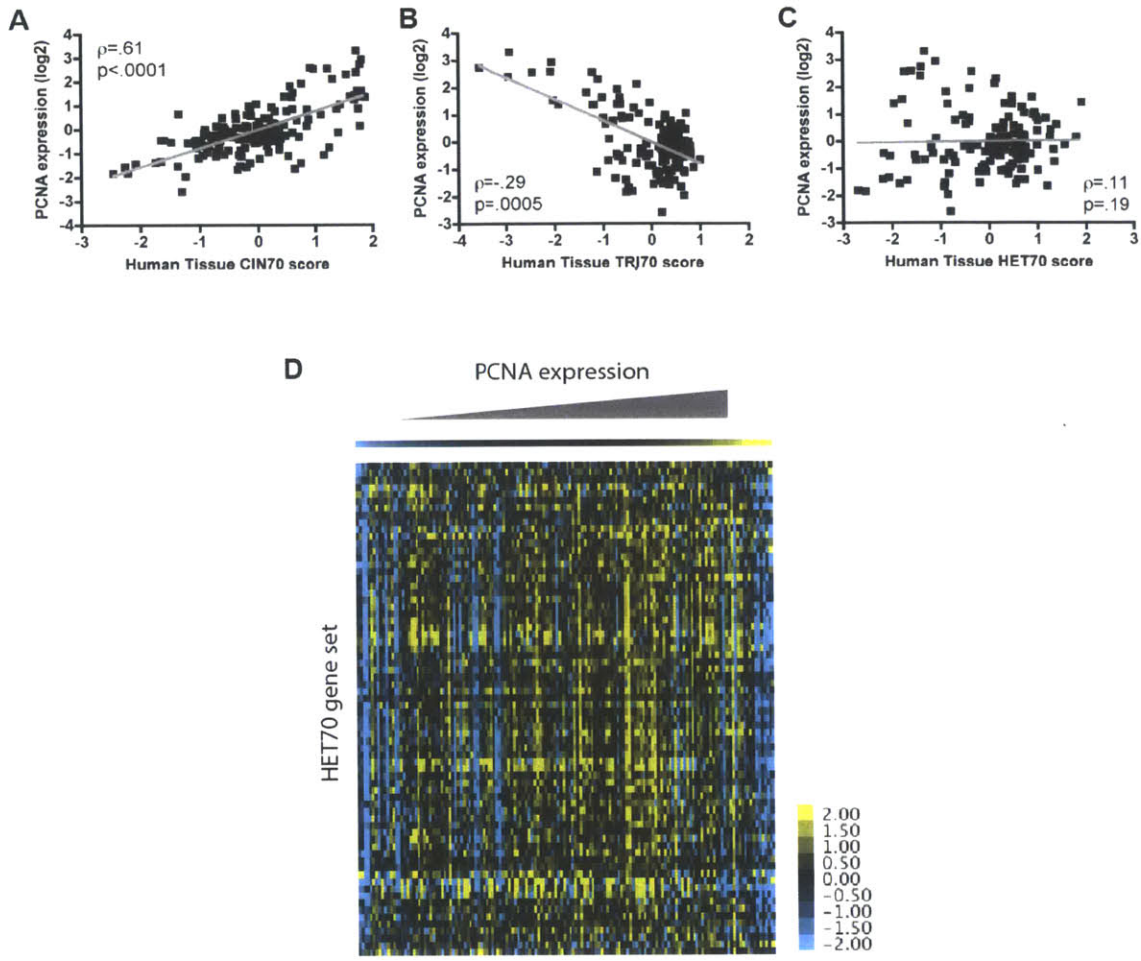


Figure S5. The TRI70 and CIN70 gene sets, but not the HET70 gene set, sort human tissues according to their level of PCNA expression. (A-C) The CIN70, TRI70, and HET70 gene scores were calculated for 81 non-diseased human tissues and cell populations and plotted against the PCNA levels in each sample. The rank order correlations between CIN70 scores, TRI70 scores, and PCNA are significant, suggesting that these gene signatures can sort cells according to their proliferative index. However, HET70 is uncorrelated with PCNA, suggesting that it does not report proliferation. (D) Gene expression data from 81 non-diseased human tissues and cell populations were sorted according to their level of PCNA expression. A clustergram of HET70 gene expression levels, sorted according to the level of PCNA expression in each tissue, is displayed.

Prognostic relevance of aneuploidy and karyotype heterogeneity gene sets in cancer

I next assembled 27 cancer gene expression datasets from published clinical cohorts. Datasets were chosen to represent a variety of solid tumor types, microarray designs, and outcome measurements (i.e., overall survival, recurrence-free survival, etc.). The PCNA25, CIN70, TRI70, and HET70 gene sets were used to stratify patients into "above mean" and "below mean" groups in each clinical cohort, and Kaplan-Meier survival curves were calculated for each gene set (Fig. 4, Table S18, and Fig. S6). I found that high PCNA25 was significantly associated with poor survival in 15 out of 27 cohorts. PCNA25 was particularly informative in breast cancers (significant in 6/6 cohorts), brain cancers (significant in 4/4 cohorts), and bladder cancers (significant in 2/3 cohorts). PCNA25 did not provide significant patient stratification in lung cancer (1/5 cohorts), ovarian cancer (0/3 cohorts), or colorectal cancer (0/3 cohorts). These results are consistent with the known value of proliferation markers in breast cancer (47,48), and suggest that similar gene sets may be useful for other cancer types as well. In cancer types in which PCNA25 was not predictive, other facets of tumor biology, including immune evasion and metastatic potential, may be of greater importance for patient survival than proliferation rates.

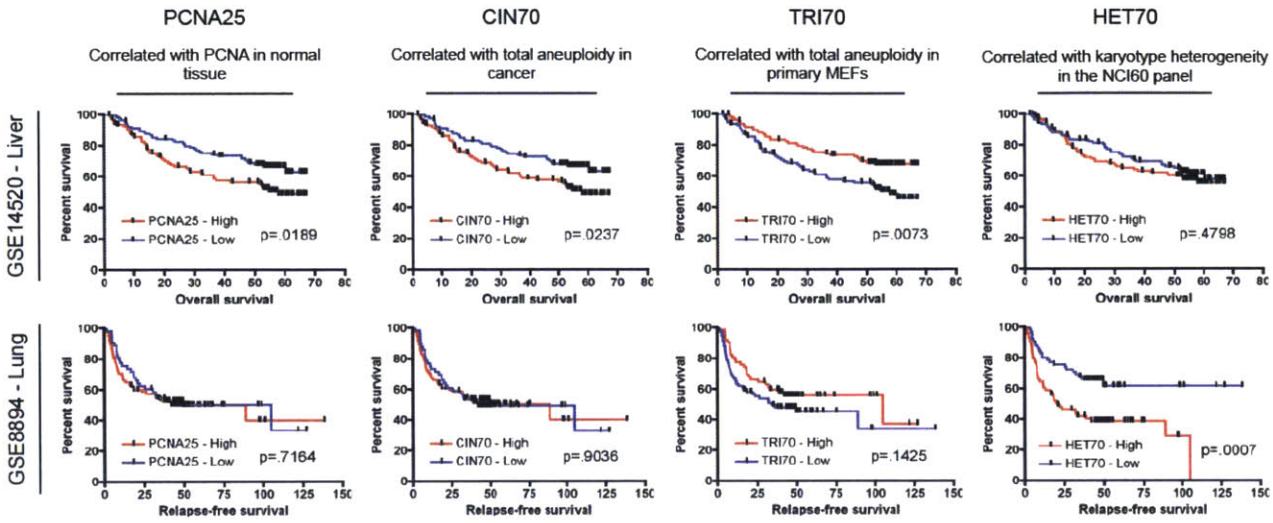


Figure 4. Stratification of patient risk based on proliferation and aneuploidy gene signatures. Patients from each clinical cohort were divided into below-mean and above-mean subsets for each gene signature, and significant differences in survival time were identified using a logrank test. Representative Kaplan-Meier curves for GSE14520 and GSE8894 are displayed. In general, PCNA25, CIN70, and TRI70 were able to classify the same patient cohorts, while HET70 classified some cohorts in which proliferation-related markers were not prognostic. The X axes indicate survival time in months. Complete results are presented in Table S18 and Figure S6.

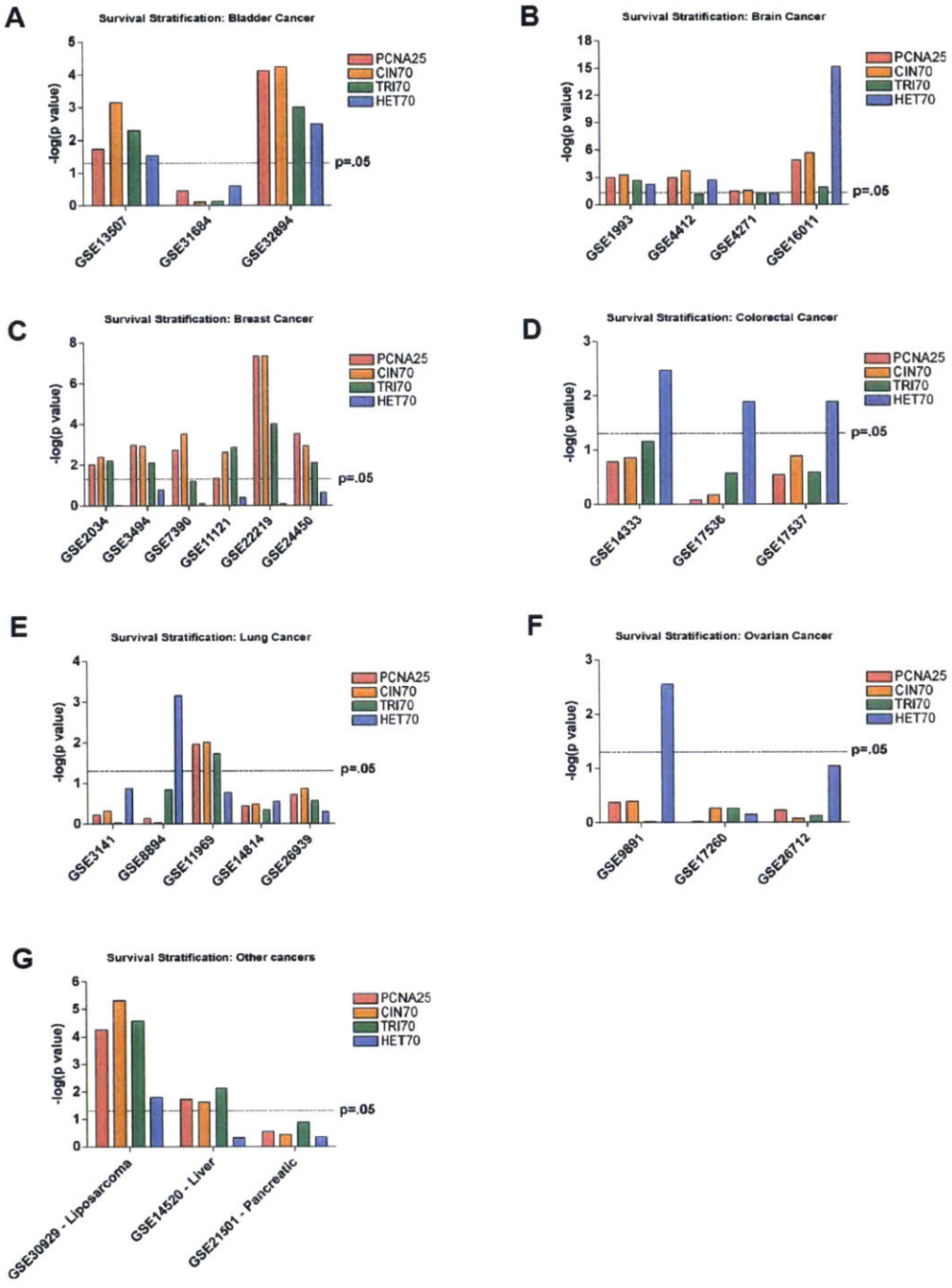


Figure S6. Patient stratification in 27 tumor gene expression cohorts. Patients from each clinical cohort were divided into below-mean and above-mean subsets for each gene signature, and p values for each comparison were calculated using a logrank test. The dotted line indicates stratification at a p=.05 significance level. Hazard ratios and further clinical information and provided in Table S18.

Subsequently, I found that CIN70 provided significant stratification of patient risk in 15 out of 27 cohorts (Fig. S6). Importantly, every cohort in which CIN70 was a significant predictor of survival was one in which PCNA25 was predictive as well. Across all datasets, the hazard ratios generated by CIN70 and PCNA25 were highly correlated (Fig. S7A; ρ=.98, $p<10^{-18}$). Despite having only 12 genes in common, the individual patient scores generated by CIN70 and PCNA25 were nearly identical (Fig. S7; ρ=0.97-0.99). I conclude that CIN70 functions primarily by reporting proliferative capacity, and that, in general, CIN70 does not provide any information beyond that which is provided by a smaller number of PCNA-associated markers.

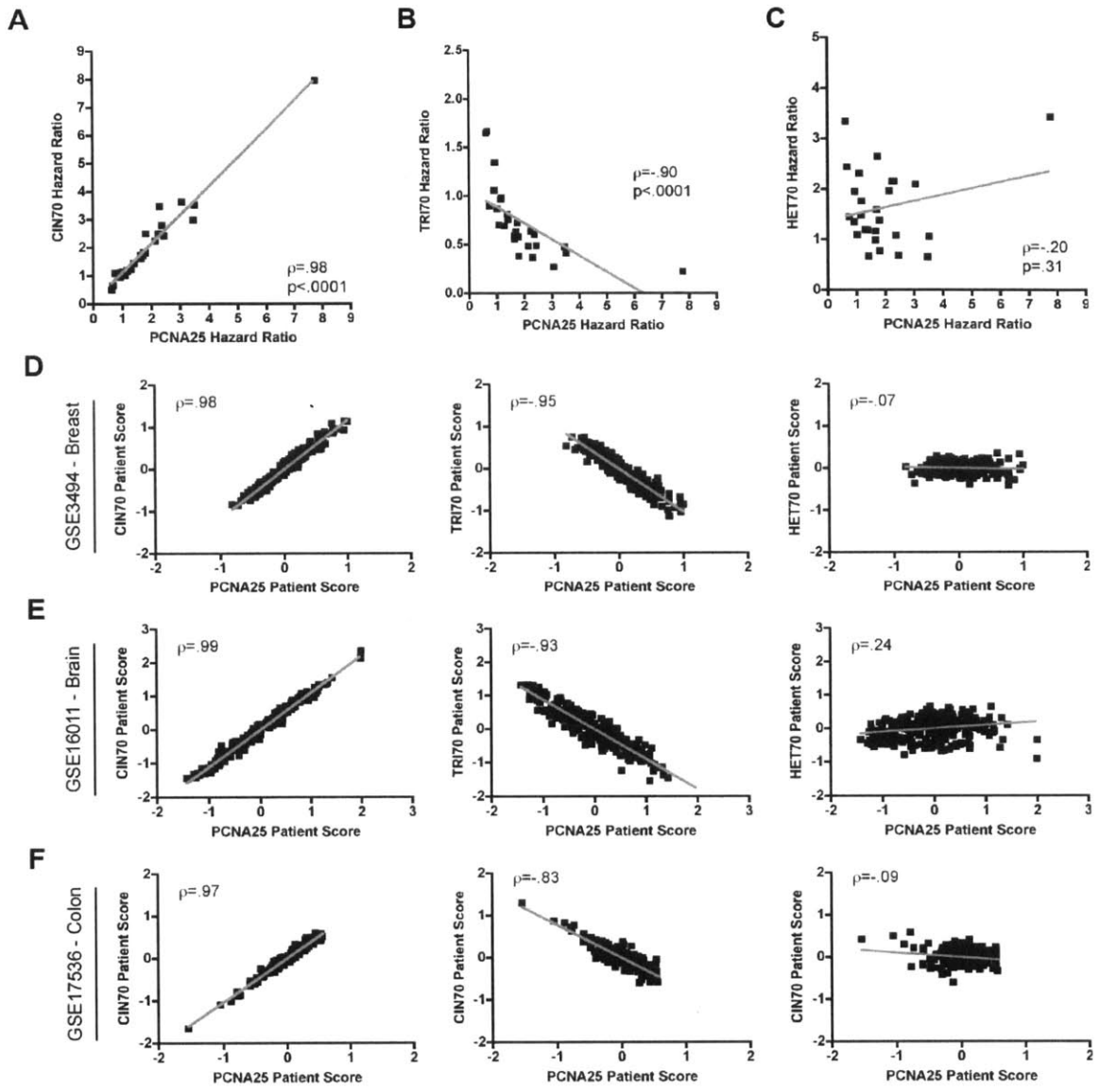


Figure S7. Classification similarities between PCNA25, CIN70, and TRI70. (A-C) PCNA25 hazard ratios from each clinical cohort were compared to hazard ratios generated by CIN70, TRI70, or HET70. While PCNA25, CIN70, and TRI70 generated very similar hazard ratios for each cohort, the hazard ratios generated by HET70 were uncorrelated with those generated by PCNA25. (D-F) Individual patient scores generated by PCNA25 were compared to patient scores generated by CIN70, TRI70, or HET70. While PCNA25 and CIN70 patient scores were nearly identical, HET70 scores were poorly correlated or uncorrelated with PCNA25.

TRI70 was found to be a significant predictor of survival in 12 out of 27 patient cohorts. Unlike PCNA25 and CIN70, high TRI70 was associated with favorable outcomes and low TRI70 was associated with disease progression and death. All cohorts in which TRI70 was predictive were also stratified by PCNA25, and hazard ratios calculated using the two gene signatures were highly correlated (Fig. S7B; ρ=-.90, $p<10^{-10}$). Thus, consistent with our above analysis, the most prominent gene expression changes caused by aneuploidy in primary cells are directly related to slow proliferation, which can predict survival in various cancer types.

High HET70 scores were significantly associated with poor prognosis in 12 out of 27 patient cohorts. Surprisingly, 5 of these cohorts were not stratified by PCNA25: HET70 was able to stratify colorectal cancer (3/3 patient cohorts), which PCNA25 failed to do. Across all datasets the hazard ratios generated by HET70 and PCNA25 were uncorrelated (Fig S7C), and individual patient scores generated by HET70 and PCNA25 were poorly correlated or uncorrelated (Fig. S7D-S7F). I conclude that the gene expression changes associated with chromosomal instability *in vitro* can serve as proliferation-independent markers for poor prognosis *in vivo*. Though I cannot ascertain at this time whether HET70 scores stratify patients according to the chromosomal instability of their tumors, I hypothesize that chromosomal instability is one mechanism by which the increased expression of these genes can arise.

HET5: A five gene signature with prognostic relevance in multiple cancer types

In order to narrow the HET70 gene signature into a smaller set of genes which may have clinical relevance, I calculated univariate Cox proportional hazard scores for each HET70 gene in each cancer dataset. I then sorted the HET70 genes by their average Z-score across the bladder, brain, colon, lung, and ovarian cancer datasets. The five genes with the highest average Z-scores were *LGALS1*, *LEPRE1*, *FN1*, *CALU*, and *PLOD2*, which I denote as the HET5 gene set. Among the cohorts examined above, these five genes were able to stratify

patients into low-risk and high-risk subgroups in 2/3 bladder cancer cohorts, 4/4 brain cancer cohorts, 3/3 colorectal cancer cohorts, 1/1 liposarcoma cohort, 1/1 liver cancer cohort, 3/5 lung cancer cohorts, and 1/3 ovarian cancer cohorts (Table S19). I then assembled a second set of published clinical expression arrays which were not used in the initial derivation of the HET5 signature. Within this test set, HET5 provided significant stratification in 8/15 cohorts, including 1/1 bladder cancer cohort, 2/2 brain cancer cohorts, and 2/4 lung cancer cohorts (Fig. S8 and Table S20). Remarkably, HET5 was also able to stratify patient risk in 3/4 melanoma cohorts, despite the fact that no melanoma datasets were present in the original training sets. Thus, the HET5 genes in particular may warrant further investigation as proliferation-independent markers of cancer progression and patient survival.

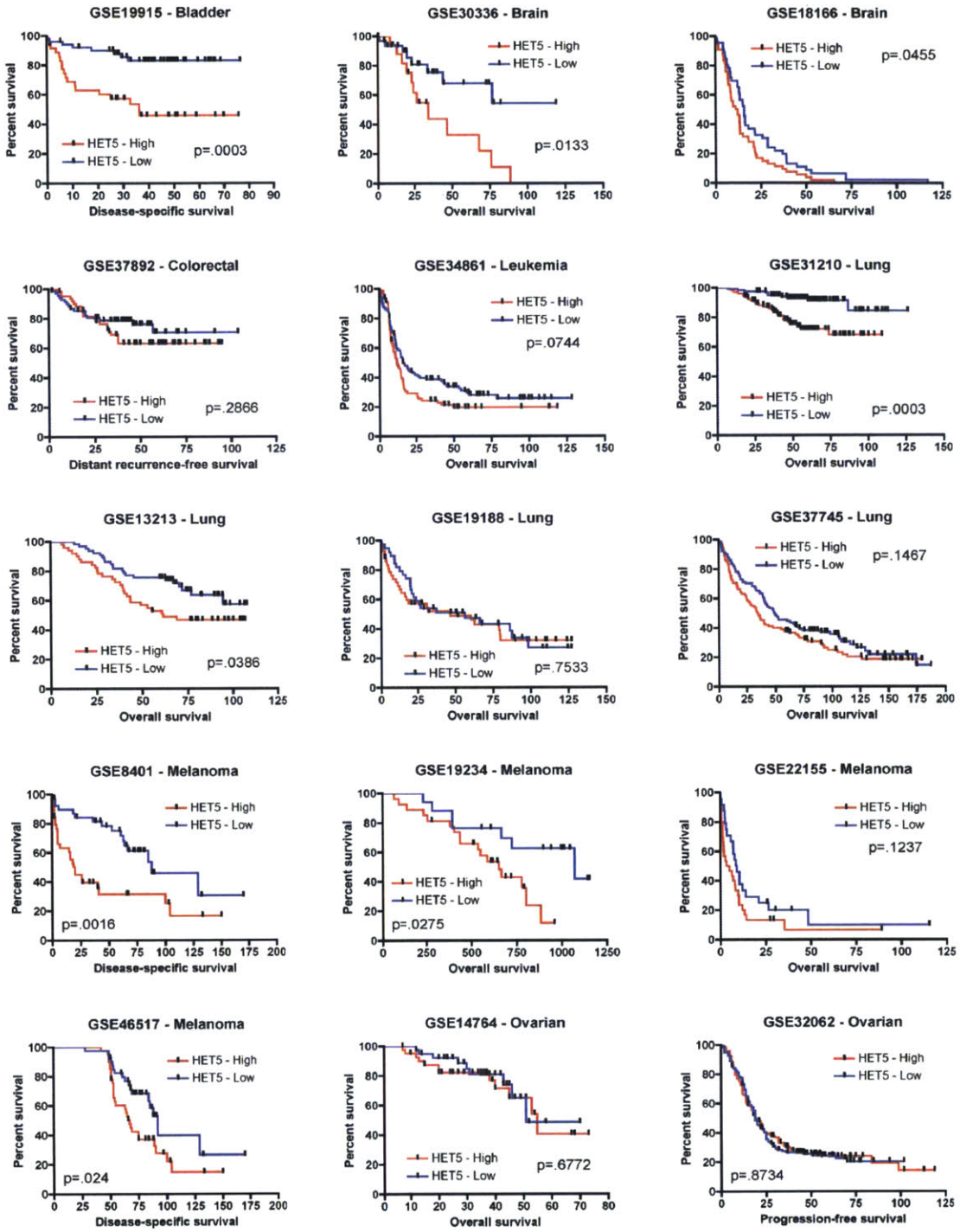


Figure S8. Kaplan-Meier analysis using the HET5 gene set. Patients from the indicated clinical cohorts were divided into above-mean and below-mean subsets, and significant differences in survival time were identified using a logrank test. The X axes indicate survival time in months.

Discussion

Many cancers display chromosomal instability, defined as an increased rate of chromosome missegregation (45). Our results suggest that aneuploidies resulting from chromosome segregation errors impose a fitness cost on cancer, *in vitro* if not also *in vivo* [Fig. 1, Fig. 2, and (22)]. CIN cancer cell lines divide more slowly than chromosomally-stable lines, downregulate transcripts associated with progression through the cell cycle, and upregulate stress- and signaling-related genes [Fig. 2A, Table S2, and (36)]. A similar transcriptional response is evident in CIN primary cells and in a stable cancer cell line that has been induced to gain extra chromosomes (Fig. S3). Additionally, CIN cancers appear to have increased metabolic requirements: they consume more glutamine and glucose per cell, and produce more citric acid cycle byproducts, than stable cancer cell lines (Fig. 3). Thus, in these respects, chromosomally-unstable cancer cells resemble the stressed state of primary aneuploid cells.

Yet, these stresses are not apparent in all aneuploid cancers. When normalized to diploid tumors, aneuploid tumors display transcriptional patterns that are opposite those observed in untransformed aneuploid cells. Furthermore, increasing levels of clonal aneuploidy within specific tumor types are tightly correlated with markers of rapid proliferation (Fig. S5). To reconcile these results, I hypothesize that there are two "types" of aneuploidy that are common to cancer cells: clonal aneuploidy, which is present in the bulk of a tumor and arises due to the selective advantages that it confers, and spontaneous aneuploidy, which results from chromosome missegregation and generally induces fitness costs on a cell.

The CIN70 gene signature was initially constructed by identifying genes that correlated most strongly with clonal aneuploidy in cancer: tumors with high CIN70 scores were the most structurally complex, while tumors with low CIN70 scores displayed the fewest alterations from the diploid state (41). However, our findings demonstrate that CIN70 is a marker of cell proliferation, rather than chromosomal instability: 100% of CIN70 genes are significantly

correlated with PCNA expression in normal tissue, and CIN70 scores accurately differentiate between rapidly dividing and slowly dividing karyotypically-normal human tissue (Fig. S5). These results can be explained by hypothesizing that structural complexity in tumors results from an ongoing evolutionary process. Genetic alterations, including mutations, deletions, and chromosomal duplications, arise during a cancer's growth. Those changes that induce a proliferative advantage are selected within the tumor and rise to clonal levels. While aneuploidy is generally associated with a fitness cost, the few chromosomal alterations that do confer proliferative advantages – such as the gain of a chromosome arm carrying an oncogene – are naturally selected and become the dominant tumor cell population over time. Thus, tumors with more aneuploid regions are likely to have acquired more growth-promoting genetic alterations, explaining the tight link between CIN70 and cell proliferation.

In contrast, aneuploidies resulting from random segregation errors have not been directly selected for. Whole-chromosome aneuploidy alters the copy numbers of hundreds or thousands of genes after a single mitotic event. Human cells lack a global dosage compensation system for autosomal aneuploidy; most copy number variation has proportional effects on the levels of mRNA's and proteins encoded on the aneuploid chromosome (29). Thus, it has been hypothesized that aneuploidy induces a fitness cost by wasting energy on the transcription, translation, and degradation of unnecessary proteins, and by interfering with the formation or function of stoichiometry-sensitive protein complexes (49). Consistent with the results presented here, we have shown that both primary aneuploid cells and chromosomally-unstable (but not chromosomally-stable) cancers are sensitive to energy stress- and proteotoxic stress-inducing compounds (12). Note that I do not intend to suggest that CIN is always antagonistic to robust proliferation. Some mutations – such as loss of p53 – can increase both proliferation and chromosomal instability (50). As p53 loss is a late event in many cancers (51–53), aggressive stage III or stage IV tumors may display both CIN and heightened mitotic

activity. Nonetheless, the overall transcription patterns observed in CIN cells, or stable cells that have been induced to gain chromosomes, are consistent with unselected aneuploidy placing a substantial burden on cellular homeostasis.

The stressed state of CIN cells may have further relevance for its role in tumor development and evolution. Chromosomal instability has been linked with drug resistance and cancer relapse (54,55). The ability of CIN cells to generate genomically-heterogeneous progeny can likely contribute to a tumor's ability to evolve and metastasize despite treatment. However, as genotoxic chemotherapy selectively kills rapidly-dividing cells, it may also be the case that CIN cells escape elimination due to their inherently stressed state and slow growth. The physiological differences between chromosomally-stable and unstable cells could be exploited to develop drugs that selectively kill high-CIN cells, which may be added to existing chemotherapy regimens to minimize relapse (12).

More broadly, there is growing interest in the use of genomic technology to stratify patient risk and identify appropriate treatments for cancer (56). While many published gene signatures have demonstrated the ability to identify aggressive tumors, further analysis has established that some of these signatures are prognostic only due to their ability to detect proliferation rates (46). Indeed, I have confirmed here that proliferation, measured by PCNA25, is a significant but not universal risk factor in diverse cancers, and that one previously-published signature (CIN70) is tightly coupled to proliferative index. However, I demonstrate that HET70 - the 70 genes most strongly upregulated in karyotypically heterogeneous cancers - function as a proliferation-independent risk factor in several different patient cohorts. How strongly these genes correlate with CIN *in vivo* is at present unknown, and I suspect that CIN is one of many factors that can drive their expression. Furthermore, the HET70 signature can be narrowed to five genes which maintain significant prognostic utility in multiple cancer types. The HET5 gene set includes *FN1*, encoding fibronectin, and *LGALS1*, encoding Galectin-1, which have

extensive roles in cancer growth, metastasis, and immune evasion (57–60). Three other genes which comprise HET5 (*CALU*, *PLOD2*, and *LEPRE1*) have functions that are less well-understood. Further analysis of these genes may shed light on the clinical progression of diverse cancers. Moreover, these findings provide a mechanistic link between CIN and aggressive disease, and suggest genes and biological processes that could be targeted to contravene CIN-induced cancer progression.

Materials and Methods

Data Sources. Gene expression data from trisomic MEFs were acquired from (9), and the set of probes classified as "expressed" were used in this study. Gene expression data from the NCI60 cell lines analyzed on HG-U133A arrays were downloaded from CellMiner (24). Probes were updated using Release 32 of the NetAffx probe annotations (25). Karyotype heterogeneity values were acquired from (26), and missing values were kindly provided by A. Roschke (NCI, Bethesda, MD). Doubling times of the NCI60 panel were acquired from (27). Gene expression values from cancer/normal tissue pairs were acquired from the Gene Expression Omnibus (GSE15641, GSE27943, GSE3167, GSE3189, and GSE5364). Gene expression values from breast tumors of known ploidy were acquired from Gene Expression Omnibus (GSE12071). Consumption/release rates of metabolites in the NCI60 panel were acquired from (28). Gene expression values from tetrasomic HCT116 cells were acquired from (29). Gene expression values from normal human tissues were acquires from the Gene Expression Omnibus (GSE1133). Gene expression values from clinical cancer specimens were acquired from the sources cited in Table S18 and S20.

Mouse Husbandry and MEF Culture. All animal studies and procedure were approved by the MIT Institutional Animal Care and Use Committee. $Cdc20^{AAA}$ mice were kindly provided by Dr. Pumin Zhang (Baylor College of Medicine). $BubR1^{H}$ mice were kindly provided by Dr. Jan van Deursen (Mayo Clinic). Embryos were derived via timed matings between heterozygous animals. Embryos were harvested on day E12.5 for $Cdc20^{AAA}$ crosses and E13.5-E14.5 for $BubR1^{H}$ crosses. MEFs were isolated and cultured as described in (9).

MEF Microarrays. RNA was isolated from early passage MEFs using Trizol (Invitrogen) followed by purification with RNeasy Mini Kits (Qiagen). cDNA was constructed and labeled using the Ovation RNA Amplification System V2 and the FL-Ovation cDNA Biotin Module V2 (NuGEN) according to the manufacturer's specifications. cDNA was hybridized to Mouse 430A 2.0 Arrays (Affymetrix) for 16 hours and then scanned on an Affymetrix GeneChip scanner according to the manufacturer's specifications. Data was summarized and normalized using gcRMA, and has been deposited in GEO (GSE49894).

Data Analysis. Analysis was performed in Excel, Python, and R. For each microarray experiment, probesets were log2-transformed and then normalized by subtracting the average expression level of that probeset. Probesets annotated to the same gene were collapsed by averaging. For primary MEFs and tetrasomic HCT116 cells, aneuploid chromosomes were excluded from analysis. For most samples, genes annotated to sex chromosomes were excluded from analysis, and, on the Affymetrix platform, non-specific probesets (those annotated with an "x") were excluded from analysis. The exception to this is the CIN70 gene signature in Fig. S4, in which some of the genes are present on sex chromosomes or are only measured by non-specific probesets.

GO term enrichment analysis was performed using GProfiler with a Benjamini-Hochberg corrected p value of .05 and a maximum p value of 10^{-2} (30). Enrichments were performed against the relevant background gene set, i.e. against all genes not present on an aneuploid or sex chromosome. Mouse and human orthologs were identified using GProfiler. Only genes that displayed a one-to-one orthology relationship between mouse and human were utilized for analysis. Gene Set Enrichment Analysis was performed using a Pearson ranking metric, a gene set size range of 5 to 5000, and 1000 permutations of gene set labels (31). Transcription factor motif enrichments were identified using MSigDB (32). Metabolic pathway enrichments were identified with MetaboAnalyst using a maximum P value of 10^{-2} (33). Gene sets were clustered in Cluster 3.0 (34) and visualized in Java Treeview (35). For single-gene correlation analysis, a cutoff of >0.3 or <-0.3 was uniformly applied. For two-sample comparisons, a combined t test ($p<.05$) and mean fold change (>1.3 or <-1.3) threshold was used to identify differentially-expressed genes. As only the mean mRNA expression values were reported for the HCT116+5 cells, genes expressed ±1SD from the overall mean were considered differentially expressed.

Analysis of the NCI60 panel of cancer cell lines. Data from 57 NCI60 cell lines were used for this analysis. Lung cancer line H23 was excluded as it was not profiled on HG-U133A arrays, MDA-N was excluded as it lacked karyotype information, and NCI-ADR-RES was excluded as it is a derivative of OVCAR-8. To identify transcripts whose expression correlated with chromosomal instability, I used a strategy similar to that described in (36). Levels of chromosomal instability were determined using each cell line's index of numerical heterogeneity (INH), as reported in (26). INH values were derived from the cell-to-cell variation in the copy number of each chromosome, as determined by spectral karyotype analysis of metaphase spreads. A chromosome whose copy number varied in different cells was judged to be

heterogeneous in that cell line. (Chromosomes that appeared to be lost in two or fewer metaphase spreads or gained in a single metaphase spread were not counted as heterogeneous, due to the possibility that these variations were experimental artefacts). Fractional index of numerical heterogeneity values ranged from 0 (no chromosomes exhibited varying copy numbers) to 1 (every chromosome exhibited varying copy numbers), in multiples of 1/23, as the X and Y chromosomes were grouped together for this analysis. Lastly, to exclude genes that showed tissue-of-origin-dependent expression patterns, I performed a 9-way ANOVA in TM4 MeV (37). A Bonferroni-corrected p value of .05 was used to exclude genes that showed tissue-specific expression patterns.

Analysis of clinical gene expression data. Gene expression and survival data were downloaded from GEO and normalized as described above. Probeset annotations were acquired from GEO. PCNA25, CIN70, and HET70 gene scores were calculated by averaging the expression levels of these genes for each patient. TRI70 gene scores were calculated by subtracting the average of the genes that are downregulated by primary aneuploidy from the average of the genes that are upregulated by primary aneuploidy. For each patient cohort, samples were divided into "high expression" and "low expression" classes depending on whether the gene set scores were above or below the mean, respectively, for each gene set. P values and hazard ratios were determined using the survival package in R or in Prism. Cox proportional hazard models were constructed using the coxph function in R.

Acknowledgments

I thank Joan Clara Smith for coding assistance, Pumin Zhang and Jan van Deursen for providing mice, and Erica Chung for help with MEF culture. I am grateful to Angelika Amon, Maitreya Dunham, Mike Laub, Aviv Regev, Yitzhak Pilpel, and members of the Amon lab for comments on the manuscript.

References

1. Holland AJ, Cleveland DW. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. Nat Rev Mol Cell Biol. 2009;10:478–87.

2. Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, et al. Aneuploidy Drives Genomic Instability in Yeast. Science. 2011;333:1026 –1030.

3. Thompson SL, Compton DA. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. J. Cell Biol. 2010;188:369–81.

4. Duesberg P, Rausch C, Rasnick D, Hehlmann R. Genetic instability of cancer cells is proportional to their degree of aneuploidy. PNAS. 1998;95:13692 –13697.

5. Zhu J, Pavelka N, Bradford WD, Rancati G, Li R. Karyotypic Determinants of Chromosome Instability in Aneuploid Budding Yeast. PLoS Genet. 2012;8:e1002719.

6. Weaver BA, Cleveland DW. Does aneuploidy cause cancer? Current Opinion in Cell Biology. 2006;18:658–67.

7. McGranahan N, Burrell RA, Endesfelder D, Novelli MR, Swanton C. Cancer chromosomal instability: therapeutic and diagnostic challenges. EMBO reports. 2012;13:528–38.

8. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, et al. Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. Science. 2007;317:916–24.

9. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, et al. Aneuploidy Affects Proliferation and Spontaneous Immortalization in Mammalian Cells. Science. 2008;322:703–9.

10. Thompson SL, Compton DA. Examining the link between chromosomal instability and aneuploidy in human cells. J Cell Biol. 2008;180:665–72.

11. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, et al. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature. 2010;468:321–5.

12. Tang Y-C, Williams BR, Siegel JJ, Amon A. Identification of Aneuploidy-Selective Antiproliferation Compounds. Cell. 2011;144:499–512.

13. Nižetić D, Groet J. Tumorigenesis in Down's syndrome: big lessons from a small chromosome. Nature Reviews Cancer. 2012;12:721–32.

14. Sotillo R, Hernando E, Díaz-Rodríguez E, Teruya-Feldstein J, Cordón-Cardo C, Lowe SW, et al. Mad2 Overexpression Promotes Aneuploidy and Tumorigenesis in Mice. Cancer Cell. 2007;11:9–23.

15. Sotillo R, Schvartzman J-M, Socci ND, Benezra R. Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. Nature. 2010;464:436–40.

16. Li M, Fang X, Baker DJ, Guo L, Gao X, Wei Z, et al. The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. Proceedings of the National Academy of Sciences. 2010;107:14188–93.

17. Weaver BAA, Silk AD, Montagna C, Verdier-Pinard P, Cleveland DW. Aneuploidy Acts Both Oncogenically and as a Tumor Suppressor. Cancer Cell. 2007;11:25–36.

18. Baker DJ, Jin F, Jeganathan KB, van Deursen JM. Whole Chromosome Instability Caused by Bub1 Insufficiency Drives Tumorigenesis through Tumor Suppressor Gene Loss of Heterozygosity. Cancer Cell. 2009;16:475–86.

19. Remeseiro S, Cuadrado A, Carretero M, Martínez P, Drosopoulos WC, Cañamero M, et al. Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. EMBO J. 2012;31:2076–89.

20. Birkbak NJ, Eklund AC, Li Q, McClelland SE, Endesfelder D, Tan P, et al. Paradoxical Relationship between Chromosomal Instability and Survival Outcome in Cancer. Cancer Research. 2011;71:3447 –3452.

21. Zasadil LM, Britigan EMC, Weaver BA. 2n or not 2n: Aneuploidy, polyploidy and chromosomal instability in primary and tumor cells. Semin. Cell Dev. Biol. 2013;

22. Sheltzer JM, Torres EM, Dunham MJ, Amon A. Transcriptional consequences of aneuploidy. PNAS. 2012;109:12644–9.

23. Ried T, Hu Y, Difilippantonio MJ, Ghadimi BM, Grade M, Camps J. The consequences of chromosomal aneuploidy on the transcriptome of cancer cells. Biochim. Biophys. Acta. 2012;1819:784–93.

24. Shankavaram UT, Varma S, Kane D, Sunshine M, Chary KK, Reinhold WC, et al. CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. BMC Genomics. 2009;10:277.

25. Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, et al. NetAffx: Affymetrix probesets and annotations. Nucl. Acids Res. 2003;31:82–6.

26. Roschke AV, Tonon G, Gehlhaus KS, McTyre N, Bussey KJ, Lababidi S, et al. Karyotypic complexity of the NCI-60 drug-screening panel. Cancer Res. 2003;63:8634–47.

27. NCI60 Discovery Services. Available from: http://dtp.nci.nih.gov/docs/misc/common_files/cell_list.html

28. Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, Souza AL, et al. Metabolite Profiling Identifies a Key Role for Glycine in Rapid Cancer Cell Proliferation. Science. 2012;336:1040–4.

29. Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Molecular Systems Biology. 2012;8.

30. Reimand J, Arak T, Vilo J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Research. 2011;39:W307–W315.

31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. 2005;102:15545–50.

32. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27:1739 –1740.

33. Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. Nucl. Acids Res. 2012;

34. De Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004;20:1453–4.

35. Saldanha AJ. Java Treeview—extensible visualization of microarray data. Bioinformatics. 2004;20:3246 –3248.

36. Roschke AV, Glebov OK, Lababidi S, Gehlhaus KS, Weinstein JN, Kirsch IR. Chromosomal Instability Is Associated with Higher Expression of Genes Implicated in Epithelial-Mesenchymal Transition, Cancer Invasiveness, and Metastasis and with Lower Expression of Genes Involved in Cell Cycle Checkpoints, DNA Repair, and Chromatin Maintenance. Neoplasia. 2008;10:1222–30.

37. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. BioTechniques. 2003;34:374–8.

38. Karlsson E, Delle U, Danielsson A, Olsson B, Abel F, Karlsson P, et al. Gene expression variation to predict 10-year survival in lymph-node-negative breast cancer. BMC Cancer. 2008;8:254.

39. Baker DJ, Jeganathan KB, Cameron JD, Thompson M, Juneja S, Kopecka A, et al. BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. Nat Genet. 2004;36:744–9.

40. Li M, Fang X, Wei Z, York JP, Zhang P. Loss of spindle assembly checkpoint-mediated inhibition of Cdc20 promotes tumorigenesis in mice. The Journal of Cell Biology. 2009;185:983–94.

41. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nature Genetics. 2006;38:1043–8.

42. Swanton C, Nicke B, Schuett M, Eklund AC, Ng C, Li Q, et al. Chromosomal instability determines taxane response. Proceedings of the National Academy of Sciences. 2009;106:8671–6.

43. Salvatore G, Nappi TC, Salerno P, Jiang Y, Garbi C, Ugolini C, et al. A Cell Proliferation and Chromosomal Instability Signature in Anaplastic Thyroid Carcinoma. Cancer Res. 2007;67:10148–58.

44. Decaux O, Lodé L, Magrangeas F, Charbonnel C, Gouraud W, Jézéquel P, et al. Prediction of Survival in Multiple Myeloma Based on Gene Expression Profiles Reveals Cell Cycle and Chromosomal Instability Signatures in High-Risk Patients and Hyperdiploid Signatures in Low-Risk Patients: A Study of the Intergroupe Francophone du Myélome. JCO. 2008;26:4798–805.

45. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature. 1998;396:643–9.

46. Venet D, Dumont JE, Detours V. Most Random Gene Expression Signatures Are Significantly Associated with Breast Cancer Outcome. PLoS Comput Biol. 2011;7:e1002240.

47. Dai H, Veer L van't, Lamb J, He YD, Mao M, Fine BM, et al. A Cell Proliferation Signature Is a Marker of Extremely Poor Outcome in a Subpopulation of Breast Cancer Patients. Cancer Res. 2005;65:4059–66.

48. Mosley JD, Keri RA. Cell cycle correlated genes dictate the prognostic power of breast cancer gene lists. BMC Med Genomics. 2008;1:11.

49. Sheltzer JM, Amon A. The aneuploidy paradox: costs and benefits of an incorrect karyotype. Trends Genet. 2011;27:446–53.

50. Levine AJ. p53, the cellular gatekeeper for growth and division. Cell. 1997;88:323–31.

51. Baker SJ, Preisinger AC, Jessup JM, Paraskeva C, Markowitz S, Willson JKV, et al. p53 Gene Mutations Occur in Combination with 17p Allelic Deletions as Late Events in Colorectal Tumorigenesis. Cancer Res. 1990;50:7717–22.

52. Lassam NJ, From L, Kahn HJ. Overexpression of p53 Is a Late Event in the Development of Malignant Melanoma. Cancer Res. 1993;53:2235–8.

53. Boyle JO, Hakim J, Koch W, Riet P van der, Hruban RH, Roa RA, et al. The Incidence of p53 Mutations Increases with Progression of Head and Neck Cancer. Cancer Res. 1993;53:4477–80.

54. Lee AJX, Endesfelder D, Rowan AJ, Walther A, Birkbak NJ, Futreal PA, et al. Chromosomal Instability Confers Intrinsic Multidrug Resistance. Cancer Res. 2011;71:1858–70.

55. Bakhoum SF, Danilova OV, Kaur P, Levy NB, Compton DA. Chromosomal Instability Substantiates Poor Prognosis in Patients with Diffuse Large B-cell Lymphoma. Clin Cancer Res. 2011;17:7704–11.

56. Raspe E, Decraene C, Berx G. Gene expression profiling to dissect the complexity of cancer biology: Pitfalls and promise. Seminars in Cancer Biology. 2012;22:250–60.

57. Rabinovich GA. Galectin-1 as a potential cancer target. British Journal of Cancer. 2005;92:1188–92.

58. Dalotto-Moreno T, Croci DO, Cerliani JP, Martinez-Allo VC, Dergan-Dylon S, Huergo SPM, et al. Targeting galectin-1 overcomes breast cancer associated immunosuppression and prevents metastatic disease. Cancer Res

59. Banh A, Zhang J, Cao H, Bouley DM, Kwok S, Kong C, et al. Tumor Galectin-1 Mediates Tumor Growth and Metastasis through Regulation of T-Cell Apoptosis. Cancer Research. 2011;71:4423–31.

60. Ruoslahti E. Fibronectin and Its Integrin Receptors in Cancer. In: George F. Vande Woude and George Klein, editor. Advances in Cancer Research. 1999;1-20.

Chapter 5: Single-chromosome aneuploidy commonly functions as a tumor suppressor

Sheltzer JM, Ko J, Habibe Burgos NC, Chung E, Meehl CM, et al.

The experiments in Figures 5 and 6 were performed by JK.
All other experiments were performed by JS.

Abstract

Whole-chromosome aneuploidy is a hallmark of human malignancies. The prevalence of chromosome segregation errors in cancer – first noted more than 100 years ago – has led to the widespread belief that aneuploidy plays a crucial role in tumor development. Here, we set out to test this hypothesis. We transduced congenic euploid and trisomic primary fibroblasts with 14 different oncogenes or oncogene combinations, thereby creating genetically-identical cancer cell lines that differed only in karyotype. Surprisingly, most aneuploid cell lines divided slowly *in vitro*, formed few colonies in soft agar, and grew poorly as xenografts, relative to their matched euploid strains. Similar results were obtained when comparing a euploid human colorectal cancer cell line with derivatives of that line that harbored extra chromosomes. Only a few aneuploid strains grew at near-wild-type levels, and no aneuploid strain exhibited greater tumorigenic capabilities than its euploid counterpart. These results demonstrate that, rather than promoting tumorigenesis, aneuploidy can very often function as a tumor suppressor.

Introduction

In the early 20th century, Theodor Boveri proposed that abnormal karyotypes altered the equilibrium between pro- and anti-proliferative cellular signals, and were therefore capable of transforming primary cells into cancer cells (1). "Boveri's hypothesis" was one of the first genetic explanations of cancer development, and it helped motivate a century of research into the origins and consequences of chromosome segregation errors. Since Boveri's time, it has been established that approximately 90% of solid tumors and 75% of hematopoietic cancers display whole-chromosome aneuploidy (2). However, the precise relationship between aneuploidy and tumorigenesis remains unclear.

A preponderance of current evidence supports Boveri's hypothesis (3,4). First, individuals with Down syndrome (trisomy 21) frequently develop pediatric leukemia, suggesting a clear link between the gain of chromosome 21 and leukemogenesis (5). Secondly, many human cancers exhibit recurrent aneuploidies (6,7), and computational modeling has suggested that these patterns of chromosomal alterations reflect an evolutionary process in which cancer cells increase the copy number of loci encoding oncogenes and decrease the copy number of loci encoding tumor suppressors (8). Finally, genetically-engineered mice that harbor alleles which cause chromosomal instability (CIN) typically develop tumors at accelerated rates (9–14), particularly when combined with mutations in the p53 tumor suppressor (15). Nonetheless, several observations suggest that the relationship between aneuploidy and cancer may be more complex than previously believed. While individuals with Down syndrome are at an increased risk of developing leukemia and germ cell tumors, they are at a significantly decreased risk of developing many other common solid tumors (16). Trisomies of regions orthologous to human chromosome 21 in the mouse have also been found to suppress tumor development (17–19). Moreover, though mouse models of CIN are generally tumor-prone, in certain organs or when combined with certain oncogenic mutations, CIN mice exhibit reduced tumor burdens (20,21). Thus, aneuploidy may have tumor-protective as well as tumor-promoting consequences.

In order to further our understanding of the effects of aneuploidy on cell and organismal physiology, our lab and several others have developed systems to generate primary cells with a range of aneuploid karyotypes (22–27). These cells have been constructed without CIN-promoting mutations, thereby allowing us to study aneuploidy absent other genetic perturbations. This research has demonstrated the existence of a set of phenotypes that are shared among many different aneuploid cells and are largely independent of the specific chromosomal alteration: aneuploid cells display reduced fitness (22–24), are deficient at maintaining proteostasis (28,29), and exhibit a specific set of gene expression changes that include the down-regulation of cell-cycle transcripts and the up-regulation of a stress-response program (30–32). A crucial question, however, is in what way(s) the cellular changes induced by aneuploidy affect (and possibly drive) tumorigenesis. Aneuploid cells may be poised on the brink of transformation due to their increased dosage of oncogenes and decreased dosage of tumor suppressors (8), the inherent instability of aneuploid genomes (33), or due to a general misregulation of cell metabolism and other biological processes (34). To test these ideas, we compared the tumorigenicity of a series of genetically-matched euploid and aneuploid cells. Surprisingly, we found that nearly every aneuploid cell line that we examined displayed reduced tumor-forming potential relative to control euploid cell lines. These results necessitate a significant revision of our understanding of the relationship between aneuploidy and cancer.

Results

Single-chromosome aneuploidy is insufficient to induce neoplastic phenotypes

We took advantage of naturally-occurring Robertsonian translocations to generate mouse embryonic fibroblasts (MEFs) trisomic for chromosome 1, 13, 16, or 19, as well as sibling-matched euploid controls (23). Read depth analysis from low-pass whole genome

sequencing of each line demonstrated that the MEFs harbored complete trisomies without other chromosomal alterations and that the extra chromosomes were present clonally within the cell population (Fig. 1A). Various oncogenes are encoded on these chromosomes, including BCL2 (mChr1), FGFR4 (mChr13), Jak2 (mChr19), and many others (Table S1). Gain of these oncogenes, or some other consequence of aneuploidy, could drive malignant growth or otherwise generate cancer-like phenotypes in primary cells. We therefore set out to discover whether single-chromosome aneuploidy in MEFs was sufficient to induce neoplastic or pre-neoplastic behavior in untransformed cells.

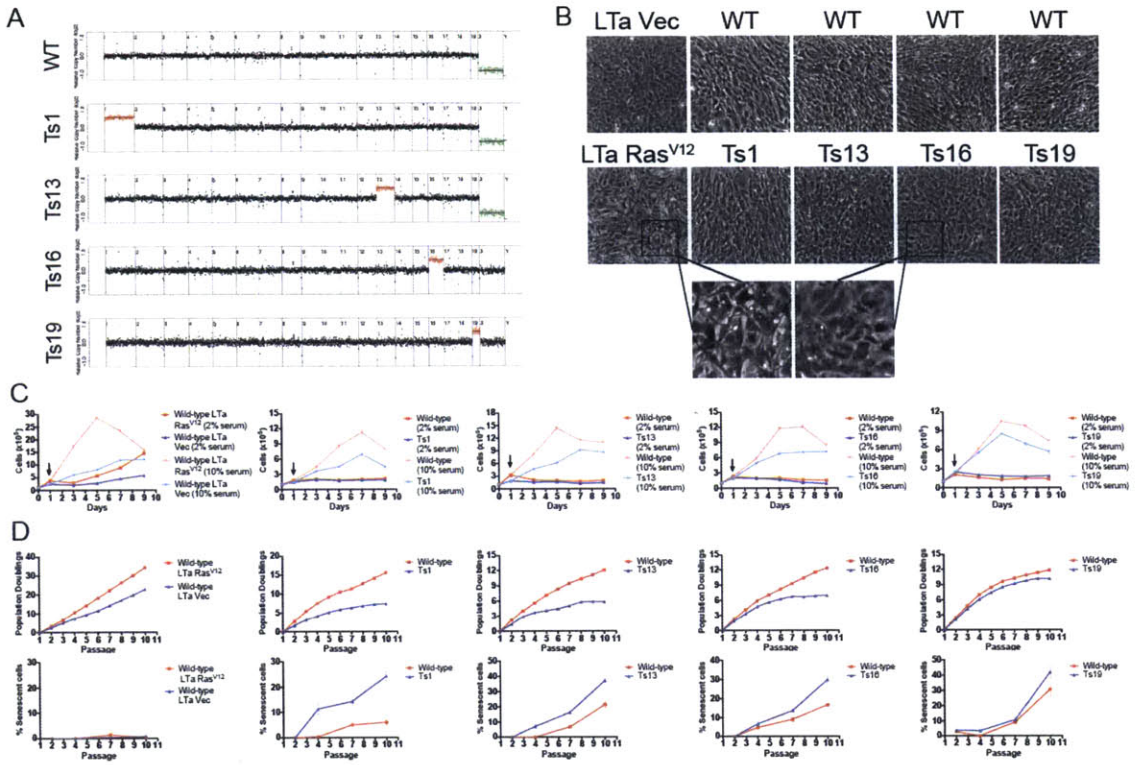


Figure 1. Single-chromosome aneuploidy is insufficient to induce neoplastic phenotypes. (A) MEF lines that were euploid or trisomic for chromosomes 1, 13, 16, or 19 were subjected to low-pass whole genome sequencing. Normalized read depths across 500kb bins are displayed. Note that only one euploid cell line is shown, though each trisomic MEF line had a separate euploid line that was derived from a wild-type littermate. (B) Photomicrographs of monolayers of the indicated cell lines. LTa+Ras^{V12} MEFs, but not trisomic MEFs, lose contact inhibition when grown to confluence. (C) Growth curves of the indicated cell lines are displayed. Cells were first plated in normal (10% serum) media, then 24 hours after plating the cells were re-fed or switched to reduced (2% serum) media (indicated by an arrow). LTa+Ras^{V12} MEFs, but not trisomic MEFs, continue to divide in low serum media. (D) The indicated strains were passaged, counted, and plated in triplicate every third day for 10 passages (top row). On passages 2, 4, 7, and 10, MEFs were also stained for beta-galactosidase expression to detect senescent cells (bottom row). LTa-transduced MEFs exhibit negligible levels of senescence, but trisomic cell lines senesce at an early passage.

As a positive control for cancer-like growth, we generated MEF lines that had been stably-transduced with the Large T antigen (LTa), which inhibits the Rb and p53 tumor suppressor pathways (35), and with either an empty vector or an activated allele of H-Ras (Ras^{V12}). As expected, the LTa+Ras^{V12} MEFs exhibited several neoplastic phenotypes: the cells were not contact-inhibited, and instead piled on top of each other when grown to confluence (Fig. 1B), they formed colonies from single cells when plated at low density (Fig. S1), they grew independently of pro-proliferative signals, as evidenced by their increase in cell number when plated in low-serum medium (Fig. 1C), and they grew robustly without senescing over 10 passages in culture (Fig. 1D). The LTa+vector MEFs displayed an intermediate, pre-neoplastic phenotype: the cells maintained contact inhibition and grew very poorly following serum withdrawal, but were mildly clonogenic and doubled without noticeable senescence over 10 passages in culture. In contrast, both euploid and trisomic MEFs failed to display any cancer-like phenotypes: they exhibited appropriate contact inhibition, failed to proliferate in low-serum media, were non-clonogenic, and senesced after 7 to 10 passages in culture. Interestingly, trisomies 1, 13, and 16 senesced at earlier passages and to a significantly higher degree than their matched euploid strains, as judged by β-galactosidase staining (Fig. 1D). We conclude that single-chromosome aneuploidy is insufficient to generate neoplastic phenotypes, and many aneuploidies in fact induce a premature growth arrest.

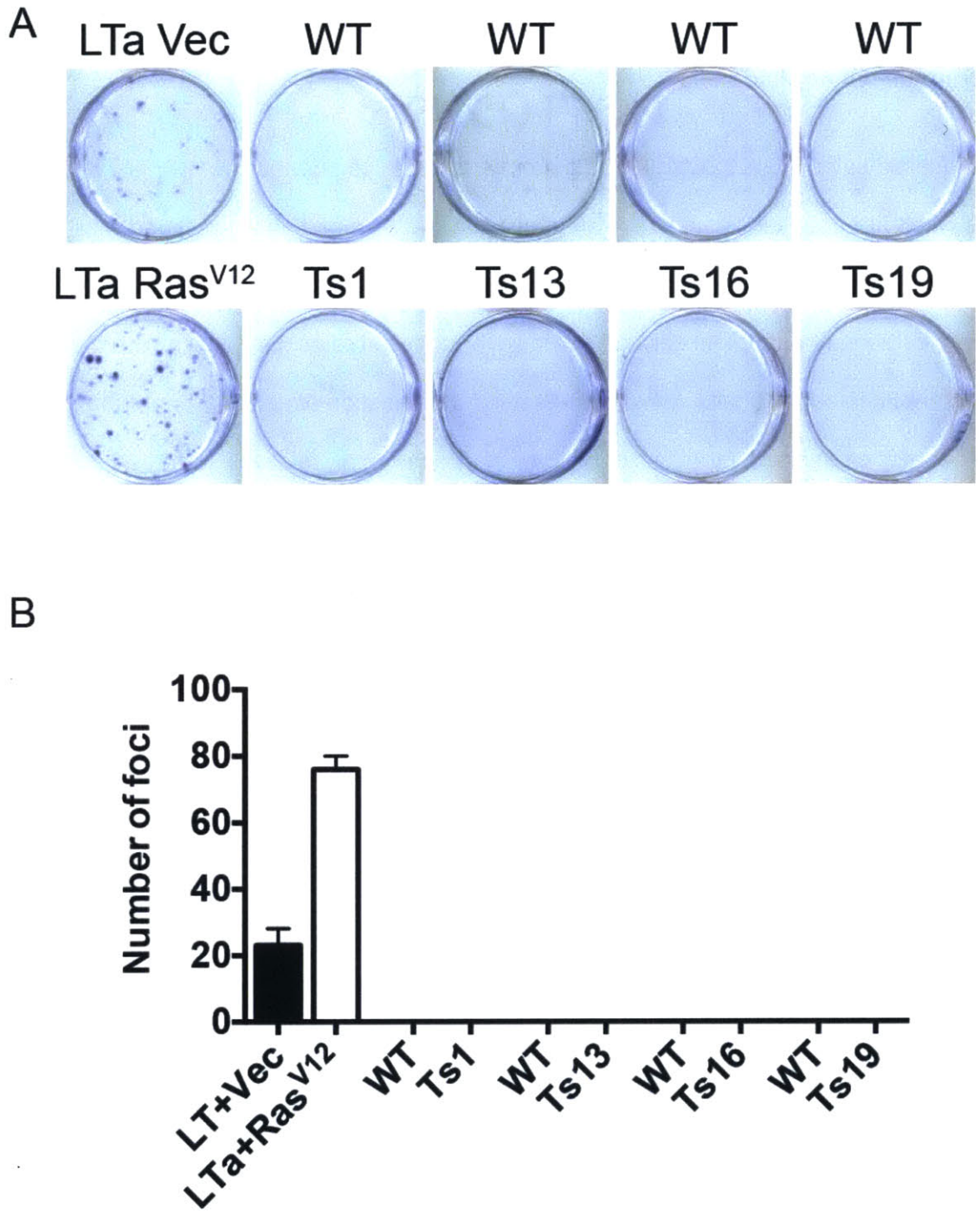


Figure S1. Primary aneuploid cells are non-clonogenic. (A) 1000 cells of the indicated strains were plated and then allowed to grow for 10 days before being stained with crystal violet. LTa-transduced MEFs are capable of forming colonies from single cells, but primary euploid and trisomic MEFs are non-clonogenic. (B) Quantification of (A).

Oncogene-transduced trisomic cells exhibit reduced proliferation and fitness relative to oncogene-transduced euploid cells

As aneuploidy alone was insufficient to generate cancer-like phenotypes, we next set out to determine whether aneuploidy could have synergistic, growth-promoting interactions with oncogenic mutations. In particular, loss of the p53 tumor suppressor has been linked with heightened proliferation of aneuploid cells (27). However, these studies did not examine the effects of loss of p53 in euploid cells, leaving unresolved the question of whether or not the proliferative benefits conferred by the loss of p53 are specific to (e.g., relatively greater) in aneuploid cells. To address this question, and to screen for synergistic interactions between aneuploidy and common oncogenic mutations, we stably transduced trisomic and euploid MEFs with plasmids encoding various oncogenes or with matched empty vectors. Trisomies 1, 13, 16, and 19, as well as matched euploid cell lines, were transduced with a dominant negative allele of p53 [p53dd (36)], the E1a oncogene [which inhibits the Rb tumor suppressor, among several other cellular pathways (37)], the Large T oncogene, or with the MYC oncogene. Following retroviral transduction and selection, the behavior of euploid and trisomic cell lines were tested in several assays that serve as *in vitro* and *in vivo* proxies for tumorigenic capacity. Additionally, the karyotypes of 10 cell lines were determined by low-pass whole genome sequencing at early passage following transduction (Fig. S2). All 10 lines maintained initial karyotypes, confirming that we had constructed oncogene-expressing cell lines with single, defined chromosomal copy number alterations.

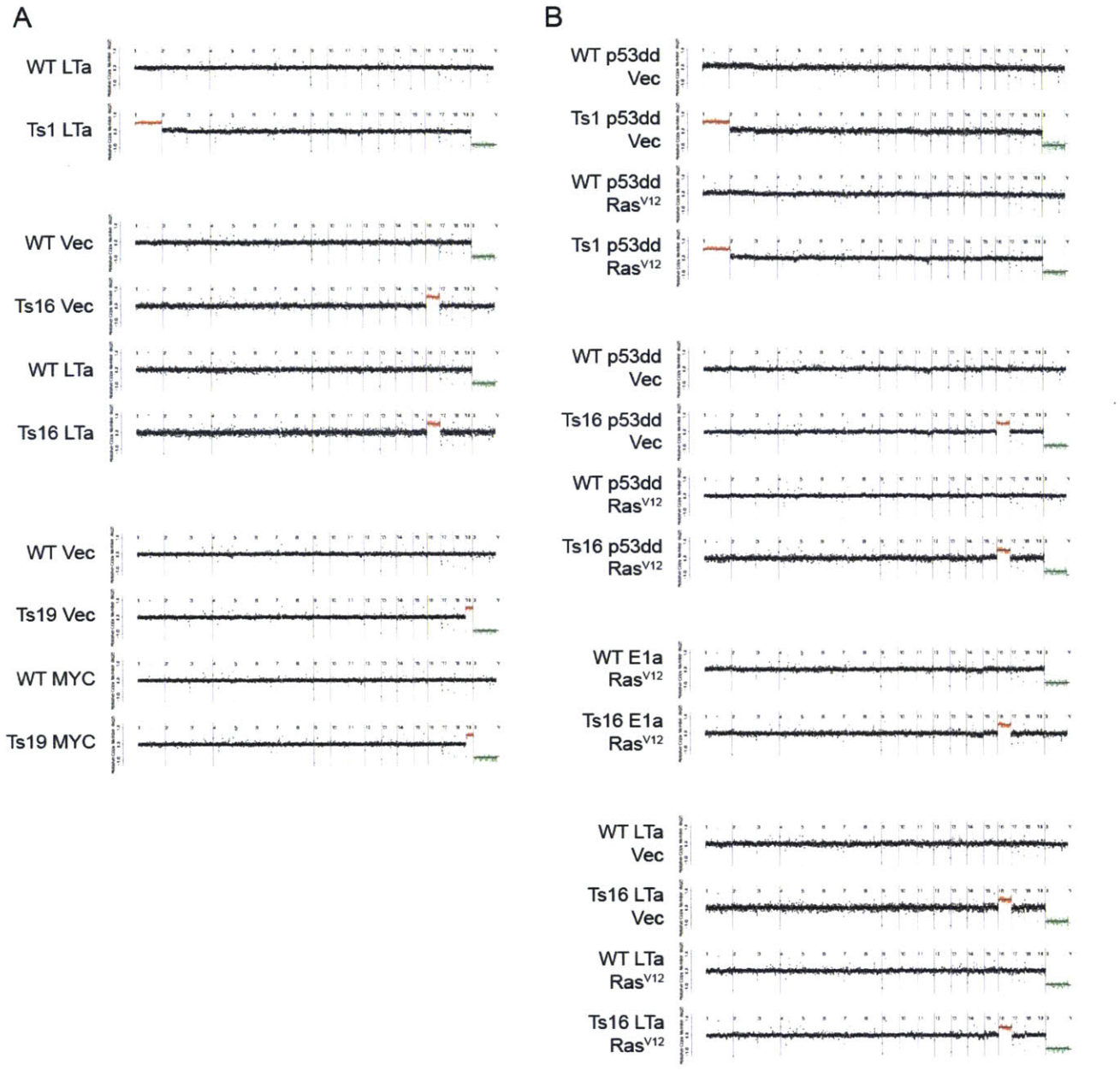


Figure S2. Trisomic karyotypes are maintained during retroviral transduction and selection. MEF lines that were transduced with (A) one oncogene or (B) two oncogenes were subjected to low-pass whole genome sequencing. 24 out of 24 cell lines were found to maintain their initial karyotype.

Oncogene-transduced euploid and trisomic MEFs were counted and passaged 10 times over the course of 30 days following selection. Empty vector-transduced trisomic MEFs generally underwent fewer population doublings than matched euploid lines (Figure 2, compare dark red and dark blue lines). Transduction with oncogenes significantly enhanced the growth of trisomic cells, and the resultant lines doubled more frequently over the course of the experiment than the vector-matched controls. However, these oncogenes also enhanced the growth of euploid cell lines. In the majority of cases, the oncogene-transduced euploid cell line underwent more population doublings than the corresponding oncogene-transduced aneuploid line, and in only a few instances did we observe equivalent proliferation between euploid and trisomic MEFs (Figure 2, compare light red and light blue lines). For instance, over 10 passages in culture, E1a-transduced Ts16 MEFs doubled about 11 times, significantly fewer than the 17 doublings underwent by an E1a-transduced euploid line. Transduction with Large T demonstrated the most significant growth rescue in trisomic strains, though LTa-expressing cells trisomic for chromosomes 1 and 16 still proliferated significantly more slowly than LTa-expressing euploid lines.

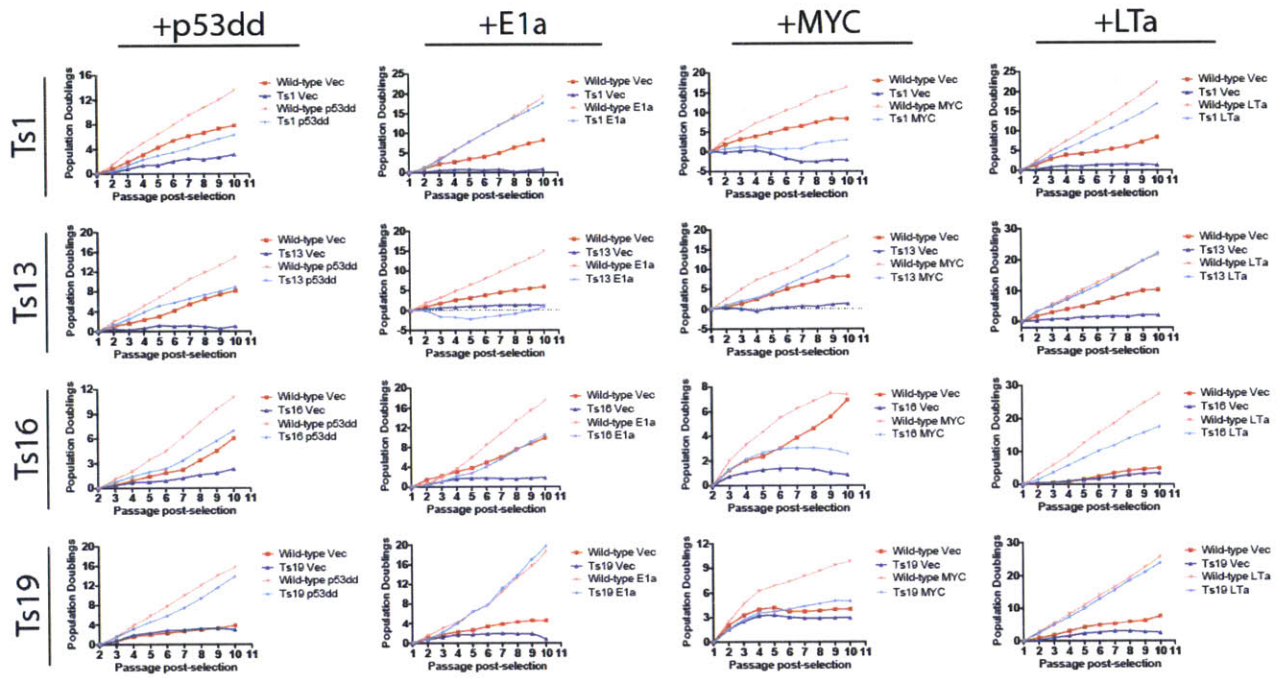


Figure 2. Aneuploidy impedes the growth of oncogene-transduced cell lines. Euploid and trisomic cell lines were stably transduced with plasmids harboring the indicated oncogene or a matched empty vector. Following selection, the cell lines were passaged every third day for up to 10 passages, and the cumulative population doublings over the course of each experiment are displayed.

We next tested several additional oncogenes in a subset of trisomies: euploid cell lines transduced with a stabilized allele of MYC (MYC^{T58A}) outgrew matched cell lines trisomic for chromosomes 13 or 16, while expression of an activated allele of BRAF ($BRAF^{V600E}$) or an allele of CDK4 resistant to p16 inhibition ($CDK4^{R24C}$) resulted in senescence in both euploid and trisomic cell lines (Fig. S3).

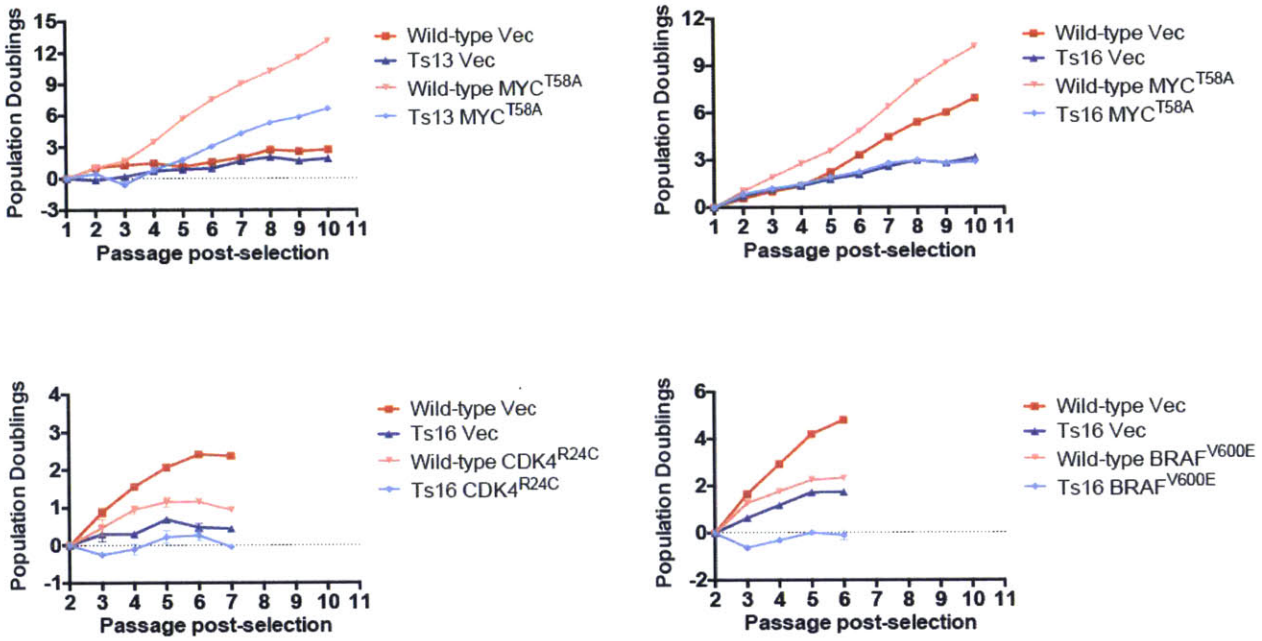


Figure S3. Effects of MYC^{T58A}, BRAF^{V600E}, and CDK^{R24C} on trisomic cells. Euploid and trisomic cell lines were stably transduced with plasmids harboring the indicated oncogene or a matched empty vector. Following selection, the cell lines were passaged every third day for up to 10 passages, and the cumulative population doublings over the course of each experiment are displayed. Experiments with BRAF^{V600E}, and CDK4^{R24C} were terminated prematurely as both euploid and trisomic cells senesced following transduction.

As the experiments described above were conducted using initial populations of primary cells, we selected nine experiments to repeat on independently-derived cell lines (Fig. S4A). Replicate experiments displayed some line-to-line variability, but recapitulated the major features of our initial results: oncogene-transduced trisomic cells grew poorly relative to euploid cells, with only LTa-expressing lines and lines trisomic for chromosome 19 growing at close to wild-type levels (Fig. S4A). Our prolonged culture period also allowed us to follow the dynamics of aneuploid cell populations over time: rapidly-dividing subpopulations could potentially arise during 30 days in culture that would enhance the apparent proliferative capacity of the aneuploid MEFs. We found that one Ts19+LTa cell line proliferated more rapidly over time (Fig. S4A), while an independent Ts19+LTa line grew at approximately the same rate over the course of the experiment (Fig. 2). The reason for this discrepancy is explored below. In general, we found that few trisomic cell lines doubled more rapidly over successive passages, while many trisomic cell lines (particularly those transduced with MYC or p53dd) senesced in culture instead. Across multiple experiments, we did not observe any instances in which transduction with an oncogene generated a trisomic cell line that reproducibly outgrew its matched euploid counterpart. We conclude that euploid lines harboring common oncogenic mutations generally proliferate more rapidly than trisomic lines harboring the same lesions.

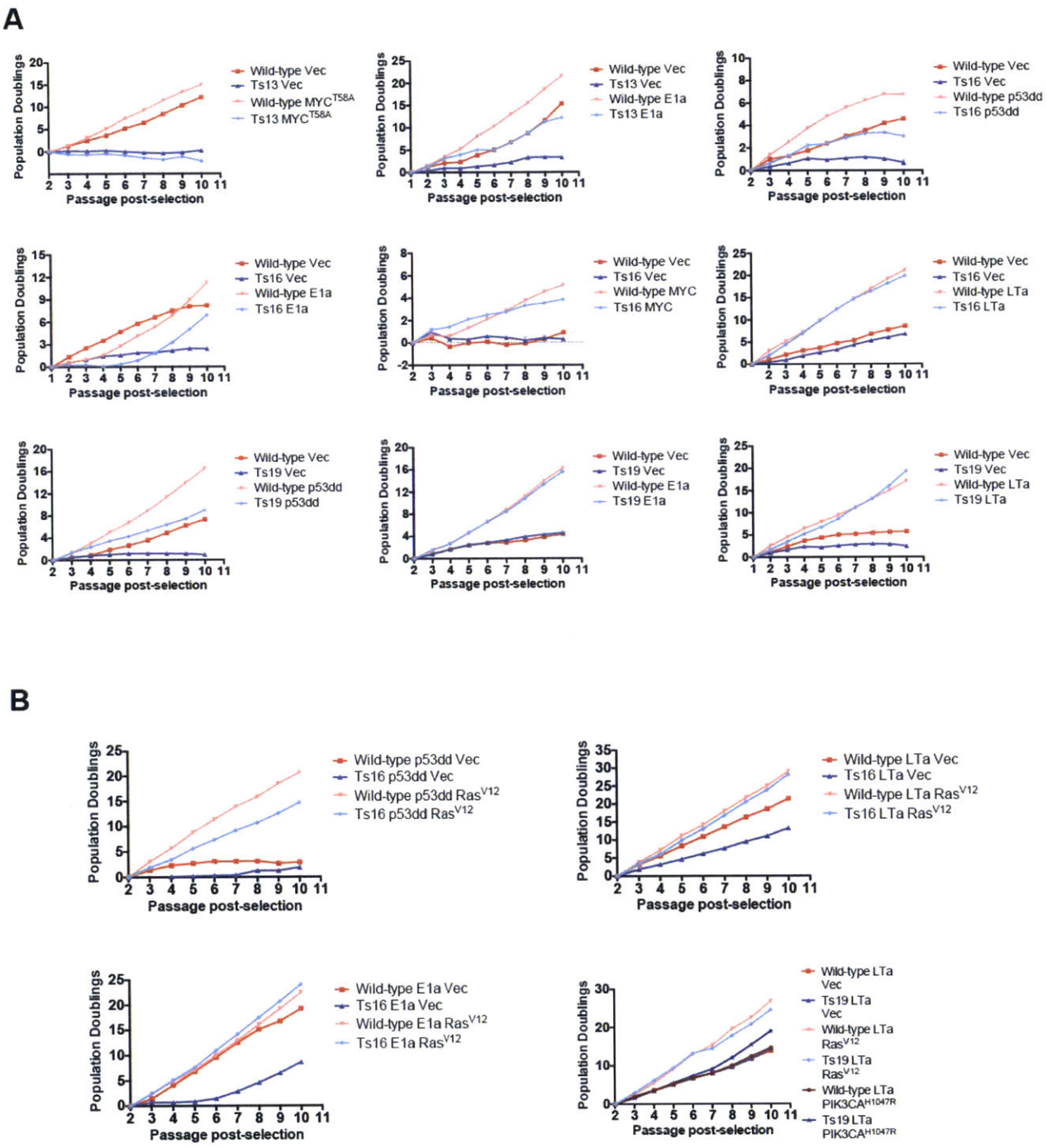


Figure S4. Replicate oncogene-transduction experiments on independently-derived cell lines. Nine single-oncogene transduction experiments (A) and four two-oncogene transduction experiments (B) were repeated in different, independently-derived cell lines. These lines were stably transduced with plasmids harboring the indicated oncogene(s) or a matched empty

vector, and then passaged every third day for up to 10 passages. Note that some variability exists between replicates (e.g., compare Ts16+MYC in Fig. S4 and Fig. 2), but no trisomy+oncogene combination was found to reproducibly outgrow a matched euploid strain.

Though significant proliferative differences between aneuploid and euploid lines persisted following oncogene transduction, it remained possible that the oncogenes provided a relatively greater growth advantage to aneuploid cells than to euploid cells. To test this, we quantified the benefit conferred by each oncogene by comparing the number of cells recovered at every passage from oncogene-transduced and vector-transduced cell lines (Fig. S5A). For most oncogene-trisomy combinations, the fold change growth enhancement was equivalent between euploid and aneuploid lines. For instance, transduction with dominant-negative p53 resulted in an approximately 1.6 to 1.8-fold increase in doublings per passage, relative to the vector-transduced controls, in all euploid and trisomic cell lines. Thus, the abrogation of p53 signaling does not specifically enhance the growth of these trisomic cells. In contrast, MYC had a proportionally greater effect on euploid cells than on trisomy 19 cells, while E1a had a proportionally greater effect on Ts1 MEFs than on euploid MEFs, and Large T had a proportionally greater effect on Ts13 13 MEFs than on its euploid control (Fig. S5). These results suggest that synergistic interactions between oncogenes and trisomies are rare, and many oncogenes confer similar, karyotype-independent growth advantages.

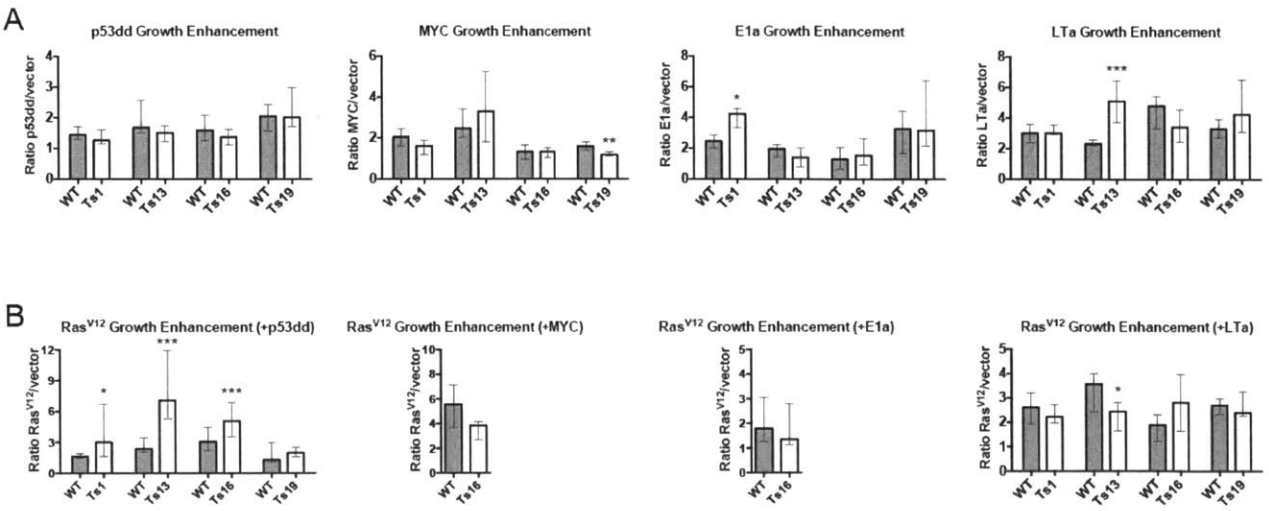


Figure S5. Relative growth enhancement conferred by oncogenes in euploid and aneuploid cell lines. (A and B) The number of cells recovered from oncogene-transduced MEFs was divided by the number of cells recovered from vector-transduced MEFs at every passage. Bar graphs display the median ratios and the interquartile ranges. * p<.05; *** p<.0005 (Wilcoxon rank-sum test).

As an additional test of the proliferative capabilities of the oncogene-transduced lines, we assessed the focus formation ability of cells that had been transduced with E1a or LTa (MYC and p53dd-transduced lines remained non-clonogenic). We found that transduced euploid lines exhibited uniformly superior colony-forming ability relative to the trisomic lines, even in instances when the euploid and aneuploid lines demonstrated approximately equal doubling times in culture (Fig. S6). The differences between the colony formation and population doubling assays likely reflect the fact that forming a colony is a relatively greater challenge to a cell than doubling in a monolayer, and this challenge exacerbates the fitness differential between euploid and trisomic cells.

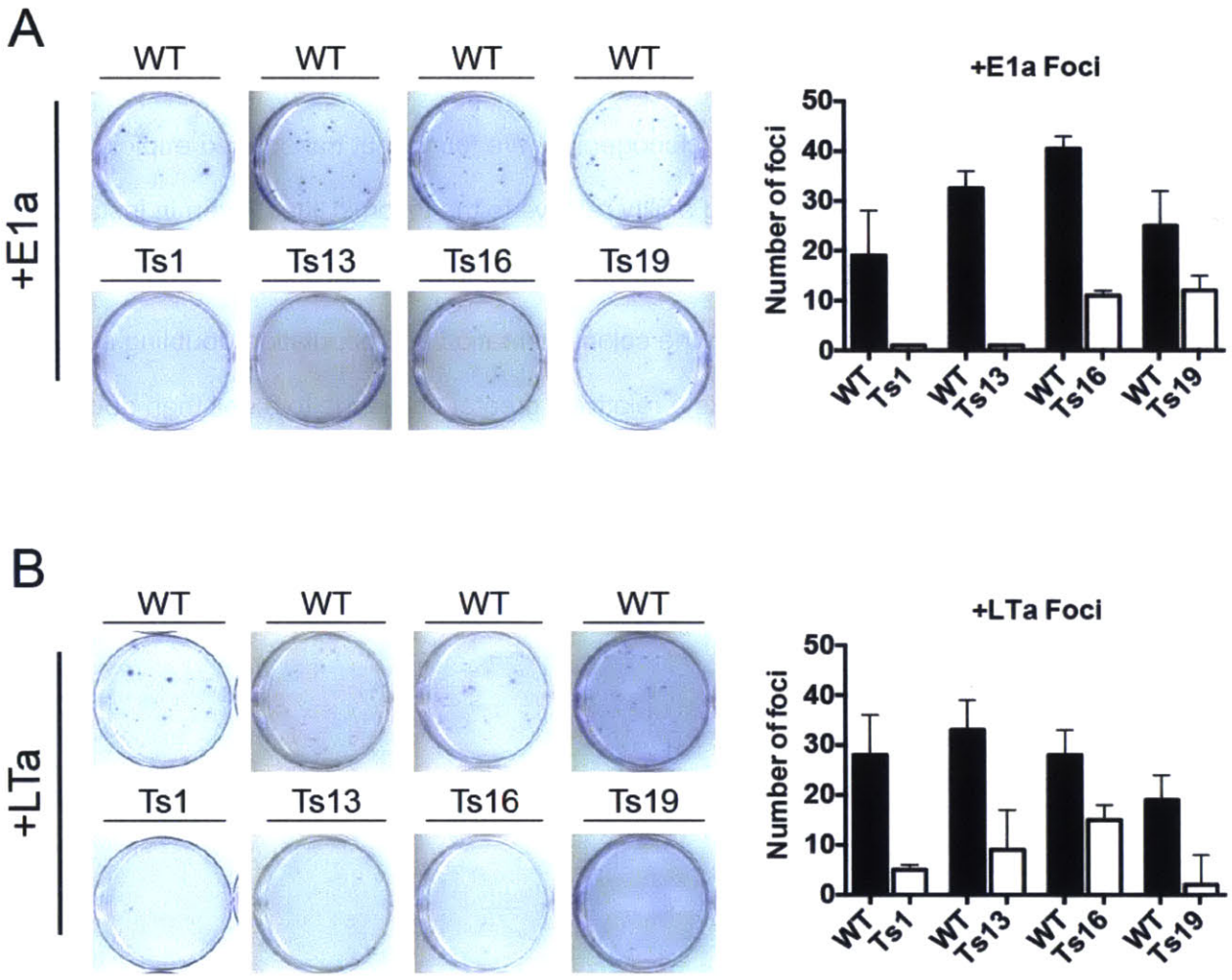


Figure S6. Oncogene-transduced aneuploid cell lines exhibit reduced clonogenicity. (A and B) 1000 cells of the indicated strains were plated and then allowed to grow for 10 days before being stained with crystal violet. Representative plates are shown on the left, while average colony counts are displayed on the right. In each experiment, the trisomic cell lines were found to exhibit a significantly reduced focus formation ability relative to the matched euploid cell line ($p<.01$, Student's t test).

Ras^{V12}-transformed aneuploid cells exhibit reduced tumorigenicity

Singly-transduced MEFs are not fully-transformed; complete transformation of rodent cells requires two cooperating oncogenes (38). We therefore took p53dd- or LTa-transduced cell lines from each trisomic and matched euploid strain and then stably transduced them with plasmids harboring H-Ras^{V12} or a control empty vector. Whole genome sequencing revealed that 14 out of 14 tested cell lines maintained their initial karyotype following two rounds of retroviral transduction and selection, demonstrating that our protocol does not result in loss of the trisomic chromosome (Fig. S2B).

Ras^{V12}-transduced cell lines doubled significantly faster than vector-transduced lines over 10 passages in culture, and Ras^{V12} narrowed or in some cases abolished the proliferative difference between euploid and trisomic MEFs (Fig. 3A). However, no Ras^{V12}-expressing trisomic line grew faster than its euploid counterpart. In a subset of trisomies, we also examined the effects of Ras^{V12} expression in E1a- or MYC-transduced cell lines (Fig. S7A), and we tested the expression of $BRAF^{V600E}$, $PIK3CA^{H1047R}$, or MYC in lieu of Ras^{V12} as a driver oncogene (Fig. S7B). The latter oncogenes typically had little effect in this assay compared to experiments using Ras^{V12} that were performed in parallel. Replicates of four experiments on independently-derived cell lines revealed some cell line-specific variability but no major differences in experimental outcome. One line of Ts19+LTa+$PIK3CA^{H1047R}$ grew slightly better than its euploid control, though an independent line of Ts19+LTa+$PIK3CA^{H1047R}$ did not exhibit this phenotype (Fig. S7, and see Fig. 7 below). No oncogene cocktail tested resulted in reproducibly superior growth in a trisomic strain compared to its euploid counterpart. Interestingly, in a p53dd background, Ras^{V12} had a relatively greater effect on trisomies 1, 13, and 16 than it did on the euploid strains (Fig. S5B). This is likely due at least in part to increased senescence of the p53dd+vector doubly-transduced trisomic MEFs, and a similar selective growth enhancement was not observed in LTa-, MYC-, or E1a-transduced trisomic strains.

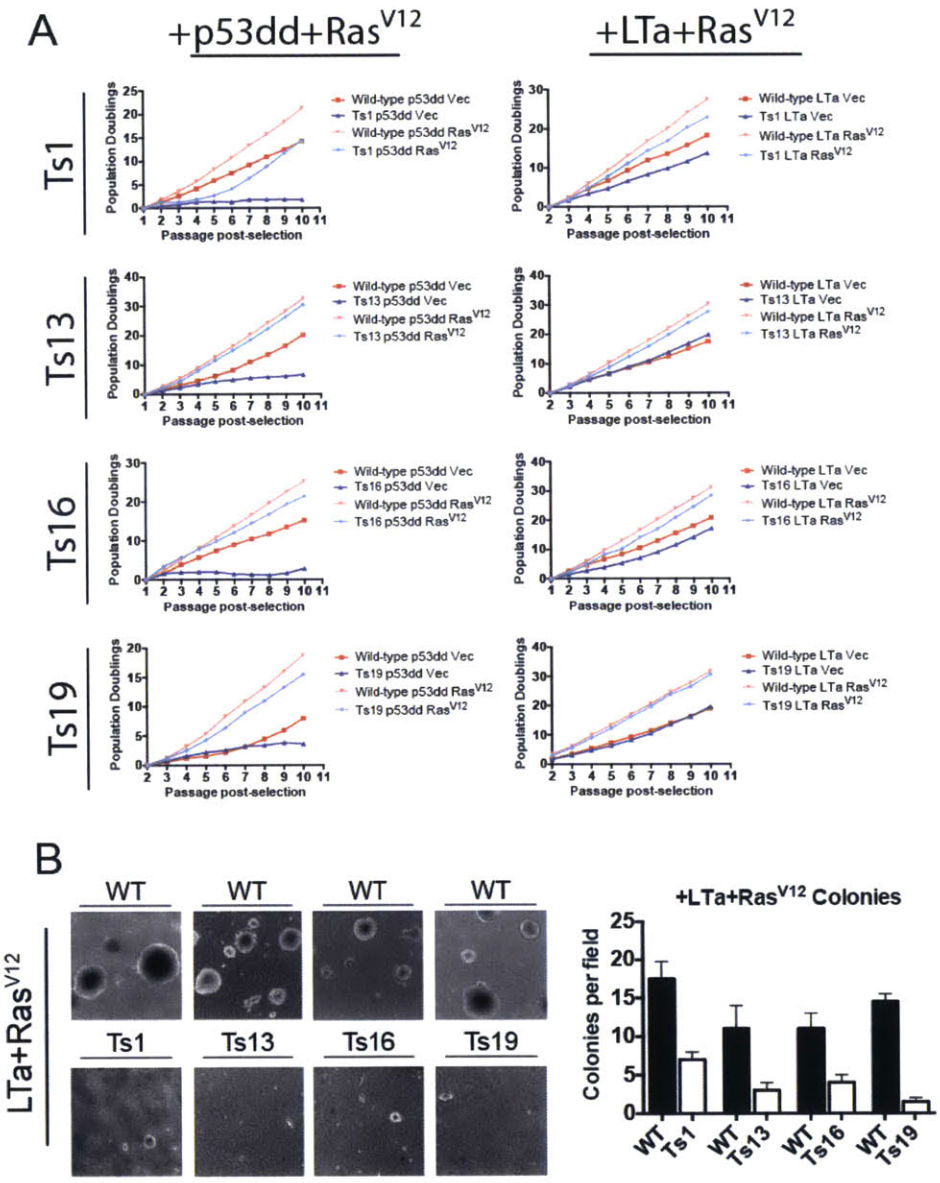
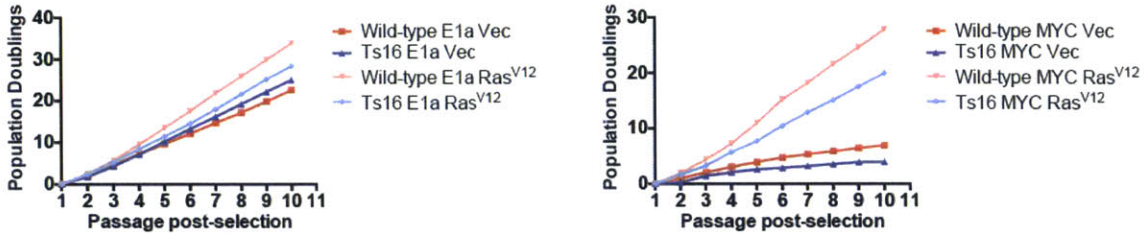


Figure 3. Ras^{V12} rescues proliferation but not colony formation in trisomic MEFs. (A) Euploid and trisomic cell lines were first stably transduced with p53dd or with LTa, and then were transduced a second time with plasmids harboring Ras^{V12} or a matched empty vector. The cell lines were passaged, counted, and plated in triplicate up to 10 passages following the second round of selection. (B) 20,000 cells of the indicated strains were plated in soft agar and then allowed to grow for 20 days. For each comparison, the euploid MEFs formed more colonies than the trisomic MEFs ($p<.01$, Student's t test).

A



B

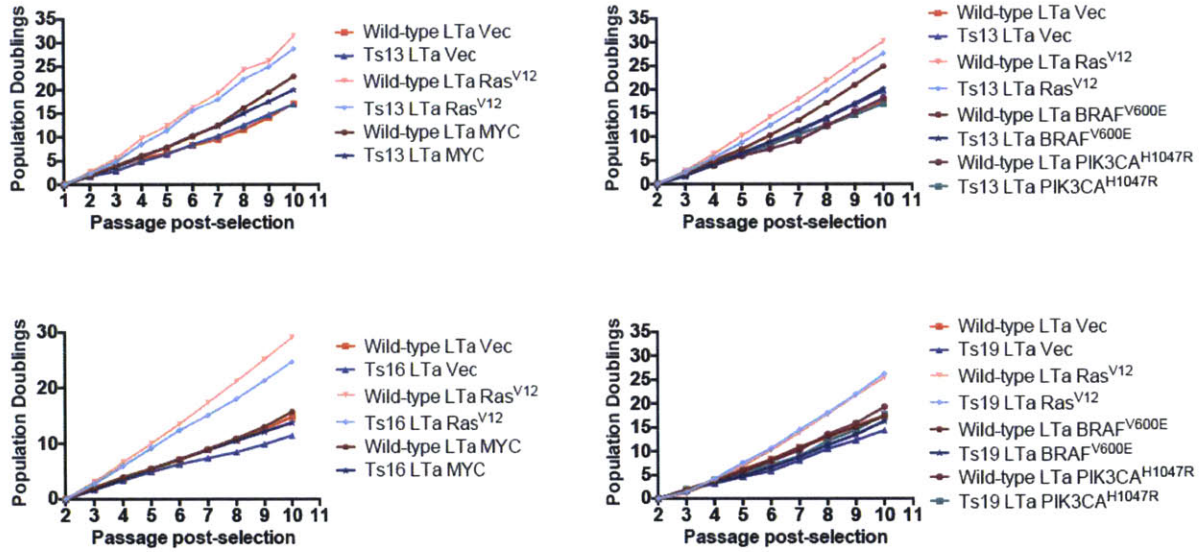


Figure S7. Additional oncogene cocktails in aneuploid and euploid MEFs (A and B) Euploid and trisomic cell lines were stably transduced with plasmids harboring the indicated oncogene or a matched empty vector. Following selection, the cell lines were passaged every third day for up to 10 passages, and the cumulative population doublings over the course of each experiment are displayed. Note that in B, experiments with Ras^{V12} were also performed in parallel, and are shown here for comparison.

As the Ras^{V12}-transduced trisomic lines displayed equivalent or nearly-equivalent proliferative capacity as the euploid lines, we hypothesized that Ras^{V12} could suppress the fitness differential between aneuploid and euploid MEFs. However, as we observed with the singly-transduced cell lines, Ras^{V12}-transduced euploid MEFs formed more colonies from single cells than trisomic strains did, even when the strains were observed to proliferate at the same rate in culture (Fig. S8A and S8B). Fully transformed cell lines are also competent to grow in soft agar, a phenotype that strongly correlates with *in vivo* tumorigenicity (39). We tested the ability of LTa+Ras^{V12}-transduced cell lines to form colonies in soft agar, and in each case the euploid lines exhibited higher colony-forming ability than the equivalently-transduced trisomic strains (Fig. 3B).

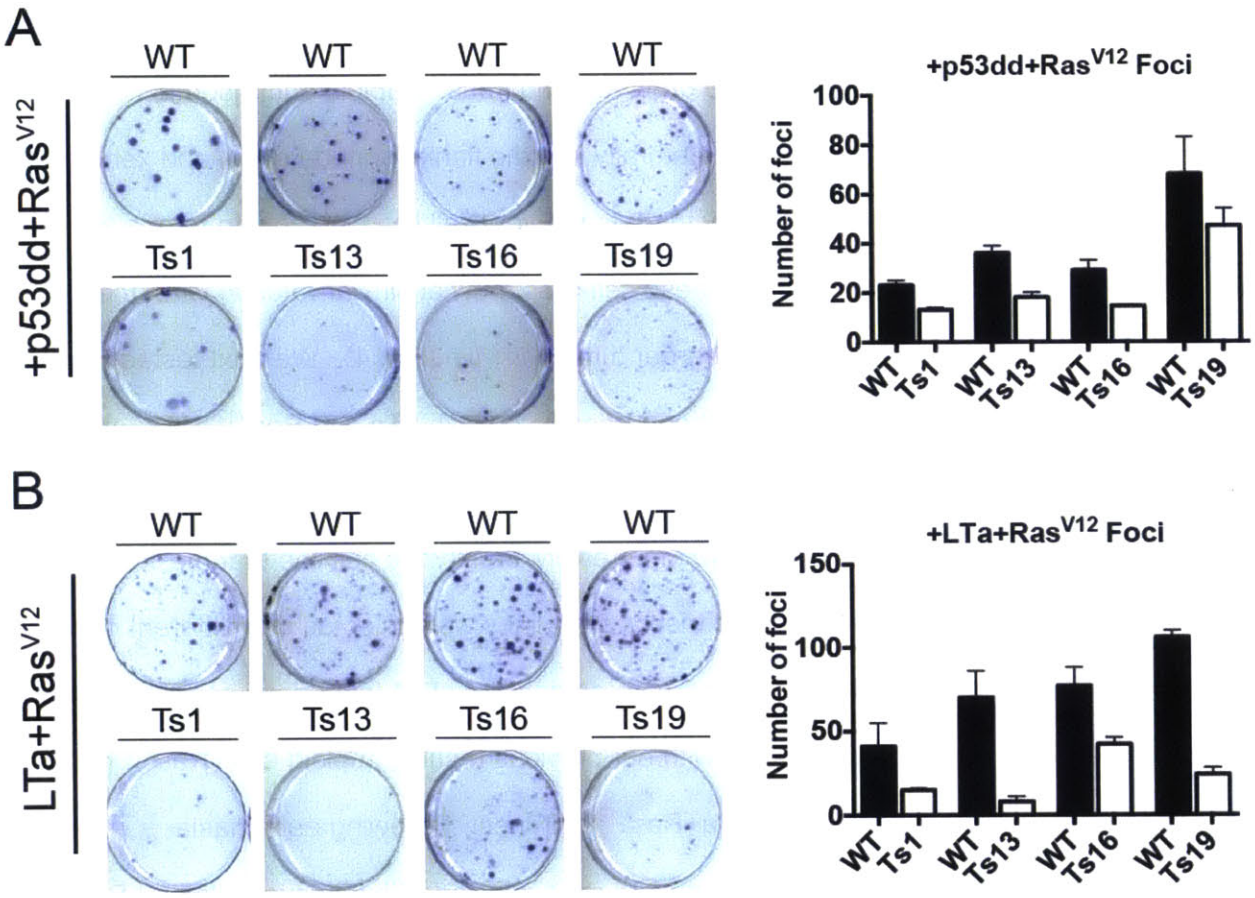


Figure S8. Transformed aneuploid cells display poor clonogenicity and a reduced ability to form colonies in soft agar. (A and B) 1000 cells of the indicated strains were plated and then allowed to grow for 10 days before being stained with crystal violet. For each comparison, the euploid MEFs formed more foci than the trisomic MEFs ($p<.01$, Student's t test).

As a final test of the tumorigenicity of trisomic MEFs, we examined the ability of Ras^{V12}-transduced euploid and trisomic cell lines to grow as xenografts. Euploid and trisomic p53dd+Ras^{V12} cell lines were injected contralaterally into the flanks of nude mice, and xenograft volume measurements were obtained every third day. Every cell line was able to form a tumor at the site of the injection. However, while these cell lines grew at similar rates *in vitro*, the euploid lines invariably formed significantly larger tumors *in vivo* (Fig. 4). We next tested the tumorigenicity of trisomic cell lines that had been transduced with LTa and Ras^{V12}. In preliminary experiments, a fraction of mice injected with these cell lines developed cachexia that resulted from metastatic disease. In order to more easily identify the cell line that the metastases were derived from, euploid and trisomic cell lines were injected into different mice during single experiments. Additionally, mice were euthanized at 11 to 15 days post-injection, as cachexia began to develop. Consistent with our previous results, euploid LTa+Ras^{V12} cell lines formed larger tumors than trisomic LTa+Ras^{V12} cell lines. Following euthanasia, necropsies were performed on 29 mice: 3 out of 14 mice injected with trisomic cells and 5 out of 15 mice injected with euploid cells exhibited evidence of gross metastases (p=.68, Fisher's exact test). Metastatic lesions were commonly observed on the stomach, spleen, liver, and pancreas of these mice, with no apparent difference in organ colonization between mice that had been injected with euploid or trisomic MEFs. Histological analysis identified the primary tumors and metastatic lesions as poorly-differentiated fibrosarcomas, consistent with their embryonic fibroblast origins (Fig. 4C). No gross differences in histology were apparent between euploid and trisomic tumors. In total, these results demonstrate that tumors derived from euploid and trisomic cells form histologically similar structures, but euploid cells typically outgrow genetically-identical trisomic cells *in vivo*.

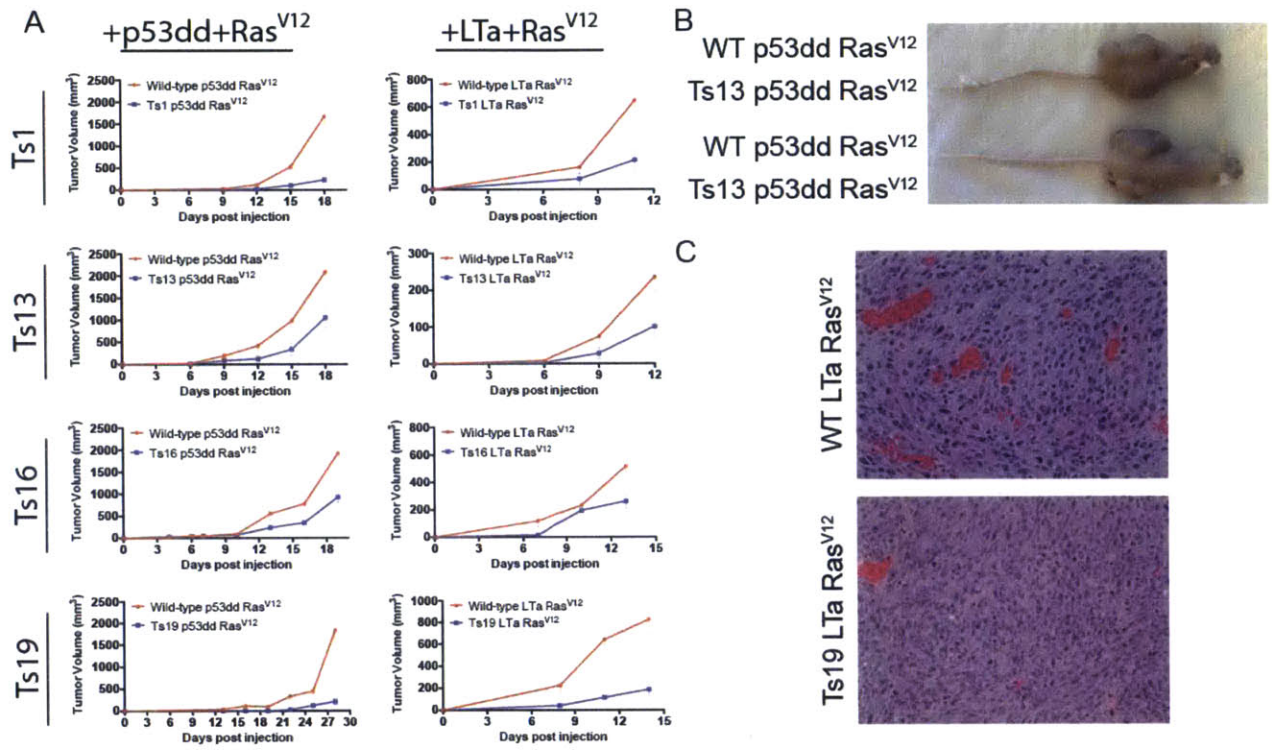


Figure 4. Trisomic cell lines grow poorly as xenografts. (A) 10^6 euploid or aneuploid cells transduced with either p53dd and Ras^{V12} or with LTa and Ras^{V12} were injected subcutaneously into the flanks of nude mice, and then tumor volume was measured every three days. Note that mice injected with cells transduced with LTa+Ras^{V12} had to be euthanized prematurely due to cachexia (see Main Text). (B) Representative images of mice injected contralaterally with WT+p53dd+Ras^{V12} cells or Ts13+p53dd+Ras^{V12} cells. (C) Representative H&E sections of primary tumors from WT+LTa+Ras^{V12} cells or Ts19+LTa+Ras^{V12} cells.

Aneuploidy impedes the tumorigenicity of human colorectal cancer cells

Our experiments in MEFs demonstrated that aneuploidy in primary cells commonly impedes transformation. As aneuploidy is a nearly universal occurrence in cancer, we hypothesized that the acquisition of aneuploidy in previously-transformed cells, rather than in primary cells, could have distinct, pro-tumorigenic consequences. To test this, we utilized a series of chromosomally-stable human colorectal cancer cells into which extra chromosomes had been introduced via microcell-mediated chromosome transfer (24,40). We first characterized each cell line by whole-genome sequencing (Fig. 5A). The parental cell line, HCT116, had a euploid chromosomal complement, though read-depth analysis confirmed several previously-described segmental gains on chromosomes 8, 10, 16, and 17. These amplifications were also present in the derived cell lines, and are therefore unlikely to affect the results described below. We compared the behavior of the parental HCT116 line to a cell line that was trisomic for chromosome 5 (HCT116 5/3), a cell line that had regions of trisomy and tetrasomy on chromosome 5 (HCT115 5/4), a cell line that had regions of trisomy and tetrasomy on chromosome 3 (HCT116 3/3), and two cell lines that had regions of trisomy and tetrasomy on chromosome 8 (HCT116 8/3 c4, which had gained a complete extra copy of chromosome 8, and HCT116 8/3 c3, which had gained a partial copy of chromosome 8). Oncogenes encoded on these chromosomes include beta-catenin (hChr3), PIK3CA (hChr3), TERT (hChr5), MYC (hChr8), and several others (Table S2).

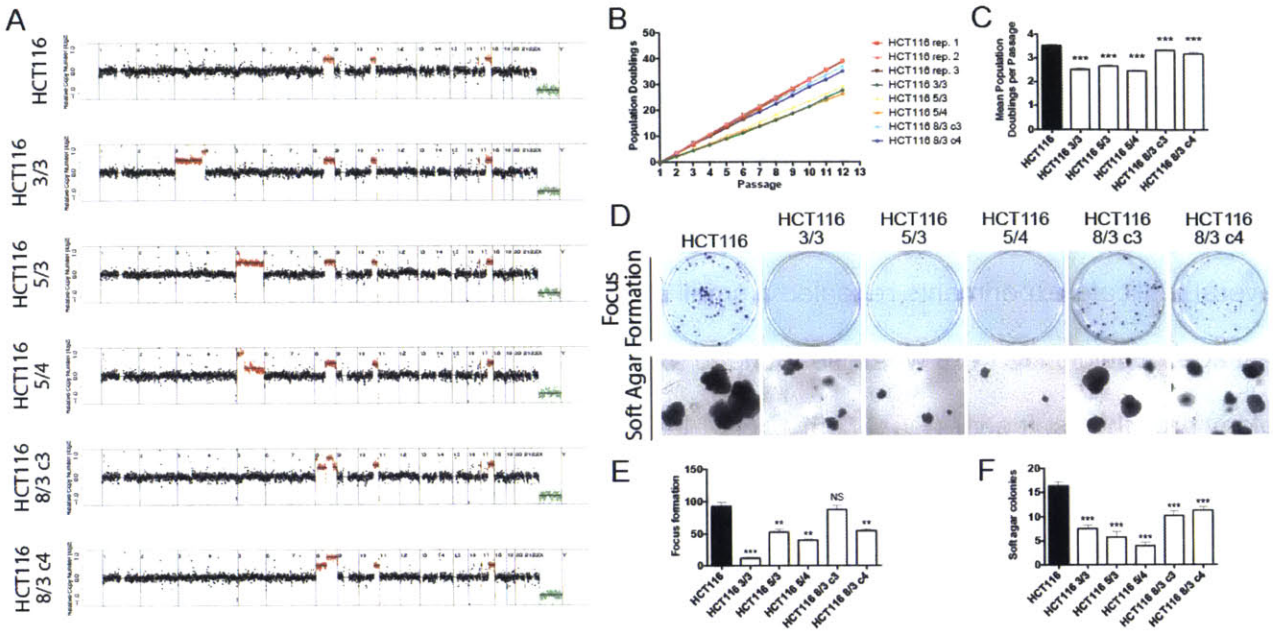


Figure 5. Aneuploidy impedes the growth of human colorectal cancer cell lines *in vitro*.

(A) Normalized read depths from whole genome sequencing of euploid and aneuploid human colorectal cancer cell lines. (B) Growth curves of euploid and aneuploid colorectal cancer cell lines counted and passaged every third day. (C) Quantification of the mean population doublings per passage of multiple replicates of the experiment shown in B. (D) 200 cells of the indicates strains were plated and allowed to grow for 14 days before being stained with crystal violet (top), or 2000 cells on the indicated strains were plated in soft agar and allowed to grow for 20 days before being imaged (bottom). (E) Quantification of the focus formation assay in D. (F) Quantification of the soft agar assay in D.

We tested the HCT116 cell lines in similar assays as described above in order to compare the relative fitness and tumorigenicity of euploid and aneuploid colon cancer cells. During serial passaging, the euploid parental line divided the most rapidly, while the HCT116 3/3, 5/3, and 5/4 lines underwent on average one fewer doubling per passage (Fig. 5B and 5C). HCT116 8/3 c3 and c4 divided at nearly the same rate as the wild-type line, although analysis of several replicate experiments revealed a small but significant decrease in doublings per passage. Euploid HCT116 was also found to exhibit the highest rates of focus formation and colony growth in soft agar (Fig. 5D-5F). HCT116 3/3, 5/3, and 5/4 displayed significant impairments in both assays. HCT116 8/3 c4 exhibited a small reduction in both assays, while HCT116 8/3 c3 grew moderately worse in soft agar but was able to form foci on plastic at wild-type levels. We conclude that the introduction of aneuploidy into cancer cells commonly antagonizes growth *in vitro*, though in some instances it can be a nearly-neutral event.

To determine how aneuploidy in these cell lines influenced tumorigenesis *in vivo*, we performed contralateral subcutaneous injections of either the euploid parental HCT116 line or the aneuploid derivative lines into flanks of nude mice. The parental HCT116 cell line formed large tumors in all animals into which it had been injected. These tumors grew at a rapid rate, and each animal had to be euthanized 30 to 35 days after injection due to tumor burden. HCT116 3/3 and 5/4 formed small nodules at the site of injection that remained stable or increased in size very slightly over the course of the experiment (Fig. 6). Mice injected with HCT116 5/3 developed tumors that grew more rapidly than either 3/3 or 5/4 but significantly less rapidly than the euploid tumors. Finally, consistent with our *in vitro* experiments, HCT116 8/3 c3 and c4 grew well as xenografts, and there was no significant difference in tumor volume between these lines and the wild-type line. We conclude that gain of chromosomes in a cancer cell line can antagonize tumor formation.

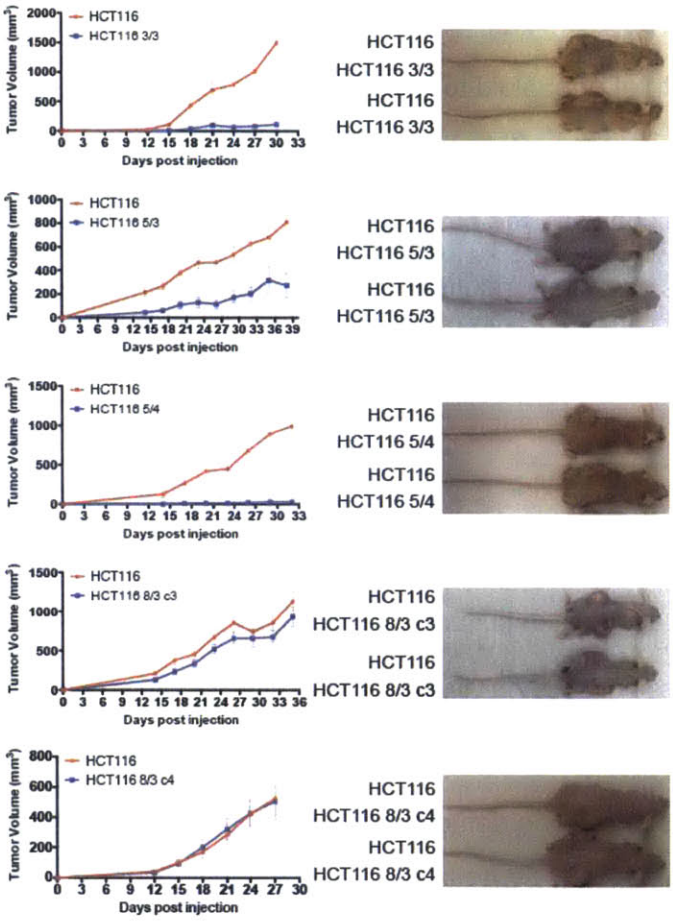


Figure 6. Aneuploidy impedes the growth of human colorectal cancer cell lines *in vivo*.

$4x10^6$ euploid or aneuploid cells were injected subcutaneously into the flanks of 5-10 nude mice, and then subsequent tumor growth was measured every third day.

Improved growth in aneuploid cells is associated with further karyotypic alterations

The robust growth of aneuploid tumors suggests that cancer cells are able to adapt to the adverse effects of aneuploidy. Our results argue that one commonly-hypothesized aneuploidy-tolerating mechanism, the activation of oncogenes and the inactivation of p53 and other tumor suppressors, is largely insufficient to equalize growth between euploid and aneuploid cells. We therefore sought to uncover other changes that could explain how cells are sometimes able to adapt to the aneuploid state.

In the oncogene-transduction experiments described above, we noted that one Ts19+LTa cell line initially proliferated at approximately the same rate as its euploid control, and then, following several passages in culture, its proliferative rate increased (Fig. 7A). An independent LTa-expressing Ts19 cell line did not exhibit this phenotype, and instead consistently doubled at the same rate over 10 passages in culture. Whole genome sequencing of the rapidly-growing Ts19 cell line demonstrated that it had unexpectedly gained an extra copy of chromosome 2 in addition to the trisomy of chromosome 19 (Fig. 7A). In contrast, the slower-growing Ts19 cell line was found to maintain its initial karyotype. Similarly, in a set of independent cell lines, we noted that one line of Ts19+LTa+PIK3CA^{H1047R} proliferated more rapidly than its euploid control, while Ts19+LTa+Vec and Ts19+LTa+Ras^{V12} proliferated at the same rate as similarly-transduced euploid MEFs. Karyotype analysis demonstrated that the Ts19+LTa+PIK3CA^{H1047R} cell line had also gained an additional copy of chromosome 2, while the other Ts19 cell lines had not (Fig. 7B). These results suggest that the gain of chromosome 2 enhances the growth rate of cells trisomic for chromosome 19.

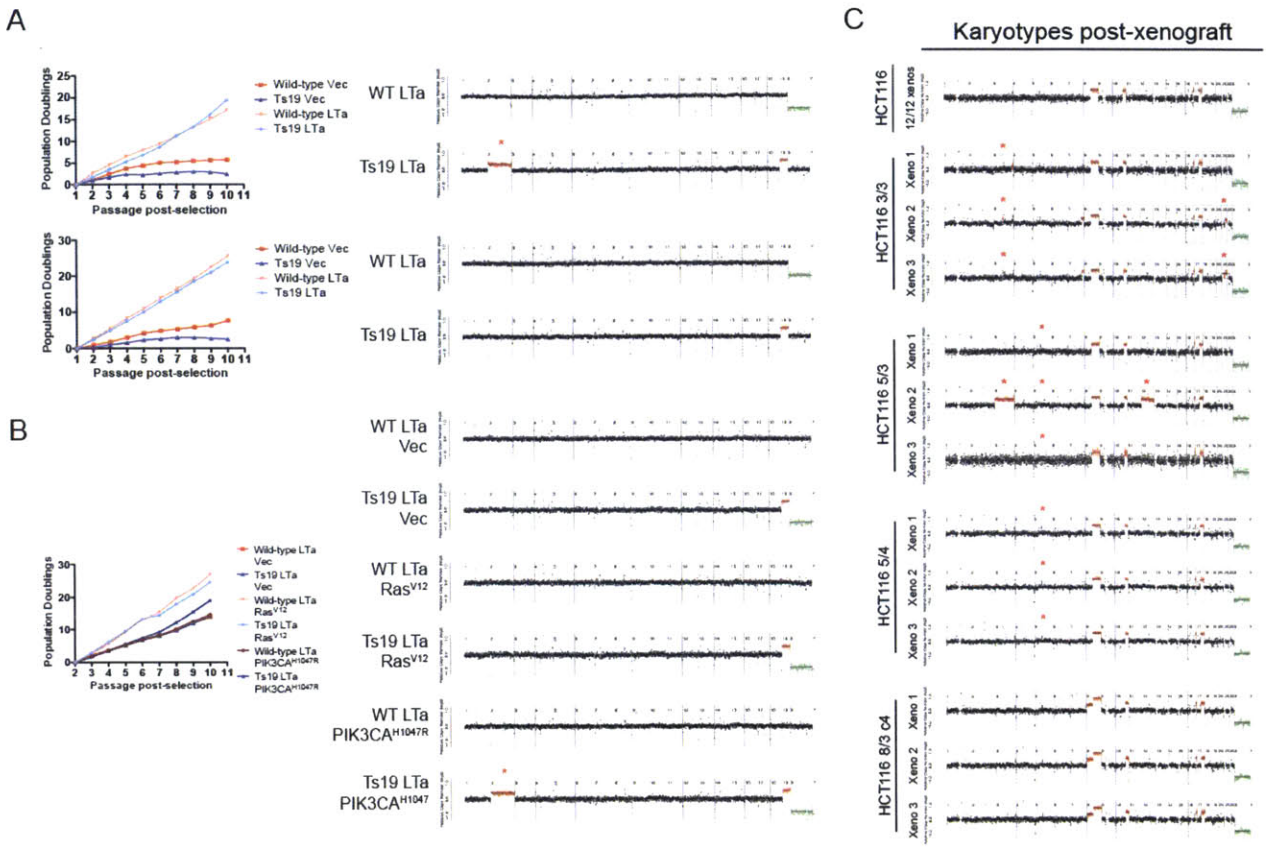
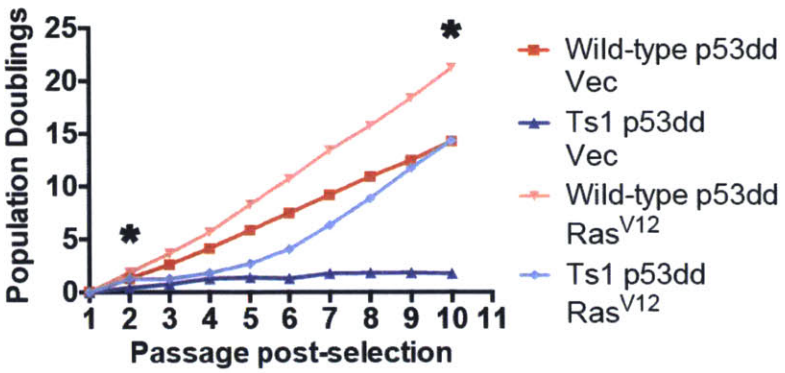


Figure 7. Karyotype evolution correlates with enhanced growth in aneuploid cell lines.

(A) Two independent Ts19 and matched euploid control cell lines were transduced with Large T or with an empty vector. Sequencing of the fast-growing Ts19 line at passage 7 revealed an extra copy of chromosome 2 (indicated with an asterisk), while sequencing of the slower-growing line revealed no further karyotypic changes. (B) An independent Ts19 and matched euploid control cell line were transduced with Large T, and then transduced a second time with an empty vector, Ras^{V12}, or PIK3CA^{H1047R}. Ts19+LTa+PIK3CA^{H1047R} grew more rapidly an equivalently-transduced euploid line. Sequencing of all six lines at passage 10 revealed that Ts19+LTa+PIK3CA^{H1047R} had acquired an extra copy of chromosome 2, while the other cell lines showed no other deviations from the expected karyotypes. (C) Xenografts from figure 6 were extracted, digested with trypsin, and then plated on plastic. Low-pass whole genome sequencing revealed that 12 HCT116 xenografts and 3 HCT116 8/3 c4 xenografts maintained their initial karyotypes. However, all HCT116 3/3, 5/3, and 5/4 xenografts lost their initial

trisomies or tetrasomies during *in vivo* growth, and several lines displayed additional chromosomal copy number alterations. Deviations from each cell line's starting karyotypes are indicated with an asterisk.

To expand our analysis beyond Ts19, we next analyzed a Ts1+53dd+Ras^{V12} cell line. These cells initially doubled every ~50 hours. After 10 passages in culture, the trisomic cell line was observed to double every ~27 hours, a rate indistinguishable from p53dd+Ras^{V12}-transduced wild-type cells. Whole genome sequencing at early passage demonstrated that the euploid and trisomic cell lines maintained their initial karyotypes following transduction. However, at passage 10, a second round of genome sequencing demonstrated that the rapidly-growing Ts1+p53dd+Ras^{V12} cell line had lost the extra copy of chromosome 1, and instead displayed several other chromosome gains and losses, which were consistent with trisomies and pentasomies in a tetraploid population (Fig. S9). The Ts1+p53dd+Vec line, which continued to proliferate at a very low rate, maintained an extra copy of chromosome 1 at late passage. Thus, improved growth of an aneuploid cell line may also result from chromosome loss and/or tetraploidization.



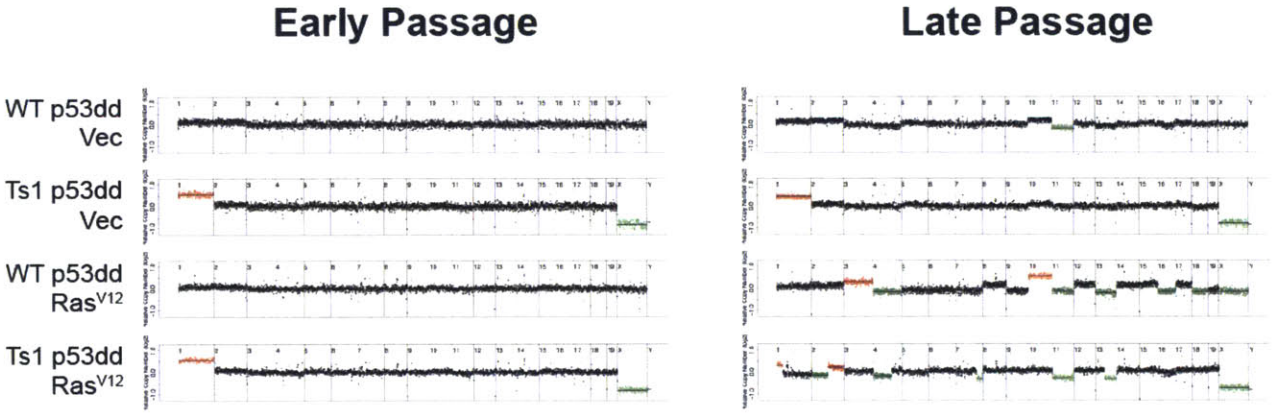


Figure S9. A rapidly-growing Ts1 line exhibits several karyotype changes. A Ts1 cell line and a matched euploid control line were transduced with p53dd and then transduced a second time with an empty vector or with Ras^{V12}. The Ts1 MEFs initially grew poorly, then evolved to grow at a rate indistinguishable from wild-type cells. Whole-genome sequencing at passage 2 revealed that the slow growing line maintained its initial karyotype, while whole-genome sequencing at passage 10 revealed that the rapidly growing line had lost the trisomy of mChr1, and displayed several further chromosomal gains and losses.

To analyze how cells adapted to aneuploidy *in vivo*, we re-derived cell lines from 12 euploid xenografts and from 3 HCT116 3/3, 5/3, 5/4, and 8/3 c4 xenografts. Following two passages in culture to deplete stromal cell contamination, each re-derived cell line was subjected to whole-genome sequencing. The euploid cell lines were chromosomally stable, and 12 out of 12 re-derived lines had karyotypes that were indistinguishable from the pre-xenograft cell line (Fig. 7C). However, every HCT116 3/3, 5/3 and 5/4 cell line was found to have lost its original trisomic or tetrasomic chromosome. 3 out of these 9 aneuploid lines exhibited other chromosomal alterations: one cell line initially trisomic for chromosome 5 gained an extra copy of chromosomes 3 and 12, while two cell lines initially trisomic for chromosome 3 gained an extra copy of chromosome 21. HCT116 8/3 c4, which proliferated at near-euploid levels, maintained its trisomy when grown as a xenograft, and did not exhibit any further chromosomal alterations. These results suggest that growth-inhibitory aneuploidies are selected against during *in vivo* tumor formation, while aneuploidies that are neutral can escape negative selection.

Discussion

Aneuploidy is a nearly-universal feature of human cancers. However, despite its frequent occurrence, we have found that in carefully-controlled experiments, aneuploid cells exhibit reduced tumorigenicity relative to genetically-matched euploid cells. Aneuploidy of 6 different chromosomes tested so far (mChr1, mChr13, mChr16, mChr19, hChr3, and hChr5) impedes cell proliferation, clonogenicity, anchorage-independent growth, and xenograft growth. Aneuploidy of only one chromosome (hChr8) has nearly-neutral consequences in these assays. Moreover, the activation of several oncogenic pathways (HRAS, BRAF, PIK3CA, and MYC) as well as the ablation of p53 and RB signaling are insufficient to overcome the fitness penalty induced by aneuploidy. We failed to detect any conditions in which aneuploidy promoted the

transformation of primary cells, synergized with an oncogenic mutation, or otherwise contributed to tumorigenesis. Indeed, we have found that several chromosome gains are actively selected against.

While these results are unexpected, they are consistent with much of what has been learned about the effects of aneuploidy on normal cell physiology. Aneuploid chromosomes are transcribed and translated proportional to their copy number (22,24,41,42), which can lead to stoichiometric imbalances in endogenous proteins and protein complexes (28,43). To compensate, cells rely on a set of protein quality control mechanisms, including the HSF1/HSP90 folding pathway (28,40), autophagy, and proteasomal degradation (41,42). The energetic cost of expressing, folding, and turning over excess proteins, as well as the downstream consequences of unmitigated protein imbalances, impose a significant fitness cost on the cell. The severity of aneuploid phenotypes is generally proportional to the degree of aneuploidy, and indeed, we found that transformed cells trisomic for mouse chromosome 1 grew much more slowly than cells trisomic for mouse chromosome 19. Moreover, while studies of mouse models of Down syndrome have identified certain genes whose triplication may have tumor-protective effects (17–19), our results suggest that whole-chromosome aneuploidy itself can function as a powerful tumor suppressor.

Why, then, is aneuploidy such a frequent occurrence in cancer? Several possibilities remain. First, it may be the case that aneuploidy nearly always results in a proliferative disadvantage, but in very rare circumstances, involving specific cell types or chromosomes, aneuploidy may confer a benefit. For example, trisomy of chromosome 21 predisposes individuals to leukemia (5), and gain of chromosome 21 is a common occurrence in sporadic leukemia (6,44), but trisomy of chromosome 21 appears to protect against the development of many other cancer types, including breast, lung, and prostate cancers (16). Thus, while all nine aneuploid strains that we examined hinder or are neutral with regard to tumor growth, it is conceivable that a wider survey of aneuploidies or cell types would reveal unusual cases in

which aneuploidy provides a growth advantage. Secondly, chromosome missegregation, rather than aneuploidy *per se*, could be a crucial driver of tumorigenesis. Lagging chromosomes can be shattered during anaphase (45) or hypermutated following encapsulation in a micronucleus (46); missegregation-induced DNA damage could therefore promote transformation while any subsequent aneuploidy exists mainly as a "passenger" mutation. Similarly, reductive mitoses from a tetraploid intermediate may generate tumor-initiating gross chromosomal rearrangements with a large number of aneuploid chromosome "passengers" (47). Finally, our study has examined the consequences of single-chromosome trisomy and tetrasomy, as monosomies and complex aneuploidies are incompatible with mouse development. Nonetheless, it remains possible that chromosome loss events, or the gain of multiple chromosomes, could promote tumorigenesis while trisomies and tetrasomies do not. Novel genetic tools will be required to model these types of karyotype changes without inducing gross CIN.

Lastly, our results suggest one way that cells may adapt to aneuploidy. Following serial passaging or growth *in vivo*, aneuploid cells frequently exhibited further karyotypic alterations, and these changes correlated with improved growth. Aneuploid cells can revert to euploidy by losing their extra chromosomes, and we have demonstrated that this is a common occurrence when the extra chromosome(s) induces a significant growth disadvantage. Alternately, cells can acquire other chromosome copy number changes, including both chromosome gains and losses. In particular, the gain of chromosome 2 correlates with enhanced growth in multiple independent experiments with trisomy 19. We hypothesize that these changes can act to buffer the negative consequences of aneuploidy, potentially by balancing certain stoichiometric imbalances caused by the initial trisomy. Consistent with this notion, many distinct chromosome copy number alterations are observed together in the same tumors more often than expected by chance (6). Chromosome gain events are more likely to be associated with other chromosome gain events, and vice-versa (6). For instance, tumors that have gained an extra copy of

chromosome 7 are significantly more likely to have also gained an extra copy of chromosome 17, while loss of chromosome 7 is correlated with loss of chromosome 17 (6). Such changes may function to maintain stoichiometry in key protein complexes or pathways. Higher-resolution studies of correlations among single-gene copy-number alterations in cancers will shed light on this topic and may explain the multiplicity of aneuploidies found in cancer.

Materials and Methods

MEF derivation, culture and transduction. A Robertsonian breeding scheme was utilized to generate sibling-matched euploid and trisomic MEFs as described in ref (23). MEFs were cultured in DMEM supplemented with 10% FBS, 2mM glutamine, and 100 U/ml penicillin and streptomycin. Cells were maintained at 37° C and 5% CO2 in a humidified environment. Cell counting was performed using the Cellometer Auto T4 system. Plasmids encoding oncogenes were obtained from Addgene (https://www.addgene.org/) and then transfected into the Phoenix-Eco cell line (48) using TransIT-LT1 (Mirus). Viral supernatants were collected 24, 48, and 72 hours post-transfection, and were applied to freshly-split passage 2 MEFs. Transduced cells were selected by FACS, or by the addition of puromycin (1.6 μg/ml), hygromycin (200 μg/ml), or G418 (1 mg/ml).

Human colon cancer cell culture. Aneuploid cell lines derived from HCT116 were previously described in refs (24) and (40). Cells were cultured in DMEM supplemented with 10% FBS, 2mM glutamine, and 100 U/ml penicillin and streptomycin. Cells were maintained at 37° C and 5% CO2 in a humidified environment.

Low-pass whole genome sequencing. Sequencing reactions were performed at the MIT BioMicro Center. 50 ng of purified DNA from each cell line was prepared and barcoded using Nextera reagents (Illumina), and tagmented material was PCR amplified for seven cycles. Libraries were quantified using an AATI Fragment Analyzer before pooling. Libraries were sequenced (40bp read length) on an Illumina HiSeq2000. Reads were demultiplexed using custom scripts allowing single mismatches within the reference barcode.

Sequence reads were trimmed to 40 nucleotides and aligned to the mouse (mm9) or human (hg19) genomes using BWA (0.6.1) with default options(49). HMMcopy (0.1.1) was used to detect copy number alterations by estimating copy number in 500-kb bins controlling for mappability [downloaded from UCSC Genome Bioinformatics (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/ or http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeMapability/)] and GC content (calculated by HMMcopy gcCounter) (50).

Cell proliferation and tumorigenicity assays. For proliferation assays, MEFs and HCT116 cells were passaged using a modified 3T3 protocol (51). $3x10^5$ cells were plated in three wells of a 6-well plate, and cells were combined, counted, and re-plated at the same density every third day. For focus formation assays, 1000 cells (MEFs) or 200 cells (HCT116) were plated in triplicate on 10cm plates, and then allowed to grow for 10 days (MEFs) or 14 days (HCT116). Subsequently, colonies were fixed with ice-cold 100% methanol for 10 minutes, and then stained with a solution of 0.5% crystal violet in 25% methanol for 10 minutes. For soft agar assays, a 1% base layer of Difco Agar Noble was prepared and then mixed with an equal amount of 2X DMEM. The solution (0.5% agar in 1X DMEM) was then added to each well of a 6-well plate and allowed to solidify. Subsequently, a top layer of 0.7% agar was prepared and mixed with an equal volume of a 2X solution of DMEM containing 10,000 cells (MEFs) or 2000

cells (HCT116) and added to the base layer in triplicate. The plates were incubated for 20 days at 37° C before being imaged.

For xenograft studies, 5 to 10 female, 5-week old Nu/J mice (Jackson Laboratory Stock 002019) were utilized for each experiment. Cells to be injected were harvested and concentrated to 10^7 (MEFs) or $4x10^7$ (HCT116) cells/ml in PBS. 100µl of the solution was injected subcutaneously into the rear flanks of each mouse using a 25 gauge needle. Euploid and aneuploid cell lines were typically injected contralaterally, with the exception of experiments involving cell lines transduced with LTa and Ras^{V12}, in which only one cell line was injected into each animal. Tumors were measured every third day using calipers, and tumor volumes were calculated using the formula $0.5 \times A \times B^2$, where A is the longer diameter and B is the shorter diameter. All animal studies and procedures were approved by the MIT Institutional Animal Care and Use Committee.

Beta-galactosidase staining. 5000 cells of each strain were plated in triplicate in a 48-well plate, allowed to attach overnight, and then stained using a Senescence Histochemical Staining Kit (Sigma-Aldritch). Cells were incubated in the X-gal solution overnight at 37° C and then imaged the subsequent day.

References

1. Boveri T. Concerning the Origin of Malignant Tumours by Theodor Boveri. Translated and annotated by Henry Harris. J Cell Sci. 2008;121:1–84.

2. Weaver BA, Cleveland DW. Does aneuploidy cause cancer? Curr Opin Cell Biol. 2006;18:658–67.

3. Holland AJ, Cleveland DW. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. Nat Rev Mol Cell Biol. 2009;10:478–87.

4. Gordon DJ, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer. Nat Rev Genet. 2012;13:189–203.

5. Seewald L, Taub JW, Maloney KW, McCabe ERB. Acute leukemias in children with Down syndrome. Mol Genet Metab. 2012;107:25–30.

6. Ozery-Flato M, Linhart C, Trakhtenbrot L, Izraeli S, Shamir R. Large-scale analysis of chromosomal aberrations in cancer karyotypes reveals two distinct paths to aneuploidy. Genome Biol. 2011;12:R61.

7. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 2013;45:1134–40.

8. Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, et al. Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns to Shape the Cancer Genome. Cell. 2013;155:948–62.

9. Michel LS, Liberal V, Chatterjee A, Kirchwegger R, Pasche B, Gerald W, et al. MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. Nature. 2001;409:355–9.

10. Sotillo R, Hernando E, Díaz-Rodríguez E, Teruya-Feldstein J, Cordón-Cardo C, Lowe SW, et al. Mad2 Overexpression Promotes Aneuploidy and Tumorigenesis in Mice. Cancer Cell. 2007;11:9–23.

11. Sotillo R, Schvartzman J-M, Socci ND, Benezra R. Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal. Nature. 2010;464:436–40.

12. Li M, Fang X, Wei Z, York JP, Zhang P. Loss of spindle assembly checkpoint-mediated inhibition of Cdc20 promotes tumorigenesis in mice. J Cell Biol. 2009;185:983–94.

13. Park I, Lee H, Choi E, Lee Y-K, Kwon M-S, Min J, et al. Loss of BubR1 acetylation causes defects in spindle assembly checkpoint signaling and promotes tumor formation. J Cell Biol. 2013;202:295–309.

14. Jeganathan K, Malureanu L, Baker DJ, Abraham SC, Deursen JM van. Bub1 mediates cell death in response to chromosome missegregation and acts to suppress spontaneous tumorigenesis. J Cell Biol. 2007;179:255–67.

15. Li M, Fang X, Baker DJ, Guo L, Gao X, Wei Z, et al. The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. Proc Natl Acad Sci USA. 2010;107:14188–93.

16. Nižetić D, Groet J. Tumorigenesis in Down's syndrome: big lessons from a small chromosome. Nat Rev Cancer. 2012;12:721–32.

17. Sussan TE, Yang A, Li F, Ostrowski MC, Reeves RH. Trisomy represses ApcMin-mediated tumours in mouse models of Down's syndrome. Nature. 2008;451:73–5.

18. Baek K-H, Zaslavsky A, Lynch RC, Britt C, Okada Y, Siarey RJ, et al. Down's syndrome suppression of tumour growth and the role of the calcineurin inhibitor DSCR1. Nature. 2009;459:1126–30.

19. Reynolds LE, Watson AR, Baker M, Jones TA, D'Amico G, Robinson SD, et al. Tumour angiogenesis is reduced in the Tc1 mouse model of Down's syndrome. Nature. 2010;465:813–7.

20. Weaver BAA, Silk AD, Montagna C, Verdier-Pinard P, Cleveland DW. Aneuploidy Acts Both Oncogenically and as a Tumor Suppressor. Cancer Cell. 2007;11:25–36.

21. Silk AD, Zasadil LM, Holland AJ, Vitre B, Cleveland DW, Weaver BA. Chromosome missegregation rate predicts whether aneuploidy will promote or suppress tumors. Proc Natl Acad Sci U S A. 2013;110:E4134–41.

22. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, et al. Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. Science. 2007;317:916–24.

23. Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, et al. Aneuploidy Affects Proliferation and Spontaneous Immortalization in Mammalian Cells. Science. 2008;322:703–9.

24. Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol Syst Biol. 2012;8.

25. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, et al. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature. 2010;468:321–5.

26. Thompson SL, Compton DA. Examining the link between chromosomal instability and aneuploidy in human cells. J Cell Biol. 2008;180:665–72.

27. Thompson SL, Compton DA. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. J Cell Biol. 2010;188:369–81.

28. Oromendia AB, Dodgson SE, Amon A. Aneuploidy causes proteotoxic stress in yeast. Genes Dev. 2012;26:2696–708.

29. Tang Y-C, Williams BR, Siegel JJ, Amon A. Identification of Aneuploidy-Selective Antiproliferation Compounds. Cell. 2011;144:499–512.

30. Sheltzer JM, Torres EM, Dunham MJ, Amon A. Transcriptional consequences of aneuploidy. Proc Natl Acad Sci. 2012;109:12644–9.

31. Sheltzer JM. A Transcriptional and Metabolic Signature of Primary Aneuploidy Is Present in Chromosomally Unstable Cancer Cells and Informs Clinical Prognosis. Cancer Res. 2013;73:6401–12.

32. Dürrbaum M, Kuznetsova AY, Passerini V, Stingele S, Stoehr G, Storchová Z. Unique features of the transcriptional response to model aneuploidy in human cells. BMC Genomics. 2014;15:139.

33. Duesberg P, Rasnick D, Li R, Winters L, Rausch C, Hehlmann R. How aneuploidy may cause cancer and genetic instability. Anticancer Res. 1999;19:4887–906.

34. Rasnick D, Duesberg PH. How aneuploidy affects metabolic control and causes cancer. Biochem J. 1999;340:621–30.

35. Ahuja D, Sáenz-Robles MT, Pipas JM. SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. Oncogene. 2005;24:7729–45.

36. Shaulian E, Zauberman A, Ginsberg D, Oren M. Identification of a minimal transforming domain of p53: negative dominance through abrogation of sequence-specific DNA binding. Mol Cell Biol. 1992;12:5581–92.

37. Gallimore PH, Turnell AS. Adenovirus E1A: remodelling the host cell, a life or death experience. Oncogene. 2001;20:7824–35.

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

39. Shin SI, Freedman VH, Risser R, Pollack R. Tumorigenicity of virus-transformed cells in nude mice is correlated specifically with anchorage independent growth in vitro. Proc Natl Acad Sci. 1975;72:4435–9.

40. Donnelly N, Passerini V, Dürrbaum M, Stingele S, Storchová Z. HSF1 deficiency and impaired HSP90-dependent protein folding are hallmarks of aneuploid human cells. EMBO J. 2014;33:2374–87.

41. Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, et al. Identification of Aneuploidy-Tolerating Mutations. Cell. 2010;

42. Dephoure N, Hwang S, O'Sullivan C, Dodgson SE, Gygi SP, Amon A, et al. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. eLife. 2014;3.

43. Sheltzer JM, Amon A. The aneuploidy paradox: costs and benefits of an incorrect karyotype. Trends Genet TIG. 2011;27:446–53.

44. Loncarevic IF, Roitzheim B, Ritterbach J, Viehmann S, Borkhardt A, Lampert F, et al. Trisomy 21 is a recurrent secondary aberration in childhood acute lymphoblastic leukemia with TEL/AML1 gene fusion. Genes Chromosomes Cancer. 1999;24:272–7.

45. Janssen A, van der Burg M, Szuhai K, Kops GJPL, Medema RH. Chromosome Segregation Errors as a Cause of DNA Damage and Structural Chromosome Aberrations. Science. 2011;333:1895–8.

46. Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova EV, Pan Y, et al. DNA breaks and chromosome pulverization from errors in mitosis. Nature. 2012;482:53–8.

47. Fujiwara T, Bandi M, Nitta M, Ivanova EV, Bronson RT, Pellman D. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. Nature. 2005;437:1043–7.

48. Swift S, Lorens J, Achacoso P, Nolan GP. Rapid production of retroviruses for efficient gene delivery to mammalian cells using 293T cell-based systems. Curr Protoc Immunol Ed John E Coligan Al. 2001;Chapter 10:Unit 10.17C.

49. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754–60.

50. Ha G, Roth A, Lai D, Bashashati A, Ding J, Goya R, et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome Res. 2012;22:1995–2007.

51. Todaro GJ, Green H. Quantitative Studies of the Growth of Mouse Embryo Cells in Culture and Their Development into Established Lines. J Cell Biol. 1963;17:299–313.

Chapter 6: Conclusions and Future Directions

Aneuploidy, Genomic Instability, and Cancer

In Chapter 2, I reported that unbalanced karyotypes were sufficient to cause genomic instability in budding yeast, which could potentially accelerate the acquisition of growth-promoting genetic alterations. I demonstrated that single chromosome gains increased the rates of forward mutation, mitotic recombination, chromosome loss, and double-strand break formation. I also established that these phenotypes were due to stoichiometric imbalances in proteins encoded on the extra chromosomes (1). The chromosomal instability (CIN) phenotype of aneuploid yeast has since been confirmed by the Li lab (2), and further work in our lab has linked the DNA damage phenotypes that I initially observed to widespread defects in DNA replication (3). A crucial question, however, is to what extent these phenotypes are present in aneuploid mammalian cells, and whether they affect the genomes of aneuploid tumors.

Several recent studies have now demonstrated that single-chromosome aneuploidy in human cells is sufficient to induce a low level of CIN: FISH experiments in lymphocytes from individuals with Turner syndrome (an XO karyotype), trisomy 13, trisomy 18, or trisomy 21 revealed a greater degree of karyotype heterogeneity than was observed in lymphocytes from euploid individuals (4,5). FISH studies of amniocytes from fetuses with Klinefelter syndrome (an XXY karyotype), trisomy 18, or trisomy 21 reached a similar conclusion (6). The most comprehensive characterization of CIN in aneuploid human cells has been reported by Nicholson et al. (2015), who demonstrated via several independent methods that aneuploidy in a karyotypically-stable colon cancer cell line increased the frequency of lagging chromosomes and led to occasional cytokinesis failure (7). They further demonstrated that in cells trisomic for chromosome 13, an extra copy of a single gene present on chromosome 13 (SPG20) was responsible for this phenotype, as over-expression of the gene had a dominant-negative effect on the cytokinetic machinery. In total, these results suggest that our earlier findings in yeast are also conserved in mammalian cells.

The strong link between aneuploidy and CIN in mammals suggests one resolution to a key paradox concerning tumorigenesis: despite the high levels of aneuploidy frequently found in tumors, very few mutations have been identified in tumor genomes that are sufficient to cause CIN. Tumors generally have functional spindle checkpoints and rarely acquire mutations in the spindle assembly checkpoint (SAC) (8). Mutations in the cohesin subunit *STAG2* have been proposed to contribute to aneuploidy in human tumors (9). However, *STAG2* mutations are rare (occurring in <2% of cancers), and loss of Stag2 expression is not associated with aneuploidy in several tumor types (10–12). Thus, how cancers become aneuploid remains largely unknown. Our results in yeast, coupled with the above studies in mammalian cells, suggest that aneuploidy may be an auto-catalytic event: the acquisition of aneuploidy, even in a chromosomally-stable background, can cause CIN, thereby destabilizing the karyotype and leading to more aneuploidy. Testing this will require analyzing the long-term fates of matched euploid and aneuploid cell populations to determine at what rates the karyotypes diverge.

In addition to documenting CIN in aneuploid yeast, I also found that disomic yeast strains displayed elevated levels of spontaneous DNA damage and mitotic recombination. Whether aneuploidy can also cause structural DNA damage in mammalian cells is unknown, and some evidence suggests that this phenotype is not conserved beyond yeast. I found that trisomic MEFs do not harbor a higher level of DNA breaks, as judged by H2AX and 53BP1 staining (unpublished results). Furthermore, when I induced DNA damage with radiation or with a transient treatment with phleomycin, H2AX and 53BP1 foci resolved in the trisomic MEFs with wild-type kinetics (unpublished results). These findings were not entirely unexpected, as trisomy in diploid yeast caused significantly less genomic instability than disomy in haploid yeast did (Chapter 2). Furthermore, it was previously known that p53 was not constitutively activated in trisomic MEFs, which DNA damage would be expected to induce (13). It could be the case that the specific trisomies that I examined (mChr 13, 16, and 19) do not accumulate DNA

damage while other trisomies do. Alternately, it may be the case that more complex aneuploidies induce DNA damage, while single-chromosome trisomies do not. We have not tested these questions, and to the best of my knowledge no other investigators have either. However, examining DNA damage in high-complexity aneuploid strains will be crucial to understanding the link between numeric and structural chromosome changes in cancer.

Lastly, it is important to understand the mechanism(s) inducing genomic instability in aneuploid cells. We found that buffering aneuploidy-induced stoichiometric imbalances suppresses aneuploidy-induced genomic instability. However, the severities of these phenotypes are not correlated with the size of the disomic chromosomes, unlike nearly every other phenotype that we have documented in these strains. Thus, it could be the case that single-copy over-expression of individual genes (randomly distributed across several chromosomes) causes genomic instability. One way to potentially address this would be to transform a wild-type strain with a pooled yeast genomic library on CEN plasmids, grow the transformants +/- phleomycin, and then sequence the final populations to identify genes on CEN plasmids that dropped out during growth in phleomycin. Alternately, various forms of genomic instability could be measured in single-gene over-expression experiments, as Bonney et al. (2015) did for proliferation (14). These experiments could identify genes and biological pathways whose duplication drives genomic instability.

The transcriptional consequences of aneuploidy in primary cells

In Chapter 3, I examined the effects of aneuploidy on gene expression in aneuploid primary cells from various species. I reported that aneuploidy induces a transcriptional stress response that is independent of the identity of the extra chromosome and is remarkably well-conserved among eukaryotes. Aneuploid cells in different organisms down-regulate an

overlapping set of cell cycle and RNA processing transcripts while up-regulating membrane, extracellular, and stress response genes. Following the publication of the results described in Chapter 3, Stingele et al. (2012) came to similar conclusions based on their study of aneuploidies in human RPE-1 and HCT116 cells (15). Namely, aneuploidies of different chromosomes induced similar gene expression changes, and cell cycle and membrane-related transcripts were particularly affected by aneuploidy. Stingele et al. also applied SILAC to measure proteome changes in aneuploid strains, and demonstrated that these gene expression patterns were present at the protein level as well. Some of these changes make sense in light of what we know about aneuploid biology: the down-regulation of mitotic transcripts likely reflects the fact that aneuploid cells progress slowly through the cell cycle. However, other changes have more mysterious origins and consequences. The Gene Ontology term most up-regulated in aneuploid primary cells is the extracellular matrix, and some extracellular matrix genes are expressed 50 to 100-fold higher in trisomic MEFs than in euploid MEFs (data not shown). Why this pathway is induced by aneuploidy is unknown. Future work on this question will involve examining the extracellular matrix in aneuploid cells and determining how aneuploid cells interact with their environment. Do aneuploid cells have different adhesive properties? Do they exhibit motility alterations, or increased invasiveness? Do they induce any type of cell non-autonomous response in surrounding cell populations? Additionally, several of the most up-regulated genes in trisomic cells should be knocked down to determine the effect(s) of their depletion on aneuploid cell populations.

We also found that the transcriptional changes induced by aneuploidy were significantly correlated across different eukaryotic species. We believe that these changes represent a universal environmental stress/slow growth response, and are how cells respond to less-than-ideal growth conditions. It is currently unknown whether some of the gene expression changes exhibited by trisomic MEFs represent an "aneuploidy-specific" response. This could be

investigated by looking for differences between the transcriptomes of trisomic MEFs and MEFs grown under various other stresses. Additionally, the aneuploidy stress signature was most strongly apparent in cells grown *in vitro*, and was masked in neural tissue from trisomic embryos. We believe that the transcriptional signature that we detected represents the effects of aneuploidy when cells are stimulated to proliferate without restraint (e.g., bathed in nutrients and mitogenic signals in tissue culture). Under such conditions, the growth suppression caused by aneuploidy would be most apparent. The transcriptional consequences of aneuploidy *in vivo* – for instance, in a homeostatic tissue where most cells are in G0/G1 and are not dividing - could be different and should be investigated more deeply. At one point, we considered the possibility that discovering universal gene expression markers of aneuploidy could lead to the development of novel prenatal diagnosis tools, though the subsequent success of maternal plasma sequencing at identifying trisomic embryos has rendered this idea rather moot (16).

In our analysis, the transcriptional responses to aneuploidy between species were similar but not identical. Ribosome subunits were most strongly down-regulated in yeast and plants, while cell cycle transcripts were most strongly down-regulated in mice and humans. Apparently, growth in plants and fungi is controlled by ribosome abundance, while growth in mammals is controlled by cell cycle transcript abundance. (Alternately, the causality could be reversed: cells could respond to slow growth by conserving energy by down-regulating ribosomal or cell cycle genes). Why these differences exist is unknown, and the evolutionary history of this difference could be fascinating to explore. At what point in the evolutionary tree did translation cease being the most energetically-costly biological process? Why is the downregulation of cyclins and DNA replication factors not a part of the yeast environmental stress response? Is the proliferation rate of yeast insensitive to the levels of these genes? We could begin to address the former question by conducting a meta-analysis on gene expression profiles, +/- several stresses, in species from diverse evolutionary lineages. The latter question

could be addressed by looking for haploinsufficiency in diploid strains harboring single copies of these genes.

The transcriptional consequences of aneuploidy in cancer

In Chapter 4, I reported that the acquisition of aneuploidy in various cancer cell lines mimicked the transcriptional stress response that I observed in aneuploid primary cells. Trisomic MEFs, trisomic human cells, CIN MEFs, and CIN cancer cells were found to share a limited and overlapping set of transcriptional changes. Additionally, I demonstrated that a previous gene expression signature of CIN, called CIN70, measured cancer cell proliferation, rather than CIN. Lastly, I found that transcriptional markers of aneuploidy stress correlated with poor clinical prognosis in a variety of cancer types in a proliferation-independent manner. One conclusion that I draw from this work is that there is an important difference between clonal and subclonal aneuploidy in tumors. The CIN70 gene signature was constructed by identifying genes whose expression was increased in structurally-complex tumors (17). However, I found that structural complexity does not actually correlate with chromosomal instability, as determined by measuring cell-to-cell karyotype variability. Thus, while CIN70, structural complexity, and proliferation rate are significantly correlated, real CIN is correlated with slow growth and a distinct set of transcriptional changes. These findings suggest to me that there are two types of aneuploidy, broadly defined, in cancer. The first type of aneuploidy is clonal, and consists predominantly of focal or megabase-scale aneuploidies targeting specific oncogenes and tumor suppressors. In certain circumstances, aneuploidies of this type, like amplifications that encompass *MYC*, or deletions that encompass *TP53*, would be beneficial, and therefore have a selective advantage within tumors. These aneuploidies could rise to clonal levels within a tumor due to the proliferative advantage that they confer, and this would explain why the CIN70 gene signature reports tumor doubling times. On the other hand, sub-

clonal aneuploidy would result from chromosome segregation errors, and would not be selected for by tumors. As these aneuploidies are random and unselected, they still induce a transcriptional stress response that is analogous to the stress response in primary aneuploid cells. At the same time, gene sets derived from legitimate CIN in cancer cell lines (HET70 and HET5) are also associated with poor prognosis in several cancer types. Thus, sub-clonal aneuploidy could provide an evolutionary benefit as well, perhaps by diversifying tumor phenotypes through chromosomal copy number changes. It is important to note that this analysis was conducted on cancer cell lines grown *in vitro*, and it remains possible that, *in vivo*, CIN exerts a different effect on transcription. The effects of CIN on gene expression *in vivo* could be directly tested in xenografts or autochthonous mouse cancer models that harbor inducible hairpins targeting *MAD2* or *MPS1*.

One puzzling discovery that I reported concerned the high degree of tissue specificity for the stratification of patient risk by PCNA25 (the gene signature that reports the average expression level of the 25 genes whose expression correlates most strongly with the proliferative marker PCNA). While high PCNA25 expression was significantly correlated with poor survival in breast, bladder, brain, and liver cancers as well as sarcomas, PCNA25 was uncorrelated with survival in lung, ovarian, or colorectal cancers. These results are consistent with previous reports suggesting that slowly-dividing colorectal tumors are more deadly than rapidly-dividing tumors (18–20), though to my knowledge this has not been reported in ovarian or lung cancers. It is strange that in certain tissues, slowly-dividing tumors are the most lethal (or, at least, are not associated with improved prognosis). Why this is true warrants significant investigation. It could be the case that slowly-dividing tumors are more metastatic than rapidly-dividing cancers [though the opposite is true in breast cancer (21)]. Alternately, slow division could promote immune evasion or chemotherapy resistance in cancer cells. It is important to note that slow division could be a "driver" or a "passenger" of the highly-aggressive phenotype.

A set of genomic alterations could induce both slow growth and metastatic capabilities. Or, slow growth itself, for whatever reason, could directly affect the downstream aggressiveness of the tumor. Testing these questions in mouse models could be difficult. Potentially, one could construct ovarian and colorectal cancer cell lines that have *CDK4* transcription under the control of a titratable promoter, and then one could test how different CDK4 levels affect cancer aggressiveness *in vitro* and *in vivo*. If the induction of slow growth has certain pro-metastatic consequences, then this should carefully inform how chemotherapy is applied in these cancers.

Aneuploidy can function as a tumor suppressor

In Chapter 5, I reported that transformed trisomic fibroblasts exhibit reduced tumorigenicity relative to transformed euploid fibroblasts. Similarly, the addition of extra chromosomes into a chromosomally-stable human cancer cell line inhibits or is neutral with regard to tumor growth. These results are quite surprising, given the prevalence of aneuploidy in cancer. However, this suggests that in certain circumstances aneuploidy can function as a tumor suppressor, rather than a tumor promoter.

While the preponderance of evidence suggests that aneuploidy drives tumor growth, several prior observations are consistent with the findings presented in this work. In particular, it has long been recognized that individuals with trisomy 21 are at lower risk of developing many common solid tumors (22). Additionally, it is worth noting that many studies that suggest that aneuploidy functions to promote tumor growth have been conducted in chromosomally-unstable mice or cell lines (23–25). Phenotypes observed in CIN mice are attributed to aneuploidy, though there is often little evidence to support this inference. (The work presented in this thesis could be accused of making a similar extrapolation: we have predominantly studied single-chromosome gains, but we suggest that our results may hold true for chromosome loss events and higher-complexity aneuploidies as well). Nonetheless, it is worth considering potential

distinctions between CIN and aneuploidy. Mitosis in a SAC-deficient cell could lead to the accumulation of DNA damage, if lagging chromosomes are caught by micronuclei or damaged by the cytokinesis machinery (26,27). Additionally, SAC components may have functions outside of the spindle checkpoint. For instance, BubR1 has been suggested to function in the DNA damage response and in the maintenance of sister chromatid cohesion (28,29). Tumorigenesis assays in BubR1-deficient mice are therefore complicated by the pleiotropic functions of the wild-type gene product (30). Finally, CIN mice show a wide range of phenotypes: mice harboring *BUB1B* mutations exhibit progeria, a phenotype not observed in other CIN mice, and the spectra of spontaneous tumor formation varies greatly between different CIN models (31,32). Thus, distinct CIN-inducing lesions may produce unique consequences, the reasons for which are poorly understood.

In our experimental system, several oncogene combinations (particularly involving LTa and Ras^{V12}) equalized doubling times between euploid and trisomic MEFs when grown using a 3T3 protocol. However, when challenged in other assays – colony formation, growth in soft agar, or growth as xenografts – the transduced aneuploid cell lines uniformly performed worse than the wild-type MEFs, regardless of the oncogenic alleles that they harbored. It is not clear what the differences between each of these assays are. It has been reported that anchorage-independent growth is the best proxy for *in vivo* tumorigenesis (33). At the same time, only a small percentage of tumors that arise naturally in humans can be successfully propagated in cell culture [one such study reports the generation of cell lines from only 13 of 200 tumors explanted (34)]. Moreover, our "gold standard" assay for tumor-forming potential – xenografts in nude mice – is performed with a large initial inoculum (10^6 cells), and the injected cells are grown in mice that lack the ability to mount an adaptive immune response. Thus, we recognize that our assays for the tumorigenicity of aneuploid cells are far from ideal. For that reason, in future

work, I will be developing tools to study the consequences of chromosomal imbalances on tumorigenesis *in vivo*.

Despite the aforementioned caveats, I believe that we have learned several facts about aneuploidy in cancer. First, many common oncogenic mutations are insufficient to equalize tumor-forming potential between euploid and aneuploid cell lines. In yeast, we observed that single-nucleotide mutations in several genes were sufficient to enhance the growth of aneuploid cells, and we hypothesized that similar aneuploidy-tolerating mutations would be found in cancer (35). We have examined the effects of every gene mutated in >10% of human tumors in aneuploid cells, and failed to find any that robustly suppressed the aneuploidy-induced defects or exerted a differential effect on aneuploid and euploid cells (36). Based on our results, it is unlikely that aneuploidy-tolerating events in tumors happen via single-basepair mutations, or, if they do happen by single nucleotide substitutions, then they do not affect a large fraction of aneuploid tumors. On the other hand, hundreds of genes exhibit altered copy numbers in human tumors, and the role or relevance of the vast majority of these alterations are unknown (37). Thus, some of these common but poorly-characterized deletions and amplifications (or gene expression changes occurring via other mechanisms) may contribute to aneuploidy tolerance in human tumors.

At our level of sequencing depth (0.1X), we were unable to detect small CNV's, thus could not address this potential mechanism for aneuploidy tolerance. However, we did discover that, in 7 out of 7 cases, improved growth in aneuploid cells was associated with further karyotypic changes. In two instances, gain of Chr2 correlated with improved growth in Ts19 cells. Rapidly-growing Ts1 cells were found to have lost their trisomy and apparently tetraploidized. Finally, trisomic HCT116's lost their extra chromosomes and exhibited several additional chromosome gains following growth *in vivo*. It is striking that these cell lines exhibited chromosomal copy number alterations beyond simply losing the extra chromosome(s). One

explanation for this phenomenon may be found in our experiments in yeast: disomy of yeast chromosome VI is inviable, as chromosome VI harbors the *TUB2* (β-tubulin) gene, and a single extra copy of *TUB2* is lethal (38–40). However, attempts to generate Disome VI frequently led to the recovery of strains that also contained extra copies of chromosome XIII, which harbors *TUB1* (α-tubulin). Gaining chromosomes VI and XIII at the same time restored stoichiometric balance in this crucial protein complex, and therefore the compound disomy was viable while Dis. VI alone was lethal. In aneuploid cancers, similar compensatory chromosome gains and losses may be common (41). It would be interesting to determine whether genes that encode proteins that function in the same complexes or pathways tend to be gained concurrently within tumors. I believe that a large-scale bioinformatic investigation may shed light on this question. One could collect all karyotype and CNV data from the various cancer genome databases (Mitelman, TCGA, etc.), and then perform a correlation analysis to identify loci that are commonly gained or lost together. This approach could explain why certain cancers exhibit recurrent, complex aneuploidies.

References

1. Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, et al. Aneuploidy Drives Genomic Instability in Yeast. Science. 2011;333:1026–30.

2. Zhu J, Pavelka N, Bradford WD, Rancati G, Li R. Karyotypic Determinants of Chromosome Instability in Aneuploid Budding Yeast. PLoS Genet. 2012;8:e1002719.

3. Blank HM, Sheltzer JM, Meehl CM, Amon A. Mitotic entry in the presence of DNA damage is a widespread property of aneuploidy in yeast. Mol Biol Cell. 2015;mbc.E14–10 – 1442.

4. Reish O, Brosh N, Gobazov R, Rosenblat M, Libman V, Mashevich M. Sporadic aneuploidy in PHA-stimulated lymphocytes of Turner's syndrome patients. Chromosome Res. 2006;14:527–34.

5. Reish O, Regev M, Kanesky A, Girafi S, Mashevich M. Sporadic aneuploidy in PHA-stimulated lymphocytes of trisomies 21, 18, and 13. Cytogenet Genome Res. 2011;133:184–9.

6. Biron-Shental T, Liberman M, Sharvit M, Sukenik-Halevy R, Amiel A. Amniocytes from aneuploidy embryos have enhanced random aneuploidy and signs of senescence - Can these findings be related to medical problems? Gene. 2015;562:232–5.

7. Nicholson JM, Macedo JC, Mattingly AJ, Wangsa D, Camps J, Lima V, et al. Chromosome mis-segregation and cytokinesis failure in trisomic human cells. eLife. 2015;e05068.

8. Varetti G, Pellman D, Gordon DJ. Aurea mediocritas: the importance of a balanced genome. Cold Spring Harb Perspect Biol. 2014;6:a015842.

9. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, et al. Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science. 2011;333:1039–43.

10. Balbás-Martínez C, Sagrera A, Carrillo-de-Santa-Pau E, Earl J, Márquez M, Vazquez M, et al. Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. Nat Genet. 2013;45:1464–9.

11. Djos A, Fransson S, Kogner P, Martinsson T. Aneuploidy in neuroblastoma tumors is not associated with inactivating point mutations in the STAG2 gene. BMC Med Genet. 2013;14:102.

12. Kim MS, Kim SS, Je EM, Yoo NJ, Lee SH. Mutational and expressional analyses of STAG2 gene in solid cancers. Neoplasma. 2012;59:524–9.

13. Tang Y-C, Williams BR, Siegel JJ, Amon A. Identification of Aneuploidy-Selective Antiproliferation Compounds. Cell. 2011;144:499–512.

14. Bonney ME, Moriya H, Amon A. Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. Genes Dev. 2015;29:898–903.

15. Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol Syst Biol. 2012;8.

16. Norton ME, Jacobsson B, Swamy GK, Laurent LC, Ranzini AC, Brar H, et al. Cell-free DNA Analysis for Noninvasive Examination of Trisomy. N Engl J Med. 2015;372:1589–97.

17. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet. 2006;38:1043–8.

18. Anjomshoaa A, Nasri S, Humar B, McCall JL, Chatterjee A, Yoon H-S, et al. Slow proliferation as a biological feature of colorectal cancer metastasis. Br J Cancer. 2009;101:822–8.

19. Salminen E, Palmu S, Vahlberg T, Roberts P-J, Söderström K-O. Increased proliferation activity measured by immunoreactive Ki67 is associated with survival improvement in rectal/recto sigmoid cancer. World J Gastroenterol WJG. 2005;11:3245–9.

20. Anjomshoaa A, Lin Y-H, Black MA, McCall JL, Humar B, Song S, et al. Reduced expression of a gene proliferation signature is associated with enhanced malignancy in colon cancer. Br J Cancer. 2008;99:966–73.

21. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol. 2010;11:174–83.

22. Nižetić D, Groet J. Tumorigenesis in Down's syndrome: big lessons from a small chromosome. Nat Rev Cancer. 2012;12:721–32.

23. Li M, Fang X, Baker DJ, Guo L, Gao X, Wei Z, et al. The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. Proc Natl Acad Sci USA. 2010;107:14188–93.

24. Michel LS, Liberal V, Chatterjee A, Kirchwegger R, Pasche B, Gerald W, et al. MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. Nature. 2001;409:355–9.

25. Sotillo R, Hernando E, Díaz-Rodríguez E, Teruya-Feldstein J, Cordón-Cardo C, Lowe SW, et al. Mad2 Overexpression Promotes Aneuploidy and Tumorigenesis in Mice. Cancer Cell. 2007;11:9–23.

26. Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova EV, Pan Y, et al. DNA breaks and chromosome pulverization from errors in mitosis. Nature. 2012;482:53–8.

27. Janssen A, van der Burg M, Szuhai K, Kops GJPL, Medema RH. Chromosome Segregation Errors as a Cause of DNA Damage and Structural Chromosome Aberrations. Science. 2011;333:1895–8.

28. Fang Y, Liu T, Wang X, Yang Y-M, Deng H, Kunicki J, et al. BubR1 is involved in regulation of DNA damage responses. Oncogene. 2006;25:3598–605.

29. Malmanche N, Owen S, Gegick S, Steffensen S, Tomkiel JE, Sunkel CE. Drosophila BubR1 Is Essential for Meiotic Sister-Chromatid Cohesion and Maintenance of Synaptonemal Complex. Curr Biol. 2007;17:1489–97.

30. Park I, Lee H, Choi E, Lee Y-K, Kwon M-S, Min J, et al. Loss of BubR1 acetylation causes defects in spindle assembly checkpoint signaling and promotes tumor formation. J Cell Biol. 2013;202:295–309.

31. Baker DJ, Jeganathan KB, Cameron JD, Thompson M, Juneja S, Kopecka A, et al. BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. Nat Genet. 2004;36:744–9.

32. Schvartzman J-M, Sotillo R, Benezra R. Mitotic chromosomal instability and cancer: mouse modelling of the human disease. Nat Rev Cancer. 2010;10:102–15.

33. Shin SI, Freedman VH, Risser R, Pollack R. Tumorigenicity of virus-transformed cells in nude mice is correlated specifically with anchorage independent growth in vitro. Proc Natl Acad Sci. 1975;72:4435–9.

34. Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H, et al. In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. J Natl Cancer Inst. 1973;51:1417–23.

35. Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, et al. Identification of aneuploidy-tolerating mutations. Cell. 2010;143:71–83.

36. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502:333–9.

37. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 2013;45:1134–40.

38. Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, et al. Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. Science. 2007;317:916–24.

39. Katz W, Weinstein B, Solomon F. Regulation of tubulin levels and microtubule assembly in Saccharomyces cerevisiae: consequences of altered tubulin gene copy number. Mol Cell Biol. 1990;10:5286–94.

40. Anders KR, Kudrna JR, Keller KE, Kinghorn B, Miller EM, Pauw D, et al. A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. BMC Genet. 2009;10:36.

41. Ozery-Flato M, Linhart C, Trakhtenbrot L, Izraeli S, Shamir R. Large-scale analysis of chromosomal aberrations in cancer karyotypes reveals two distinct paths to aneuploidy. Genome Biol. 2011;12:R61.

Appendix: Elite Male Faculty in the Life Sciences Employ Fewer Women

Reprinted from PNAS:

Sheltzer JM, Smith JC. Elite male faculty in the life sciences employ fewer women. Proc Natl Acad Sci U S A. 2014;111:10107–12.

Abstract

Women make up over half of all doctoral recipients in biology-related fields, but are vastly underrepresented at the faculty level in the life sciences. In order to explore the current causes of women's underrepresentation in biology, we collected publicly-accessible data from university directories and faculty websites about the composition of biology labs at leading academic institutions in the United States. We found that male faculty members tended to employ fewer female graduate students and postdoctoral researchers than female faculty members did. Furthermore, "elite" male faculty – those whose research was funded by the Howard Hughes Medical Institute, who had been elected to the National Academy of Sciences, or who had won a major career award – trained significantly fewer women than other male faculty members. In contrast, elite female faculty did not exhibit a gender bias in employment patterns. New assistant professors at the institutions we surveyed were largely comprised of postdocs from these prominent labs, and, correspondingly, the labs that produced assistant professors had an overabundance of male postdocs. Thus, one cause of the leaky pipeline in biomedical research may be the exclusion of women, or their self-selected absence, from certain high-achieving labs.

Introduction

Between 1969 and 2009, the percentage of doctorates awarded to women in the life sciences increased from 15% to 52% (1,2). Despite the vast gains at the doctoral level, women still lag behind in faculty appointments. Currently, only 36% of assistant professors and 18% of full professors in biology-related fields are women (3). The attrition of women from academic careers – known as the "leaky pipeline" problem (4) – undermines the meritocratic ideals of science and represents a significant underutilization of the skills that are present in the pool of doctoral trainees.

A variety of factors have been suggested to influence the leaky pipeline in science, technology, engineering, and math (the "STEM" fields). Early career aspirations and choice of undergraduate major are significant departure points for women in certain disciplines (5,6). For instance, women are awarded only 19% of bachelor's degrees in physics and 18% of bachelor's degrees in engineering, and correspondingly fewer women go on to graduate school in those subjects (1). In contrast, women are awarded >50% of both bachelor and doctoral degrees in biology, suggesting that major leaks in the pipeline occur at later points in professional development. Gender differences in individuals' personal aspirations may explain some attrition from the academy (7). For instance, in surveys of grad students and postdocs, women tend to rank work-life balance and parenthood-related issues as more important than men do, and the perceived difficulty of raising a family while working as a tenure-track faculty member causes more women than men to leave the academic pipeline (8–12). Such preferences are likely constrained by societal factors: male postdocs are more than twice as likely as female postdocs to expect their spouse to make career sacrifices for their benefit (8). Additionally, female scientists with children are significantly less likely to be hired for tenure-track jobs than those without children, while male scientists with children are *more* likely to be hired for tenure-track

jobs than males without children (13). Thus, a complex mixture of both free and constrained personal choices may contribute to the leaky pipeline in STEM.

In addition to the impact of gendered preference differences, the scarcity of female faculty may be in part due to persistent discrimination against women in science. Unlike systems of *de jure* (formal) discrimination, which were common up until the middle of the 20th century and often explicitly excluded women from certain career paths, discrimination in the present day more often results from *de facto* differences in the treatment of men and women. Such behavior is linked to the problem of cumulative (dis)advantages: small differences in access to scientific goods (i.e., resources, mentoring, public visibility, etc.) may spiral over time, leading to significant divergence in achievement over the course of a career (14). These biases have been documented in both correlational and experimental studies of academic science. For instance, Moss-Racusin and colleagues (2012) sent science faculty identical resumes for a lab manager position in which only the name and gender of the applicant were changed (15). The applicant with the male name was judged to be more competent, hirable, and was offered a larger starting salary than the female applicant.

How these gender biases affect the advancement of women in science is poorly understood. Moreover, in a field like biology – where women are well-represented at the doctoral and postdoctoral levels – it may be easy to assume that issues of gender are unimportant at early career stages. Yet, not all doctoral and postdoctoral positions are equivalent: vast inter-lab differences exist in terms of reputation, mentoring, access to funding and equipment, networking possibilities, and more. Scientists who receive their training in particular labs may be at a disadvantage when applying for grants or faculty positions if their PI is less well-known or if their lab tends to produce fewer "high-impact" publications. We hypothesized that the steep decline in the representation of women at the postdoc-to-PI transition could be in part explained if the most prestigious principal investigators tended to

predominantly train young male scientists. In this study, we therefore sought to explore the link between gender, lab choice, and future academic employment. We found that women are significantly underrepresented, at the graduate and postdoctoral levels, in the laboratories of high-achieving male scientists, while elite female faculty show no such gender bias in their laboratories. We are unable to ascertain to what degree these differences result from self-selection among female trainees and to what degree they result from gender biases among male faculty members. Nonetheless, this skew in lab employment represents a novel – and possibly corrigible – aspect of the leaky pipeline in life science research.

Results

A survey of employment by gender at top-ranked programs in the life sciences

To examine the gender distribution of biomedical scientists in academia, we collected information on the graduate students, postdocs, and faculty employed in 39 departments at 24 of the highest-ranked research institutions in the United States (Table S1 and Supplemental Materials and Methods). We focused on departments that study molecular biology, cell biology, biochemistry, and/or genetics. For each faculty member, we determined their gender, academic rank, and, if available, the year in which their doctorate was received. We then used their lab website or a departmental directory to find the names of graduate students and postdocs currently working in their labs. We attempted to assign a gender to each grad student and postdoc, using the internet and social network searches when a name was ambiguous. Lastly, we used three different criteria to define faculty whose labs we hypothesized would be the most prestigious: those who were funded by the Howard Hughes Medical Institute (HHMI), who were members of the National Academy of Sciences (NAS), or who had won at least one of seven

different major research awards (e.g., the Nobel Prize, the National Medal of Science; see Table S2 and Supplemental Materials and Methods).

Table 1. A survey of biology labs in the United States

	Male	Female	Total
Faculty	1557	505	2062
Professor	1023	276	1299
Associate Professor	269	121	390
Assistant Professor	265	108	373
HHMI Investigators	113	38	151
National Academy members	210	47	257
Major award winners	53	10	63
Faculty with ≥1 trainee listed	982	358	1340
Postdoctoral researchers	3013	1891	4904
Graduate students	2120	2023	4143

In total, we obtained information on 2,062 faculty members in the life sciences (Table 1). Within this sample, 21% of full professors and 29% of assistant professors were women. The gender distribution of the faculty members that we classified as elite were approximately proportional to the gender distribution of full professors, as 25% of HHMI investigators, 18% of National Academy members, and 16% of major award winners were female. Among the trainees that we counted, 49% of 4,143 graduate students and 39% of 4,904 postdocs were women. We note that the percentages of female postdocs and assistant professors in our sample are below national averages (39% vs. 43% and 29% vs. 36%, respectively), suggesting that the leaky pipeline might be comparatively worse at top-ranked research institutions (3,16). Alternately, the subfields that we harvested from might be slightly more male-biased than other fields, like ecology, that also fall under the spectrum of the life sciences (16).

Elite male faculty employ fewer female graduate students and postdocs

We next examined the gender distribution of trainees on a per-lab basis. On average, male principal investigators ran labs that had 36% female postdocs and 47% female grad students (Fig. 1A). These values were significantly lower than we observed in labs headed by women, who employed on average 46% female postdocs and 53% female grad students (Fig. 1B; $P < .0001$ for both comparisons, Wilcoxon rank-sum test). Thus, male professors run labs that have about 22% fewer female postdocs and 11% fewer female grad students than their female colleagues do (Fig. 2).

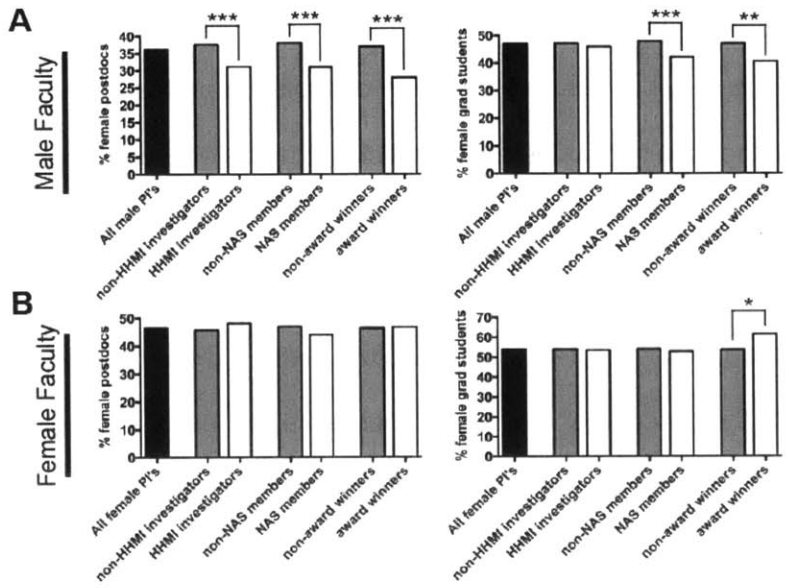


Figure 1. The gender composition of elite biology labs in the United States. The weighted average percentages of female trainees in labs with (A) male PI's and (B) female PI's who have achieved certain career milestones are displayed. Major career awards that were counted for this survey are listed in Table S2. *, P<.05; **, P<.005, ***, P<.0005 (Wilcoxon rank-sum test).

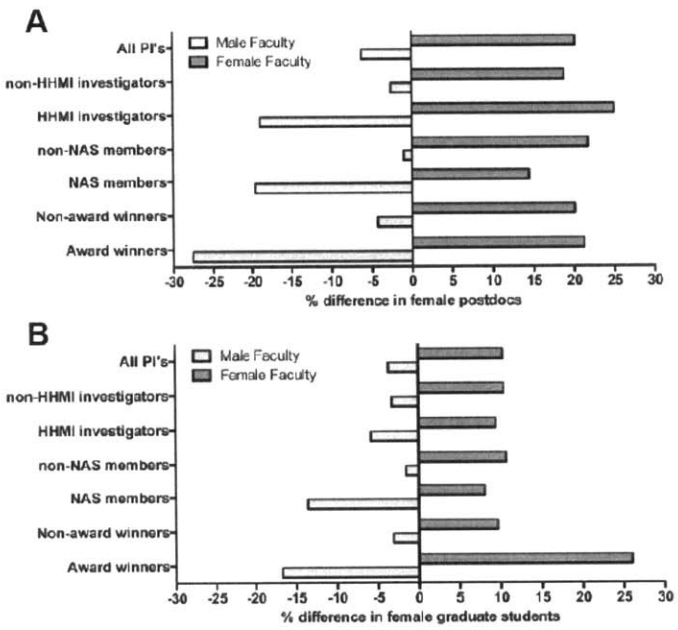


Figure 2. Elite male principal investigators employ fewer women. The percent differences in (A) female postdocs and (B) female graduate students employed by principal investigators who have achieved certain career milestones are displayed. The axis at X=0 represents employing female trainees at a rate proportional to their representation amongst all labs in this survey.

We then tested our data to determine if the most prestigious principal investigators trained fewer women. Surprisingly, for each comparison, male PI's who were funded by HHMI, elected to the National Academy, or who had won a major research award, employed significantly fewer female postdocs than the corresponding pool of other male PI's (Fig. 1A). For instance, male HHMI investigators ran labs that had on average 31% female postdocs, while men who were not HHMI investigators employed on average 38% female postdocs ($P < .0001$, Wilcoxon rank-sum test). This translates to a 19% deficit in the employment of female postdocs, relative to their representation across all labs (Fig. 2A). In contrast, female professors who had achieved the same career milestones showed no evidence of a gender bias. Women who were HHMI investigators ran labs that were 48% female postdocs, compared to 46% among those who were not funded by HHMI (Fig. 1B). Similar results were obtained when we examined women who were members of the National Academy or who had won a major research award.

At the graduate student level, we observed an analogous, though less substantial, skew in employment by gender. Male NAS members and major award winners ran labs with about 41-42% female graduate students, compared to 47-48% among other male professors (Fig. 1A). This represents a 14%-17% deficit in the employment of female graduate students by NAS and award-winning labs, respectively, relative to their representation across all labs (Fig. 2B). However, there was no difference at the graduate student level between HHMI and non-HHMI funded male PI's (Fig. 1A). Among female faculty, major award winners actually trained slightly more female graduate students than non-award winners, while HHMI funding and NAS membership did not affect the number of female grad students employed by female professors (Fig. 1B). Thus, elite male PI's, but not elite female PI's, tend to employ fewer female trainees than other faculty members who have not achieved certain career milestones.

Our dataset also included 24 Nobel Laureates in Medicine/Physiology or Chemistry. Male principal investigators who had won a Nobel Prize (n=22) ran labs that had on average 24% female postdocs and 36% female grad students, which represents a 39% and 27% deficit, respectively, relative to the pool of trainees (Fig. S1). The paucity of female Nobel Laureates prevented a meaningful comparison using this criterion, though we note that both female Laureates in our sample ran labs in which female trainees outnumbered male trainees at the time of our survey.

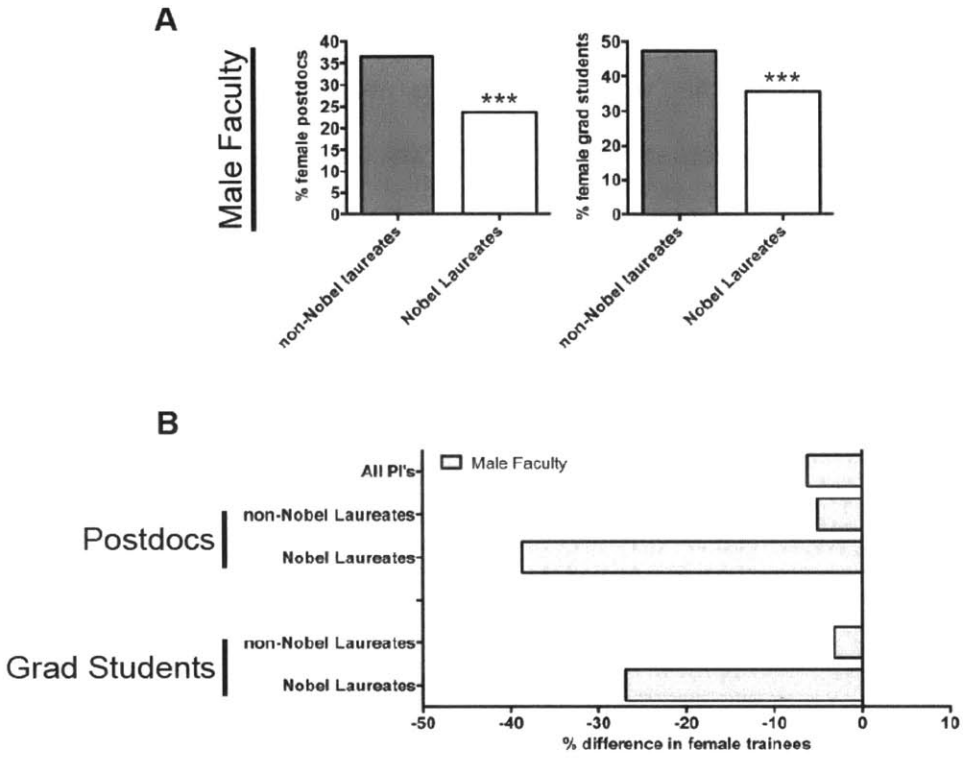


Figure S1. Male Nobel Laureates employ particularly few women. (A) The weighted average percentages of female postdocs and graduate students in labs with male PI's who have or have not won a Nobel Prize are displayed. ***, P<.0005 (Wilcoxon rank-sum test). (B) The percent differences in female postdocs and graduate students trained by principal investigators who have or have not won Nobel Prizes are displayed. The axis at X=0 represents employing female trainees at a rate proportional to their representation amongst all labs in this survey.

These results led us to consider the distribution of trainees across all lab types. We found that female trainees were much less likely to work for an elite principal investigator, particularly at the postdoctoral level (Fig. S2). Combining faculty of both genders, men were about 17% more likely to do their graduate training with a member of the National Academy, 25% more likely to do their postdoctoral training with a member of the National Academy, and 90% more likely to do their postdoctoral training with a Nobel Laureate. Thus, the gender skew in employment results in fewer women being trained in the laboratories of elite investigators.

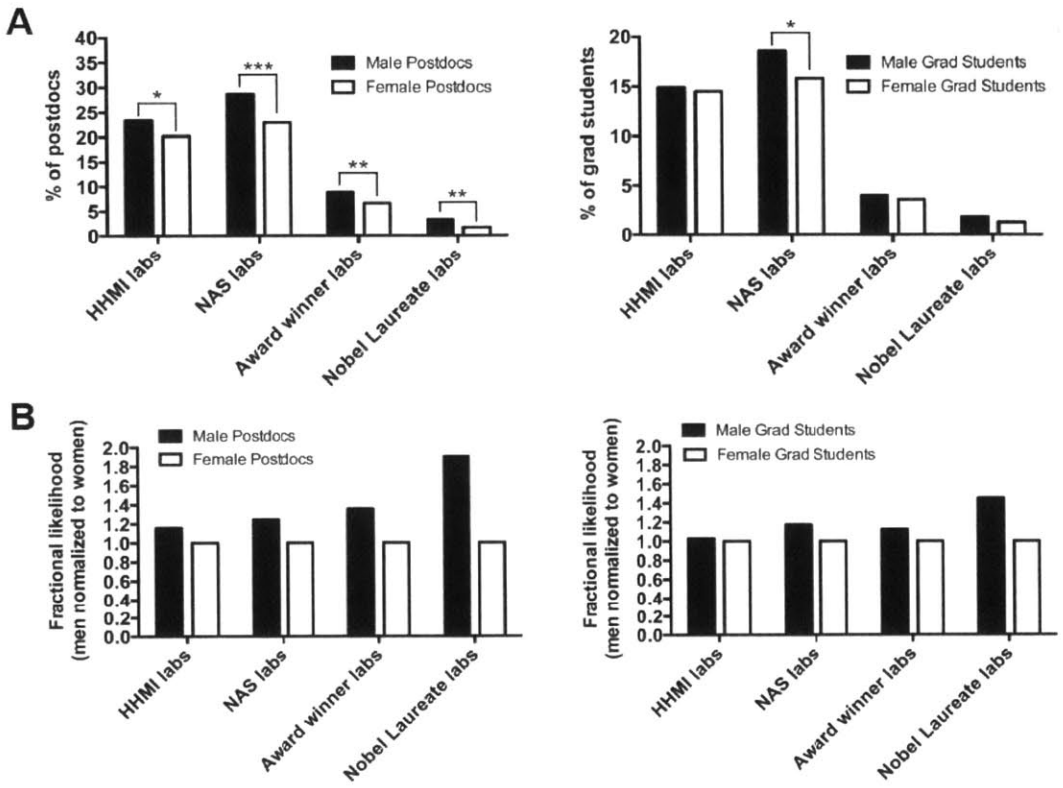


Figure S2. Women are less likely to do their graduate or postdoctoral training with an elite principal investigator. (A) The percent of female postdocs and grad students that are training with a principal investigator who has achieved one of the indicated career milestones are displayed. *, P<.05; **, P<.005, ***, P<.0005 (Fischer exact test). (B) The relative ratios of the percent of men who are training with an elite investigator, normalized to the percent of women who are training with an elite investigator, are displayed.

Other factors that affect the gender skew in biology labs

We next sought to identify other characteristics of principal investigators that correlate with altered gender distributions. Nearly every faculty member who had achieved one of the career milestones that we counted held the rank of professor, thus our results could be explained if older faculty in general trained few women. In fact, men who were full professors tended to employ fewer female postdocs, but more female graduate students, than men who were assistant professors (Fig. S3). Women who were full professors also trained fewer female postdocs than assistant professors did, but there was no difference in graduate student employment between women with different academic ranks. Nonetheless, when we restricted our analysis to only faculty holding the rank of full professor, we still observed a significant deficit in women trained specifically by elite male PI's (Fig. S3).

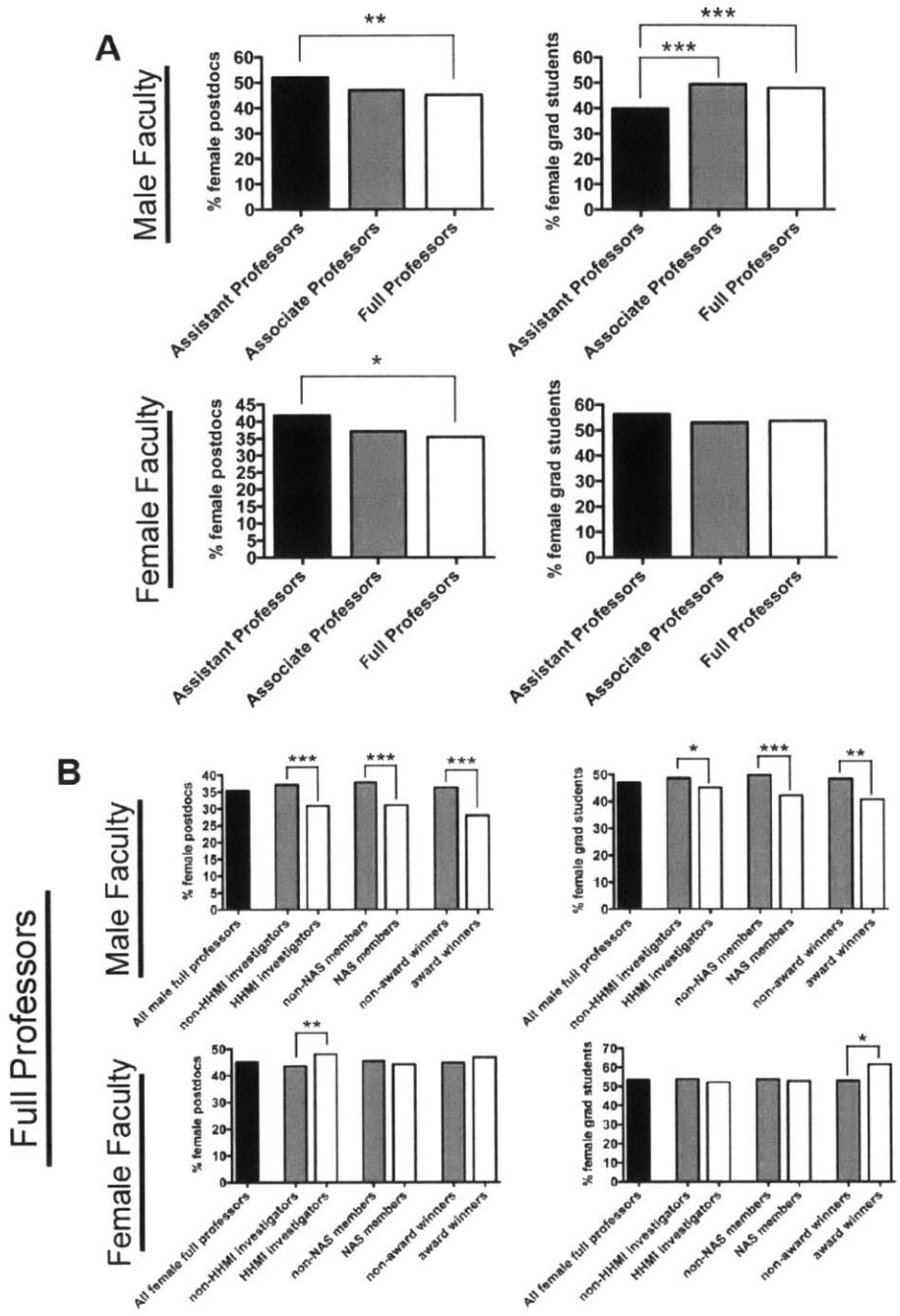


Figure S3. The influence of academic rank on employment by gender. (A) The weighted average percentages of female trainees in labs with male PI's or female PI's with different academic ranks are displayed. (B) The weighted average percentages of female trainees in labs with male PI's or female PI's who have achieved certain career milestones are displayed. *, P<.05; **, P<.005, ***, P<.0005 (Wilcoxon rank-sum test).

Among different STEM disciplines, the representation of women generally decreases in more math-intensive fields (3,17). We focused our analysis on departments of cell and molecular biology (Table S1), but our dataset does contain biophysicists, computational biologists, and other investigators who take a more quantitative approach to biological questions. HHMI investigators whose listed discipline was "biophysics", "computational biology", or "systems biology", and NAS members whose primary section was "Biophysics and Computational Biology" were found to employ particularly few female trainees (Fig. S4). However, even with these faculty members excluded, the remaining male HHMI investigators and NAS members trained fewer women than other male PI's did (Fig. S4).

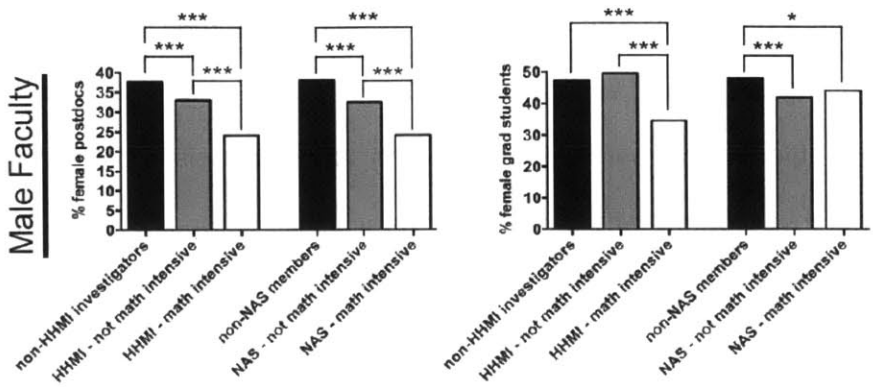


Figure S4. Excluding math-intensive laboratories does not equalize the employment of female trainees. The weighted average percentages of female trainees in labs with male PI's with different research areas are displayed. HHMI investigators whose listed discipline at www.hhmi.org was "biophysics", "computational biology", or "systems biology" were classified as "math intensive" in this figure. NAS members whose primary section in the National Academy was "Biophysics and Computational Biology" were classified as math intensive in this figure. *, P<.05; **, P<.005, ***, P<.0005 (Wilcoxon rank-sum test).

Lastly, we built various linear models from our dataset, using either the weighted percentage of female postdocs or the weighted percentage of female grad students in each lab as the dependent variable. For simplicity, we collapsed the three career achievements that we scored into a single categorical variable ("Elite status"). As expected, among male faculty, elite status was negatively correlated with the percent of female postdocs in a lab ($P < .0001$; Table S3). This remained true even when several other explanatory variables were added, including faculty rank, the years since a faculty member had received his or her Ph.D., and the total number of trainees in a lab (Table S4). As a single independent variable, "years since Ph.D." was moderately negatively correlated with the percent of female postdocs in labs with male faculty members ($P < .045$), but this effect disappeared when other variables were included in the model (Table S3-S4). This suggests that a faculty member's age is not a significant determinant of the gender makeup of their lab, and both young and old elite professors employ few women. Lab size was also negatively correlated with the representation of female postdocs, both as a single variable and in multivariable models. Regression against the percentage of female grad students in each lab revealed similar though less robust results. In multivariable models, elite status was associated with a significantly lower percentage of female grad students trained by male faculty (Table S4). However, "years since Ph.D." was correlated with an increasing representation of female grad students, while lab size was not significantly correlated in either direction. Finally, we constructed equivalent linear models for female principal investigators, but we failed to find a single variable that was significantly associated with differential representation of female trainees in these labs.

Labs that produce assistant professors employ more male postdocs

In the current funding environment, there is intense competition among postdocs for scarce tenure-track positions as assistant professors (18,19). We sought to determine how

postdoctoral lab choice and gender influenced this process. Using CV's, websites, and publication records, we determined the prior employment of 311 out of 373 assistant professors from the 39 departments that we surveyed. Of these, 276 were postdocs prior to their faculty appointments, and 144 of them completed their postdoc in one of 118 labs that we had surveyed.[1] Accordingly, we examined the characteristics of these "feeder" labs that have successfully trained postdocs who won recent faculty job searches at top universities.

The principal investigators of feeder labs were significantly more likely to be Howard Hughes investigators, National Academy members, or to have won a major research award, relative to the pool of all PI's (Fig. 3A). For instance, 13% of the professors in our dataset were members of the National Academy, but 58% of feeder lab professors were NAS members ($P < .0001$, Fisher's exact test). Our above analysis suggested that these feeder labs may therefore have skewed gender ratios. Indeed, these labs had 14% fewer female postdocs than non-feeder labs did (Fig. 3B; $P < .0001$, Wilcoxon rank-sum test). Both male and female faculty, considered separately, who had trained new professors employed fewer female postdocs than those who had not, though for female faculty the percentage was still higher than the representation of female postdocs across all labs (Fig. 3B-3D). As 71% of the assistant professors in our sample were male, these results could represent a form of selection bias. However, when we examined only the feeder labs in which female assistant professors had trained, we found that these labs still employed disproportionately few female postdocs (Fig. 3B). We conclude that principal investigators who successfully train new assistant professors employ an overabundance of male postdocs.

[1] Assistant professors who did not complete postdocs primarily held clinical positions or were independent fellows prior to their faculty appointments.

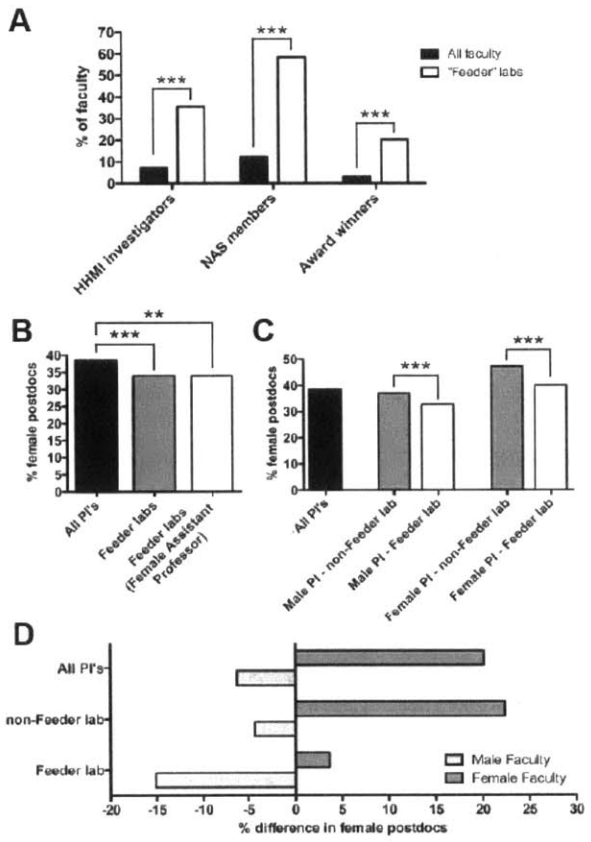


Figure 3. Feeder labs train fewer female postdocs. (A) The percent of all faculty members (black bars) or faculty members who have recently trained new assistant professors (white bars) that have achieved certain career milestones are displayed. ***, P<.0005 (Fischer exact test). (B) The percent of female postdocs employed in different labs are displayed. The black bar represents all principal investigators, the grey bar represents principal investigators who have recently trained at least one new assistant professor of either gender, and the white bar represents principal investigators who have recently trained at least one new female assistant professor. (C) The percent of female postdocs employed in different lab types, subdivided by the gender of the PI, is displayed. (D) The percent differences in female postdocs employed by different principal investigators are displayed. The axis at X=0 represents employing female postdocs at a rate proportional to their representation amongst all labs in this survey.

Discussion

Our results demonstrate that male faculty in general, and elite male faculty in particular, train fewer female grad students and postdocs relative to their representation in the pool of trainees at top universities. These findings are robust to considerations of faculty rank, age, and lab size, and cannot be explained by the exclusion of women from a small number of math-intensive labs. As the majority of assistant professors in our dataset conducted their postdoctoral research under the supervision of one of these high-achieving principal investigators, the limited number of women trained in these labs reduces the number of female candidates who would be most competitive for faculty job searches.

Notably, our current data do not demonstrate conscious bias on the part of male PI's who employ few female graduate students and postdocs. It may be the case that women apply less frequently to labs with elite male PI's. Unfortunately, data on this question are difficult to collect, as applying to do research in a lab is an unregulated and largely informal process. Interested graduate students and postdoctoral candidates typically email PI's that they seek to work with, and faculty members can easily ignore or delete requests when they so choose. Milkman et. al (2013) have demonstrated that faculty members across a host of disciplines, including the life sciences, respond to emails from prospective graduate students written under male names significantly more frequently than they respond to emails written under female names (20). Nonetheless, self-selection among female graduate students and postdocs may still contribute to the gender skew that we have documented. By graduate school, fewer women than men perceive themselves to be on an academic career track (10,21). Some women in particular may not apply to the most prestigious labs if they do not believe it to be important for their professional development. Additionally, in certain circumstances, women underrate their own skill sets (22–24), which could lead some to self-select away from elite labs.

A lack of applications from women could also reflect specific issues with a laboratory or principal investigator. Female Ph.D.'s frequently cite marriage and childbirth as reasons to opt out of scientific careers (12); faculty members who are reputed to be hostile toward maternity considerations could be implicitly discouraging women from applying to their labs. More insidiously, 16% of women employed in the academy report that they have experienced work-related sexual harassment (25), and the tolerance of sexual harassment in a laboratory could further decrease the number of female applicants. Regardless of the cause, we believe that principal investigators who receive few applications from female trainees ought to increase their efforts to proactively recruit talented women, and ought to ensure that their labs are safe spaces for female scientists. A more formalized process of applying to work as a graduate student or postdoc could also serve as a check against principal investigators who routinely fail to hire women.

Alternately, differences in quality between the male and female applicant pools could contribute to the gender gap in elite labs with male PI's. We note, however, that female investigators at the top of their respective fields run labs with just as many women as other female PI's have (Fig. 1B). Additionally, women win competitive fellowships for graduate and postdoctoral training at frequencies that are proportional to their representation among all trainees: about 55% of National Science Foundation graduate fellowships, 45% of Helen Hay Whitney postdoctoral fellowships, and 41% of Jane Coffin Childs postdoctoral fellowships are awarded to women (26–28). These observations argue against a sizeable gap in applicant quality between male and female trainees in the life sciences.

Thus, in addition to the aforementioned factors, we suggest that gender bias may contribute to the decreased employment of women in labs with elite male PI's. Several recent studies have demonstrated that gender bias remains an endemic problem in academic science. For instance, in several European countries, faculty promotion decisions are made by randomly-

chosen review committees. The promotion chances of female candidates are significantly decreased if they are assigned to an all-male review committee, while their promotion chances are equivalent or nearly-equivalent to men's chances if they are assigned to a mixed-gender committee (29,30). Similar results have been reported in the private sector: men tend to underrate women's job performances (31), while having more women in supervisory roles is associated with the increased hiring and promotion of other women (32). In an academic context, grad student and postdoc employment decisions are often made unilaterally, by the single principal investigator who runs a lab. It may be the case that in the most competitive labs, male principal investigators knowingly or unknowingly underestimate the qualifications of female applicants, and instead hire more men in their place. As mixed-gender hiring committees more accurately assess the qualifications of female applicants in a variety of settings (29–31), male PI's could potentially benefit by soliciting feedback on applicants to their labs from female postdocs or affiliated faculty members.

Irrespective of the cause of the gender disparities in elite labs, its consequences significantly shape the academic ecosystem. Our data demonstrate that these labs function as gateways to the professoriate: new generations of faculty members are predominantly drawn from postdocs trained by high-achieving PI's. Yet, these "feeder" labs employ a disproportionate number of men (Fig. 3). According to the theory of cumulative disadvantage, persistent inequalities in achievement can result from small differences in treatment over a prolonged goal-oriented process (14). In controlled studies, women in academia receive less favorable evaluations, lower salary offers, and are ignored by faculty more frequently than men are (15,20). Access to training in certain laboratories may be another level at which women are disadvantaged. The absence or exclusion of female trainees from elite labs deprives them of the resources, visibility, networking opportunities, etc., that could facilitate their professional

development. These differences may contribute to the leaky pipeline by shunting women towards labs that provide fewer opportunities for advancement in academic science.

Conclusions

The continued underrepresentation of women in the academy slows the progress of discovery by artificially excluding individuals with the ability to make significant contributions to the scientific enterprise. It is our hope that this work, along with the growing body of related evidence demonstrating gender bias in the academy (15,20,29,30), will elicit an increased awareness of the ways in which gender continues to play a role in shaping the career trajectories of young scientists. Recognition of gender disparities as a persistent problem can aid in the fair evaluation of women in hiring decisions, and can trigger active steps by individual PI's to recruit more talented women to their labs. Such steps can ensure that, in the future, an individual's gender will not hinder their ability to engage in scientific research.

Materials and Methods

Data Collection. We collected data from 39 biology departments at 24 of the top 25 graduate schools in the biological sciences, as ranked by US News and World Report (33). We did not score faculty from the Scripps Research Institute, as their laboratories were found to use different and sometimes unclear terminology to refer to postdoctoral trainees. Faculty names and ranks were collected from departmental websites. For this study, we counted all assistant professors, associate professors, and full professors. We did not score emeritus professors, adjunct professors, visiting professors, or lecturers.

Each principal investigator's trainees (graduate students and postdocs) were identified using their website or from a public departmental listing. "Graduate students", "pre-doctoral students", "Ph.D. students", and similar terms were scored as referring to grad students. We did not count rotation students, medical students, visiting students, or Master's students, though we did count MD/Ph.D. students who were performing their thesis research in that lab. "Postdoctoral associates", "postdoctoral fellows", "postdoctoral scholars", and similar terms were scored as referring to postdocs. Staff scientists, senior scientists, clinical fellows, and visiting postdocs were not counted. Gender for each individual was scored as male or female based on that individual's photo or name. For ambiguous names and/or photos, we used internet and social network searches in order to determine their gender. In rare cases (~3% of total trainees) we were unable to determine an individual's gender, and they were not counted for our study. The gender of the recipients of graduate and postdoctoral fellowships (NSF, Helen Hay Whitney, and Jane Coffin Childs) was determined in the same manner.

Laboratories were scored between 8/15/2013 and 10/10/2013. Primary data collection was done by both authors as well as by freelancers hired from the service Elance. First, a pool of potential freelancers were recruited and tested on a single department that we had previously counted. Freelancers who completed the testing set without making a single error were hired for the main study. A minimum of three freelancers were instructed to count each department. If all three freelancers agreed on the counts for a principal investigator, then that faculty member was added to the database. If not, the faculty member was re-evaluated by the freelancers or by one of the authors until a consensus count was reached.

We used three distinct criteria to define "elite" principal investigators. First, we identified faculty members whose research programs were supplemented with significant third-party funding from the Howard Hughes Medical Institute. HHMI is the largest private supporter of academic biomedical research in the United States, and investigators chosen by HHMI receive

an annual stipend of on average one million dollars (34,35). HHMI investigators are generally recognized as leaders in their fields, and 23 current or former HHMI investigators have won Nobel Prizes (36). Current Howard Hughes Medical Institute investigators were downloaded from www.hhmi.org on 10/10/2013 and matched with faculty members in our database. HHMI Professors, alumni, and Early Career Scientists were not counted as HHMI investigators.

Secondly, we identified faculty members whose research accomplishments had been recognized by their peers via election to the National Academy of Sciences. The NAS is a non-profit society charged with providing independent and objective advice to the United States government (37). Election to the National Academy is considered "one of the highest honors that a scientist can receive" (38). Members of the National Academy of Sciences were downloaded from www.nasonline.org on 10/11/2013 and matched with faculty in our database.

Lastly, we identified faculty members who had made significant discoveries that had been recognized by a third party with a major career award. We excluded awards that were presented for teaching, service, or that were reserved for young investigators, and instead selected several awards that were intended to honor scientific achievements (Table S2). The winners of these major research awards (through 2013) were downloaded from their respective websites and matched with faculty in our database.

Data Analysis. Data was analyzed using Python and Excel. The reported percentages of female graduate students and postdocs displayed in each figure are weighted by the number of graduate students and postdocs, respectively, in a given lab. Significance tests were also performed between weighted averages. Linear regression models were constructed in Python with the statsmodels and pandas packages using weighted least squares. Categorical factors

were handled using dummy coding, replacing a single categorical variable of K categories with K-1 dummy variables, each with two possible values.

Acknowledgments

We thank Angelika Amon, Terry Orr-Weaver, Frank Solomon, Kendra Albert, and members of the Amon lab for helpful comments on the manuscript.

References

1. National Science Board. Science and Engineering Indicators 2012. Available from: http://www.nsf.gov/statistics/seind12/

2. National Science Foundation, Division of Science Resources Statistics. Thirty-Three Years of Women in S&E Faculty Positions. 2008. Available from: www.nsf.gov/statistics/infbrief/nsf08308/

3. Donna J. Nelson. A National Analysis of Minorities in Science and Engineering Faculties at Research Universities. 2007. Available from: http://faculty-staff.ou.edu/N/Donna.J.Nelson-1/diversity/Faculty_Tables_FY07/FinalReport07.html

4. Pell AN. Fixing the leaky pipeline: women scientists in academia. J Anim Sci. 1996;74:2843–8.

5. Sadler PM, Sonnert G, Hazari Z, Tai R. Stability and volatility of STEM career interest in high school: A gender study. Sci Educ. 2012;96:411–27.

6. Morgan SL, Gelbgiser D, Weeden KA. Feeding the pipeline: Gender, occupational plans, and college major selection. Soc Sci Res. 2013;42:989–1005.

7. Ceci SJ, Williams WM. Understanding current causes of women's underrepresentation in science. Proc Natl Acad Sci. 2011;201014871.

8. Martinez ED, Botos J, Dohoney KM, Geiman TM, Kolla SS, Olivera A, et al. Falling off the academic bandwagon. Women are more likely to quit at the postdoc to principal investigator transition. EMBO Rep. 2007;8:977–81.

9. Ferriman K, Lubinski D, Benbow CP. Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging adulthood and parenthood. J Pers Soc Psychol. 2009;97:517–32.

10. Anders SM van. Why the Academic Pipeline Leaks: Fewer Men than Women Perceive Barriers to Becoming Professors. Sex Roles. 2004;51:511–21.

11. Postdoctoral Life at MIT: Findings from the 2010 Postdoctoral Scholar Survey. Available from: http://web.mit.edu/ir/surveys/pdf/Postdoctoral_Life_at_MIT_Report_June_2011.pdf

12. Goulden M, Mason MA, Frasch K. Keeping Women in the Science Pipeline. Ann Am Acad Pol Soc Sci. 2011;638:141–62.

13. Ginther DK, Kahn S. Does Science Promote Women? Evidence from Academia 1973-2001. National Bureau of Economic Research; 2006 Nov. Report No.: 12691. Available from: http://www.nber.org/papers/w12691

14. DiPrete TA, Eirich GM. Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments. Annu Rev Sociol. 2006;32:271–97.

15. Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. Science faculty's subtle gender biases favor male students. Proc Natl Acad Sci. 2012;201211286.

16. National Science Foundation, National Center for Science and Engineering Statistics. Women, Minorities, and Persons with Disabilities in Science and Engineering: 2013. Available from: http://www.nsf.gov/statistics/wmpd/

17. Ceci SJ, Williams WM. Sex Differences in Math-Intensive Fields. Curr Dir Psychol Sci. 2010;19:275–9.

18. Carpenter S. Tenure-Track Jobs Remain Scarce. Science. 2010; Available from: http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2010_01_15/caredit.a1000006

19. Price M. Excessive Supply, Uncertain Demand. Science. 2013; Available from: http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2013_09_16/caredit.a1300199

20. Milkman KL, Akinola M, Chugh D. Discrimination in the Academy: A Field Experiment. Rochester, NY: Social Science Research Network; 2013 Oct. Report No.: ID 2063742. Available from: http://papers.ssrn.com/abstract=2063742

21. Fox MF, Stephan PE. Careers of Young Scientists: Preferences, Prospects and Realities by Gender and Field. Soc Stud Sci. 2001;31:109–22.

22. Steinmayr R, Spinath B. What Explains Boys' Stronger Confidence in their Intelligence? Sex Roles. 2009;61:736–49.

23. Correll SJ. Gender and the Career Choice Process: The Role of Biased Self-Assessments. Am J Sociol. 2001;106:1691–730.

24. Pallier G. Gender Differences in the Self-Assessment of Accuracy on Cognitive Tasks. Sex Roles. 2003;48:265–76.

25. Ilies R, Hauserman N, Schwochau S, Stibal J. Reported Incidence Rates of Work-Related Sexual Harassment in the United States: Using Meta-Analysis to Explain Reported Rate Disparities. Pers Psychol. 2003;56:607–31.

26. Helen Hay Whitney Foundation: Current Fellowship Recipients. Available from: http://www.hhwf.org/HTMLSrc/Directory.html

27. 2013 Graduate Research Fellowships Reflect a Diversity of Fields, Institutions and Students. Available from: http://www.nsf.gov/news/news_summ.jsp?cntn_id=127538

28. THE JANE COFFIN CHILDS FUND FELLOWS 2011 – 2014. Available from: http://www.jccfund.org/about-fund/fellows

29. Zinovyeva N, Bagues M. Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment. Rochester, NY: Social Science Research Network; 2010 May. Report No.: ID 1618256. Available from: http://papers.ssrn.com/abstract=1618256

30. De Paola M, Scoppa V. Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy. 2011 Jun. Report No.: 201106. Available from: http://econpapers.repec.org/paper/clbwpaper/201106.htm

31. Bowen C-C, Swim JK, Jacobs RR. Evaluating Gender Biases on Actual Job Performance of Real People: A Meta-Analysis1. J Appl Soc Psychol. 2000;30:2194–215.

32. Kurtulus FA, Tomaskovic-Devey D. Do Female Top Managers Help Women to Advance? A Panel Study Using EEO-1 Records. Ann Am Acad Pol Soc Sci. 2012;639:173–97.

33. Best Biological Science Programs. U.S. News & World Report; Available from: http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-science-schools/biological-sciences-rankings

34. Financials | Howard Hughes Medical Institute (HHMI). Available from: http://www.hhmi.org/about/financials

35. Brower V. Room of its own. EMBO Rep. 2004;5:851–3.

36. Nobel Laureates | Howard Hughes Medical Institute (HHMI). Available from: http://www.hhmi.org/scientists/nobel-laureates

37. Overview: NAS Mission. Available from: http://www.nasonline.org/about-nas/mission/

38. Overview: Membership. Available from: http://www.nasonline.org/about-nas/membership/