A Transcriptional and Metabolic Signature of Primary Aneuploidy Is Present in Chromosomally Unstable Cancer Cells and Informs Clinical Prognosis

Jason M. Sheltzer

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 and aneuploid primary cells also share an increase in glycolytic and TCA cycle flux. 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 is clonal aneuploidy, which is selected during tumor evolution and associated with robust growth, and the other 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. Cancer Res; 73(21); 6401–12. ©2013 AACR.

Introduction

Aneuploidy and chromosomal instability (CIN) are interrelated 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. CIN 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 such as chromosome nondisjunction, micronuclei formation, and cytokinesis failure occur sporadically in normal cells.

Aneuploidy and CIN 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 CIN, CIN mice are typically tumor prone (14–16), but in some contexts CIN seems 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 CIN, including FISH, flow cytometry, and the direct analysis of mitotic figures, have been used 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 CIN on cancer transcriptomes are unknown. Here, I demonstrate that chromosomally unstable cancer cell lines are transcriptionally and metabolically similar to aneuploid primary cells. Although 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 CIN 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.

Materials and Methods

Data sources
Gene expression data from trisomic mouse embryonic fibroblasts (MEF) were acquired from ref. 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 ref. 26, and missing values were kindly provided by A. Rosckhe (NCI, Bethesda, MD). Doubling times of the NCI60 panel were acquired from ref. 27. Gene expression values from cancer/normal tissue pairs were acquired from the Gene Expression Omnibus (GEO; GSE15641, GSE27943, GSE3167, GSE3189, and GSE5364). Gene expression values from breast tumors of known ploidy were acquired from GEO (GSE12071). Consumption/release rates of metabolites in the NCI60 panel were acquired from ref. 28. Gene expression values from tetrasomic HCT116 cells were acquired from ref. 29. Gene expression values from normal human tissues were acquired from GEO (GSE1133). Gene expression values from clinical cancer specimens were acquired from the sources cited in Supplementary Tables S18 and S20.

Mouse husbandry and MEF culture
All animal studies and procedures were approved by the MIT Institutional Animal Care and Use Committee. Cdc20AAA mice were kindly provided by Dr. Pumin Zhang (Baylor College of Medicine, Houston, Texas). BubR1 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 Cdc20AAA crosses and E13.5 for BubR1 crosses. MEFs were isolated and cultured as described in ref. 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 were summarized and normalized using gcRNA and have 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, nonspecific probesets (those annotated with an “x”) were excluded from analysis. The exception to this is the CIN70 gene signature in Supplementary Fig. S4, in which some of the genes are present on sex chromosomes or are only measured by nonspecific probesets.

Gene Ontology (GO) term enrichment analysis was done using GProfiler with a Benjamini–Hochberg-corrected P value of 0.05 and a maximum P value of $10^{-2}$ (30). Enrichments were performed against the relevant background gene set, that is, 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 used for analysis. Gene Set Enrichment Analysis (GSEA) was performed using a Pearson ranking metric, a gene set size range of 5 to 5,000, and 1,000 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 more or −0.3 or less was uniformly applied. For two-sample comparisons, a combined Student t test ($P < 0.05$) 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-C cells, genes expressed $±1$ SD 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 CIN, I used a strategy similar to that described in ref. 36. Levels of CIN were determined using the index of numerical heterogeneity (INH) of each cell line, as reported in ref. 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 seemed 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 artifacts.) Fractional INH 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
transcriptional consequences of aneuploidy and CIN

chromosomes were grouped together for this analysis. Finally, 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 0.05 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. TRIT70 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 more or less than the mean, respectively, for each gene set. P values and HRs were determined using the survival package in R or in Prism. Cox proportional hazard models were constructed using the coxph function in R.

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 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 (ρ > 0.3 or ρ < −0.3), then identified GO terms enriched among these gene sets. Among upregulated 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 (Supplementary Table S1). Downregulated genes were enriched for those annotated to nucleic acid metabolism (P < 10−13), DNA repair (P < 10−12), and RNA processing (P < 10−9). These GO terms are consistent with our hypothesis that primary aneuploidy is an antiproliferative 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 cancers. Importantly, several karyotypic patterns have been determined for each of these lines, including their INH (26). The INH is a measure of cell-to-cell variability in chromosome content: cancers that exhibit CIN have large INH values, whereas 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 SCCs 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 B: transcript levels of the DNA replication factor MCM6 tend to decrease in cell lines with high CIN, whereas transcript levels of the ER-associated gene CALU are correlated with increasing CIN. Surprisingly, genes correlated or anticorrelated with CIN were enriched for many of the same functional categories as observed in aneuploid primary cells (Fig. 1C and Supplementary Table S2). Nucleic acid metabolism (P < 10−36) and the chromosome (P < 10−21) were highly enriched among genes that decreased with increasing CIN, whereas the membrane (P < 10−5) and the extracellular matrix (P < 10−6) were positively correlated with CIN. In addition, 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 anticorrelated with primary aneuploidy, all 35 were also anticorrelated with CIN.

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 GSEA, a threshold-independent method for correlating phenotypes with gene sets (31). GSEA revealed similar GO term enrichments as described earlier, particularly among downregulated genes. DNA replication, RNA splicing, and the cell cycle were significantly anticorrelated with both increasing aneuploidy and increasing CIN (Supplementary Fig. S1 and Supplementary Tables S3 and 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 biologic functions.

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 E). For instance, 19% of all genes were anticorrelated with aneuploidy in MEFs at a p < −0.3 cutoff. Among orthologous genes anticorrelated with CIN in cancer cells, a significantly greater percentage (36%) were also anticorrelated with aneuploidy in MEFs (P < 10−16, hypergeometric test). In contrast, among genes that were positively correlated with CIN, significantly fewer (10%) were anticorrelated 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 (p = 0.14, P < 10−26). When the analysis was restricted to only those genes affected by CIN in cancer cells (p < −0.3 and p > 0.3), the strength
of the correlation increased to $r = 0.28$ ($P < 10^{-13}$). 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 compared ranks between the mean fold change expression levels in the tumors and the SCC values from trisomic MEFs (Supplementary Fig. S2). Interestingly, the average expression levels from tumors were anticorrelated with the aneuploidy responsiveness of genes in MEFs (average $r = -0.16, P < 10^{-25}$; Supplementary Fig. S2A). I then identified differentially expressed genes between tumors and normal tissue using a combined Student $t$ test ($P < 0.05$) and mean fold change (±1.3) threshold. Transcripts upregulated in aneuploid primary cells were significantly more likely to be downregulated in tumors, whereas transcripts downregulated in aneuploid primary cells were likely to be upregulated in tumors (Supplementary Fig. S2B and S2C). Few GO terms were significantly enriched among the small number of genes that were similarly affected by aneuploidy and cancer (Supplementary Fig. S2D and S2E and Supplementary Table S5). Conversely, GO terms related to cell-cycle progression were strongly enriched among genes downregulated by aneuploidy but upregulated in tumors (Supplementary Fig. S2E and Supplementary Table S6).

To address the possibility that these results were due to comparing tumors with 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 anticorrelated with the expression levels from trisomic MEFs ($r = -0.12, P < 10^{-13}$; Supplementary Fig. S2F). In addition, the GO categories perturbed in aneuploid tumors were dissimilar to those affected by aneuploidy in primary cells (Supplementary 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.

**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 downregulation of cell-cycle genes and the upregulation of stress- and signaling-related genes. Consistent with this hypothesis, I found that the INH 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 ($r = 0.40, P < 0.003$). This effect remained true when cell lines derived from hematopoietic cancers were excluded ($r = 0.36, P < 0.02$). 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 ($P < 10^{-15}$; Supplementary 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 anticorrelated 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 Supplementary 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 anticorrelated with CIN than with proliferation ($|\rho_{\text{CIN}} - \rho_{\text{DT}}| > 0.3$ or $<-0.3$). Genes annotated to the mitochondrion ($P < 10^{-15}$), the mitochondrial matrix ($P < 10^{-15}$), and cellular metabolic processes ($P < 10^{-15}$) displayed a stronger correlation with CIN than with doubling time (Fig. 2C; Supplementary 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.

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

Aneuploid MEFs display various metabolic alterations: in general, trisomic cells use more glutamine, and produce more...
lactate, glutamate, and ammonium than euploid cells do. In addition, some aneuploid cells use 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 use 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, whereas 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 and malate; \( P < 10^{-5} \)) and glycolysis (e.g., pyruvate and 2-phosphoglycerate; \( P < 10^{-3} \); Supplementary 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 (Supplementary 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).

The CIN transcriptional signature is apparent in spindle checkpoint–defective MEFs and after chromosome transfer into karyotypically stable cancer cell lines

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

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 increases with doubling time (white bars) tend to show similar but smaller trends than when those metabolites are correlated with karyotype heterogeneity (black bars; \( P < 0.01 \) for both produced and consumed metabolites, paired t test).
signature. I examined the transcriptomes in chromosomally unstable primary cells and in chromosomally stable cancer cell lines in which aneuploidy was ectopically induced.

CIN 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<sup>AAA/AAA</sup> or BubR1<sup>11/11</sup>) as well as from their wild-type littermates (Cdc20<sup>+/-</sup> and BubR1<sup>+/-</sup>). 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 (ρ = 0.15, P < 10<sup>-20</sup>, and ρ = 0.26, P < 10<sup>-38</sup>, respectively). Genes upregulated in CIN MEFs tended to be upregulated in high-CIN cancer cells and in trisomic MEFs, and vice versa (Supplementary Fig. S3A–S3D). GO term enrichment analysis revealed the upregulation of extracellular and cell motility genes, and the downregulation of cell-cycle transcripts, as is also observed in the trisomic MEFs and NCI60 panel (Supplementary Table S13 and Supplementary Fig. S3E). Similar results were obtained when either Cdc20<sup>AAA/AAA</sup> or BubR1<sup>11/11</sup> MEFs were analyzed independently (data not shown). I conclude that the transcriptional patterns present in primary cells that exhibit CIN are highly similar to both primary aneuploid cells and cancer cells that are chromosomally unstable.

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 with the SCC-CIN values from the NCI60 panel and with 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 (ρ = 0.26, P < 10<sup>-47</sup>, and ρ = 0.22, P < 10<sup>-43</sup>, respectively; Supplementary Fig. S3F–S3I). The induction of aneuploidy led to the downregulation of cell-cycle transcripts and the upregulation of stress and membrane-related transcripts (Supplementary Table S14 and Supplementary Fig. S3E). Thus, inducing aneuploidy in an otherwise stable cancer cell line mimics the CIN transcriptional signature. Among 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 (Supplementary 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 anticorrelated with karyotype heterogeneity and doubling time

Carter and colleagues (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 CIN; 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 anticorrelated with karyotype heterogeneity in the NCI60 panel (ρ = –0.35, P < 0.01; Supplementary Fig. S4A). Chromosomally stable cancer cell lines (HCT116 and CCRF-CEM) generally had higher CIN70 scores than unstable cell lines (HOP-92 and TK-10). Upon inspection of the CIN70 gene list, I noted that many transcripts were well-characterized markers of proliferation (e.g., PCNA, CDK1, and 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 (ρ = –0.46, P < 0.001; Supplementary Fig. S4B). To further test the link between CIN70 and proliferation rates, I examined a microarray dataset of 81 nondiseased human tissues and cell populations. As a proxy for their proliferative index, I sorted the tissues by their level of proliferating cell nuclear antigen (PCNA) expression. While 30% of all genes were positively correlated with PCNA in this dataset, I found that 100% of CIN-related aneuploidy
CIN70 genes were positively correlated with PCNA (average \(p = 0.45, P < 10^{-5}\); Supplementary Fig. S4C). Similar correlations were observed between CIN70 and other common proliferative markers (e.g., \(\rho\)MK670 = 0.57, \(P < 10^{-13}\), \(\rho\)TOPLA0 = 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; ref. 42) were still highly correlated with PCNA (average \(p = 0.40, P < 10^{-9}\)), 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 CIN in the NCI60 panel, but instead captures information about cell proliferation rates in both the NCI60 lines and chromosomally stable normal tissue.

Construction of gene signatures associated with aneuploidy and CIN

As aneuploidy and CIN 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 nonspecific marker of cell proliferation (Supplementary Table S15; ref. 46). Second, I defined TRI70, which consists of the 70 genes that displayed the strongest absolute correlation with aneuploidy in trisomic MEFs (Supplementary Table S16). Third, I defined HET70, which consists of the 70 genes that displayed the strongest correlation with karyotype heterogeneity in the NCI60 panel (Supplementary 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 whereas 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 CIN, these gene signatures reflect cell proliferation rates (\(p = 0.61, P < 10^{-15}\), and \(p = -0.29, P < 10^{-3}\), respectively; Supplementary Fig. S5A and S5B). Interestingly, the HET70 score was neither correlated nor anticorrelated with PCNA in human tissue (\(p = 0.11, P > 0.05\); Supplementary Fig. S5C and S5D). Although 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.

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; Supplementary Table S18 and Supplementary Fig. S6). I found that high PCNA25 was significantly associated with poor survival in 15 of 27 cohorts. PCNA25 was particularly informative in breast cancers (significant in six of six cohorts), brain cancers (significant in four of four cohorts), and bladder cancers (significant in two of three cohorts). PCNA25 did not provide significant patient stratification in lung cancer (one of five cohorts), ovarian cancer (zero of three cohorts), or colorectal cancer (zero of three 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.

Subsequently, I found that CIN70 provided significant stratification of patient risk in 15 of 27 cohorts (Supplementary 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 HRs generated by CIN70 and PCNA25 were highly correlated (\(p = 0.98, P < 10^{-18}\); Supplementary Fig. S7A). Despite having only 12 genes in common, the individual patient scores generated by CIN70 and PCNA25 were nearly identical (\(p = 0.97–0.99\); Supplementary Fig. S7). 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.

TRI70 was found to be a significant predictor of survival in 12 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 HRs calculated using the two gene signatures were highly correlated (\(p = -0.90, P < 10^{-18}\); Supplementary Fig. S7B). 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 of 27 patient cohorts. Surprisingly, five of these cohorts were not stratified by PCNA25: HET70 was able to
stratify colorectal cancer (three of three patient cohorts), which PCNA25 failed to do. Across all datasets, the HRs generated by HET70 and PCNA25 were uncorrelated (Supplementary Fig. S7C), and individual patient scores generated by HET70 and PCNA25 were poorly correlated or uncorrelated (Supplementary Fig. S7D–S7F). I conclude that the gene expression changes associated with CIN 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 CIN of their tumors, I hypothesize that CIN 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

To narrow the HET70 gene signature into a smaller set of genes that 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 two of three bladder cancer cohorts, four of four brain cancer cohorts, three of three colorectal cancer cohorts, one of one liposarcoma cohort, one of one liver cancer cohort, three of five lung cancer cohorts, and one of three ovarian cancer cohorts (Supplementary Table S19). I then assembled a second set of published clinical expression arrays that were not used in the initial derivation of the HET5 signature. Within this test set, HET5 provided significant stratification in eight of 15 cohorts, including one of one bladder cancer cohort, two of two brain cancer cohorts, and two of four lung cancer cohorts (Supplementary Fig. S8 and Supplementary Table S20). Remarkably, HET5 was also able to stratify patient risk in three of four 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.

Discussion

Many cancers display CIN, 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 (Figs. 1 and 2; ref. 22). CIN cancer cell lines divide more slowly than chromosomally stable lines, downregulate transcripts associated with proliferation-independent markers for poor prognosis in vivo. Though I cannot ascertain at this time whether HET70 scores stratify patients according to the CIN of their tumors, I hypothesize that CIN is one mechanism by which the increased expression of these genes can arise.
correlated with markers of rapid proliferation (Supplementary Fig. S5). To reconcile these results, I hypothesize that there are two “types” of aneuploidies 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, whereas 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 CIN: 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 (Supplementary 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 increase to clonal levels. Although 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 CIN (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. CIN 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 physiologic 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). Although 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 that 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 that 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.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
The author thanks Joan Clara Smith for coding assistance, Pumin Zhang and Van de Wiersen for providing mice, and Erica Chung for help with the culture. The author also thanks Angelika Amon, Maitreyi Dunham, Mike Lauth, Aviv Regev, Yitzhak Finkel, and members of the Amon Laboratory for comments on the article.

Grant Support
This work was supported by NIH grant GM056060 (Angelika Amon). Angelika Amon is an investigator of the Howard Hughes Medical Institute, whose funding also supported this research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked
Transcriptional Consequences of Aneuploidy and CIN

References


27. NCBI Discovery Services [Internet]. Available from: http://dtp.nci.nih.gov/docs/misc/common_files/cell_list.html.


www.aacajournals.org Cancer Res; 73(21) November 1, 2013 6411
A Transcriptional and Metabolic Signature of Primary Aneuploidy Is Present in Chromosomally Unstable Cancer Cells and Informs Clinical Prognosis

Jason M. Sheltzer


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-13-0749

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2013/09/17/0008-5472.CAN-13-0749.DC1

Cited articles
This article cites 58 articles, 32 of which you can access for free at:
http://cancerres.aacrjournals.org/content/73/21/6401.full.html#ref-list-1

Citing articles
This article has been cited by 8 HighWire-hosted articles. Access the articles at:
/content/73/21/6401.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.