Corrections for mRNA Extraction and Sample Normalization Errors Find Increased mRNA Levels may Compensate for Cancer Haplo-Insufficiency

David A. Weaver,[1] Andrea L. Nestor-Kalinoski,[2] Kristen Craig,[2] Matthew Gorris,[2] Tejal Parikh,[2] Helen Mabry,[2] and David C. Allison[2]*

[1]Program in Bioinformatics and Proteomics/Genomics, The University of Toledo, College of Medicine and Life Sciences, Toledo, OH
[2]Department of Surgery and The University of Toledo Advanced Microscopy and Imaging Center, The University of Toledo, College of Medicine and Life Sciences, Toledo, OH

The relative mRNA levels of differentially expressed (DE) and housekeeping (HK) genes of six aneuploid cancer lines with large-scale genomic changes identified by SNP/SKY analysis were compared with similar genes in diploid cells. The aneuploid cancer lines had heterogeneous genomic landscapes with subdiploid, diploid, and supradiploid regions and higher overall gene copy numbers compared with diploid cells. The mRNA levels of the haploid, diploid, and triploid HK genes were found to be higher after correction of easily identifiable mRNA measurement errors. Surprisingly, diploid and aneuploid HK gene mRNA levels were the same by standard expression array analyses, despite the higher copy numbers of the cancer cell HK genes. This paradoxical result proved to be due to inaccurate inputs of true intra-cellular mRNAs for analysis. These errors were corrected by analyzing the expression intensities of DE and HK genes in mRNAs extracted from equal cell numbers (50:50) of intact cancer cell and lymphocyte mixtures. Correction for both mRNA extraction/sample normalization errors and total gene copy numbers found the SUIT-2 and PC-3 cell lines' cancer genes both had ~50% higher mRNA levels per single allele than lymphocyte gene alleles. These increased mRNA levels for single transcribed cancer alleles may restore functional mRNA levels to cancer genes rendered haplo-insufficient by the genetic instability of cancer.   © 2013 Wiley Periodicals, Inc.

INTRODUCTION

Chromosomal instability leads to uneven gene copy number variations (CNVs) across cancer cell genomes (Lengauer et al., 1998; Galitski et al., 1999; Singer et al., 2000; Albertson et al., 2003; Beroukhim et al., 2010; Bazeley et al., 2011; Carter et al., 2012). The effects of cancer CNVs on mRNA levels are poorly understood (Hyman et al., 2002; Platzer et al., 2002; Pollack et al., 2002; Tsafrir et al., 2006; Jiang et al., 2008; Marella et al., 2009). We recently suggested that some aneuploid CNVs may be selected to balance the effects of mutations, epigenetic silencing, and other gene losses acquired during the continuous division of chromosomally unstable cancer cells (Bazeley et al., 2011). We now show evidence for increased mRNA levels occurring in aneuploid cancer cells compared with diploid lymphocytes after correction for gene copy number differences between these cell types.

An analysis of expression array mRNA measurement errors defined the assay signal strengths needed for accurate mRNA measurements. Housekeeping (HK) genes were initially studied to minimize the effects of differentiation signals and because these genes generally produce high enough mRNA levels for accurate measurement (Chudin et al., 2001; Workman et al., 2002; Su et al., 2004; Barnes et al., 2005). Gene copy numbers were estimated by combined spectral karyotyping (SKY) and single nucleotide polymorphism (SNP) haplotyping (Greshock et al., 2007; Nestor et al., 2007) of six aneuploid cancer lines.

We found that gene expression intensity determinants (GEIDs) introduced noise into mRNA measurements, and the correction of these and other measurement errors led to finding gene

dosage effects for haploid, diploid, and triploid HK genes. This argued against HK gene mRNA levels being under feedback mechanisms causing genes with differing copy numbers to have similar mRNA levels. Surprisingly, even though the six aneuploid cancer lines had more HK gene copies than the four diploid cell types, standard expression array analyses found similar diploid and aneuploid HK gene mRNA levels.

However, a new method for correcting mRNA extraction variability and pre-measurement, sample normalization errors (Loven et al., 2012) found the mRNA levels of genes expressed in the aneuploid PC-3 and SUIT-2 cancer lines were actually higher than similar genes expressed by diploid lymphocytes. The increased mRNA levels found for the actively transcribed cancer genes, either due to elevated MYC levels (Lin et al., 2012; Nie et al., 2012) or other mechanisms, may restore functional mRNA levels to haplo-insufficient genes created by the genetic instability of cancer.

MATERIALS AND METHODS

Cell Culture and Gene Expression Measurements

Primary lymphocyte cultures, the diploid, benign BUD-8 (skin), CCD-34Lu (lung), and MRC-5(lung) cell lines, and the transformed aneuploid A549 (lung), DU 145 (prostate), LN-18 (glioblastoma), PC-3 (prostate), RWPE-2 (prostate), and SUIT-2 (pancreatic) lines were grown as previously described (Allison and Nestor, 1999; Bazeley et al., 2011). Cell counts were made with a TC-20 Automated Cell Counter (Bio-Rad, Hercules, CA).

Cells were lysed and RNA extracted with TRIzol (Invitrogen Corporation, Carlsbad, CA) followed by RNA cleanup with Qiagen RNeasy kit (Quiagen Inc., Valencia, CA). Only RNA with a purity ratio (A_{260}/A_{280}) of 1.9 to 2.1 in TE Buffer (10 mM Tris-HCl buffer, 0.5M EDTA, pH 8.0) was used. RNA samples were reverse transcribed into cDNA, synthesized to cRNA in biotinylated tri-phosphates, and labeled cRNA from each sample hybridized onto Affymetrix U133 Plus 2.0 gene chips by Affymetrix protocols (http://www.affymetrix.com, Affymetrix, Santa Clara, CA). Probe sets with present ($P < 0.04$) or marginal ($0.04 < P < 0.06$) expressions (Affymetrix MAS 5.0 algorithm) were classified as + expressions and probe sets with absent expressions ($0.06 < P < 1.0$) were classified as − expressions.

Expression Analysis Settings

Affymetrix image files were created by laser scanning of the R-Phycoerthrin Streptavidin (SAPE) fluorphore incorporated during the sample cartridge washing and staining stages. A .DAT file is first created with pixilated features which are automatically converted by an algorithm to create a .CEL file in which each individual feature is assigned a single intensity value. Once created, the .CEL image file can be evaluated by the Affymetrix MAS 5.0 software or imported into a number of third party softwares for various types of normalization and analysis.

For these studies we used All Probe Set Scaling in the Affymetrix software to assess the array data for compatibility among the various experiments. Scaling is a mathematical technique applied to the data from several different probe arrays to minimize discrepancies due to variables such as pipeting errors, varying hybridization conditions, staining effectiveness, probe array lots, and etc. The All Probe Set Scaling adjusts the trimmed mean signal of a probe array to a user-specified target signal value. Target values normally vary between 100 and 500 and in our experiments were set to 150 to keep the Scaling Factor close to 1. The Scaling Factor's for all arrays in the same project should be within a three-fold range for comparability and this applied to all comparisons in our studies. All other expression settings were default software settings.

Karyotypes, HK Genes, and SNP/SKY Analyses

Diploid lymphocyte chromosomes were confirmed by FISH karyotyping (Allison and Nestor, 1999) and the SKY karyotypes of A549, DU 145, LN-18, PC-3, and SUIT-2 have been published (Nestor et al., 2007; Bazeley et al., 2011). Complete SKY analyses of ≥24 karyotypes was performed, although only 13 normal BUD-8 karyotypes were successfully examined in this previously reported diploid line (Kushnaryov et al., 1982). HK genes were identified by functional analyses and expression in >99% of 2,263 cell types (Bazeley et al., 2011), followed by selection of 520 HK genes mapping to a single locus on chromosomes 1 to 22. SKY was performed at the Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY, and SNP haplotyping at the Gene Expression and Genotyping Facility at the Comprehensive Cancer Center of Case Western Reserve University.

Generation of Random Relative Expression Intensity (REI) Distributions to Detect Gene Expression Intensity Determinants (GEIDs)

The expression intensities (EIs) of the ten cell types were not normally distributed and therefore were analyzed by nonparametric statistical analysis. A random distribution in which all genes have a 50/50 chance of being a top or bottom REI was created to contrast with observed data (Allison et al., 1975; Allison and Nestor, 1999): a gene measured in four different cell types has 1×2^4 possible top and bottom REI expression patterns; a gene measured in five cell types has 1×2^5 random expression patterns, etc. Random proportions of genes with top-REIs in different numbers of cell types were estimated as combinations possible for each cell type number, $nC_r = n!/r!(n-r)!$ over the range of $n = 0 - r$, where C = the number of combinations, n = number of cell types, r = relative number of top-REIs in a combination, and $0! = 1$. For example, the relative proportions of genes having randomly expressed top-REIs in four cell types are 0/4, 1/4, 2/4, 3/4, and 4/4 cell types would be 1, 4, 6, 4, and 1, respectively, or $4!/(0)!(4-0)! = 4!/4! = 1$; $4!/(1)!(4-1)! = 4!/3! = 4$; $4!/(2)!(4-2)! = 4!/(2)!(2)! = 6$; $4!/(3)!(4-3)! = 4!/3! = 4$; and $4!/(4)!(4-4)! = 4!/4! = 1$, respectively. The proportions were divided by the 16 total expression patterns possible for four cell types ($2^4 = 16$), to produce a density distribution of 0.0625, 0.25, 0.375, 0.25, and 0.0625 totaling to 1. This density distribution was then multiplied by 18, or the number of HK genes mapping to chromosome 15, producing random estimates of 1.125 (0.0625×18), 4.5 (0.25×18), 6.75 (0.375) $\times$ (18), 4.5 (0.25×18), and 1.125 (0.0625×18) top-REIs occurring in 0, 1, 2, 3, and 4 cell types, respectively, for comparison with the data found for the four copies of the 18 HK genes mapping to chromosome 15. Similar random top REIs distributions were constructed to contrast with the 106, 306, 55, and 35 HK genes measured in five to eight chromosomes.

GEID Corrections

GEID noise in both aneuploid and diploid cell types was corrected by dividing the raw EI of each gene by its average EI in the four diploid cell types to produce a REI ratio. The numbers of EIs and REI ratios of haploid (1C), diploid (2C), or triploid (3C) genes > the (1C)/(2C) medians were analyzed by 2×2 χ^2 analysis. The (1C) median was set to 1, and the (2C) median value of the (1C)/(2C) comparison was used to align the (2C)/(3C) EIs and REI ratios for display.

EI Comparisons of Paired Genes

The 520 HK gene EIs expressed by all ten cell types were directly compared, gene by gene, with each other in 45 separate 2×2 χ^2 models: REI ratio transforms to eliminate GEID variability were not used in this analysis because each pair of compared EIs were from the same gene with presumably identical GEIDs. The results of the 45 chi square comparisons between each cell type were tabulated by generating a Total Expression Grade (TEG) for each cell type, with higher TEGs indicating relatively increased overall EIs for all 520 HK genes compared with all other cell types. The TEGs were calculated as follows: EIs of the same 520 HK genes were compared between two cell types at a time to generate TEGs. The highest EI of a gene pair was given a + for its cell type, and the cell type with the lower EI a 0; differences in numbers of + and 0 values for all genes in both cell types were estimated by chi square analysis ($P < 0.001$). TEGs were calculated by 2×2 χ^2 comparisons between all cell types; the cell type with significantly more + than 0 values ($P < 0.001$) was awarded a 1 and the other cell type a 0. Cell type pairs with similar EIs ($P \geq 0.001$) were each awarded 0.5. A cell type's TEG is the sum of its awarded points.

HK Gene Marker (HKGM) Assay

The HK gene marker (HKGM) assay uses DE and HK gene EIs measured in separately extracted A and B cell mRNAs and mRNAs extracted from 50:50 mixtures of intact A and B cells to correct extraction variability and sample normalization errors in estimates of the relative proportions of the A and B HK gene mRNAs. EIs of each DE gene measured in pure mRNA of the expressive cell type were divided by the average HK gene EI in the same mRNA to create a DE/HK gene EI ratio for each DE gene. All DE/HK gene ratios were multiplied by the average HK gene EI of the same HK genes in each mRNA mixture to establish baseline, 100% EIs for each DE gene in each mRNA mixture. The DE gene EIs measured in mRNA mixtures were divided by the appropriate baseline EIs, multiplied by 100, and averaged. The average DE EI %s, or HK gene markers (HKGMs), are proportional to the number of HK gene transcripts from each cell

type in the mRNA mixtures. Since the HKGMs are based on DE/HK gene transcript ratios for each cell type, if two pure mRNAs have ~similar numbers of HK gene transcripts, then both HKGMs in a 50:50 mRNA mixture will be ~50%. However, if the mRNA of one cell type has twice as many HK gene transcripts as the other, its HKGM in a 50:50 mixture would be twice that of the other.

RESULTS

Decreased Sensitivity of Low Signal-Strength mRNA Measurements

The sensitivity of the expression intensity (EI) measurements to detect 25% mRNA concentration differences at varying signal strengths was tested: The EIs of 1,931 differentially expressed (DE) genes between the A549 and DU 145 cell lines were measured in pure 100% A549 and DU 145 mRNAs, and in 75/25% and 25/75% A549 and DU 145 mRNA mixtures, with DE genes + expressed in one cell type and − expressed in the other cell type. The EIs for the DE genes measured at 100 and 75% mRNA concentrations were entered into adjacent spreadsheet columns, placing the EIs of the same genes measured at 100 and 75% mRNA concentrations adjacent to each other. The spreadsheet rows were ordered from highest to lowest of the 100% mRNA concentration EIs. If the arrays can detect ~25% mRNA differences, then the measured EIs should parallel the 100 and 75% mRNA concentrations, with measurement errors present when EIs increase between 100 and 75% mRNA concentrations (Xs, Supporting Information Table S1).

The measurement error rates increased for genes with lower EIs (Figs. 1Aa–d and Supporting Information Table S1). Since HK genes generally have higher EIs than DE genes (Su et al., 2004), many DE genes have low EIs and high measurement error rates (Fig. 1Aa,b, green), whereas 97% (503/520) of the HK genes had EIs >300 and measurement error rates of only ~2% (4/246) (Fig. 1Ad, red).

Gene Copy Numbers

We selected 520 HK genes mapping to only one locus on chromosomes 1 to 22 for analysis. Diploid cells were assumed to have two copies of each HK gene, or 1,040 HK gene copies total. Copy number estimates in aneuploid cells were made by combined SNP/SKY analyses (Supporting Information Table S2) (Bazeley et al., 2011): The copy numbers of genes mapping to aneuploid chromosomes without derivative chromosomes were assumed to equal the average number of SKY-normal homologs found per karyotype, e.g., A549 chromosomes 4 and 9 have averages of ~2 and ~2.96 normal homologs per karyotype and no derivative chromosomes in the 24 A549 karyotypes examined by SKY, giving estimates of ~2 and ~2.96 copies of these chromosomes' HK genes (Figs. 1B and 1C).

Copy number estimates were also made for 28 aneuploid chromosomes with SNP-maps having loss of heterozygosity (LOH) and heterozygous regions corresponding to SKY-identified translocations or deletions (Fig. 1D and Supporting Information Fig. S1) (Nestor et al., 2007). For example, DU 145 chromosome 1's SNP-map has a p-arm LOH with identical, homozygous SNPs (Fig. 1D, blue bars). These homozygous p-arm SNPs establish the two SKY-identified normal homologs originate from the same parent, otherwise this p-arm would also have heterozygous SNPs. Thus, the heterozygous q-arm SNPs of DU 145 chromosome 1 (red bars) must originate from chromosome 1 DNA in the t(1;4) translocation derived from the other parental homolog. Since there are ~1.88 copies of normal chromosome 1 homologs and ~0.88 copies of t(1;4) translocation per karyotype, the LOH and heterozygous regions of DU 145 chromosome 1 contain ~1.88 and ~2.76 (1.88 + 0.88) HK gene copies, respectively (Fig. 1D). The genetic map positions of the DNA fragments from LN-18 chromosome 6 in the t(4;19;6) and t(6;17) translocations cannot be resolved from the SNP map (Fig. 1E), making it unsuitable for copy number estimates.

There are a total of 3,120 HK genes in the six aneuploid cancer lines (6 × 520) and direct SNP/SKY copy number estimates were possible for 59% (1,831/3,120) of these genes (Fig. 2). The total numbers of HK gene copies in each aneuploid cancer line were estimated by (1) multiplying the average HK gene copy numbers in each line for which SNP/SKY estimates were possible by 520 (Supporting Information Table S3A) and (2) combining the total numbers of HK genes in normal homologs with varying proportions of HK gene numbers in derivative chromosomes (Supporting Information Tables S3B and S4). The copy number estimates based on SNP/SKY averages and normal homolog/derivative chromosome genes converge when derivative chromosomes were calculated to carry ~70% of normal homolog
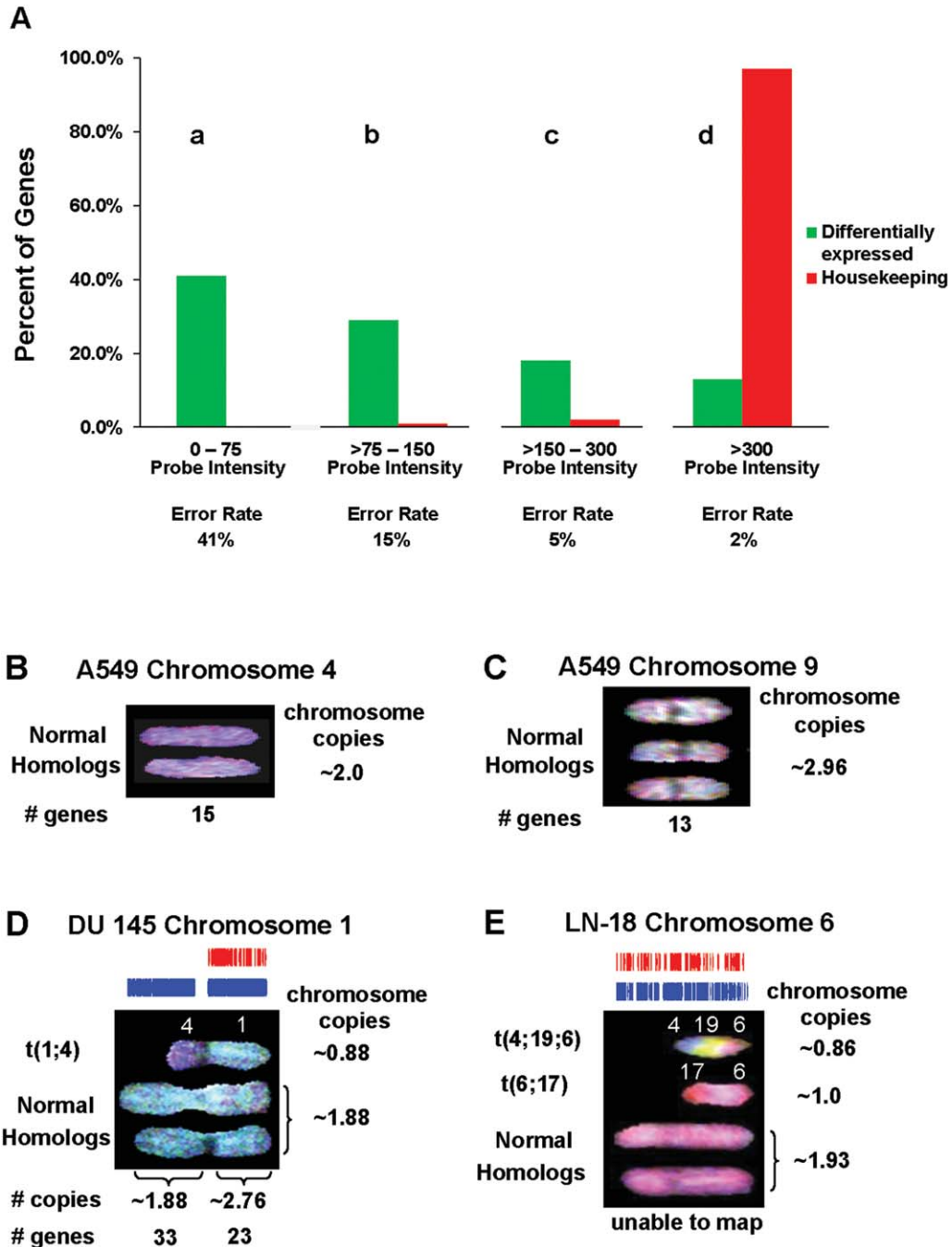
Figure 1. Signal strength measurement errors and SNP/SKY gene copy number estimates. (A) The error rate in detecting 25% mRNA differences was ~41% for genes with EI signals <75 (a), ~15% for EI signals >75–150 (b), ~5% for EI signals >150 to 300 (c), and ~2% for EI signals >300 to 18,500 (d). HK genes are shown as red bars and DE genes as green bars. (B) A549 chromosome 4 has no derivative chromosomes and approximately two normal homologs per karyotype, giving an estimate of ~2 copies for its 15 HK genes. (C) A549 chromosome 9 has no derivative chromosomes and ~2.96 normal homologs per karyotype, giving an estimate of ~2.96 copies for its 13 HK genes. (D) DU 145 chromosome 1's SNP map has a LOH (homozygous SNPs, blue bars) and a heterozygous region (heterozygous SNPs, red bars). The homozygous region establishes the ~two normal homologs (~1.88 copies/karyotype) are from the same parent and the heterozygous DNA is from the t(1;4) translocation (~0.88 copies/karyotype); giving gene copy-number estimates of ~1.88 and ~2.76 for the LOH and heterozygous regions. Also see Supporting Information Figure S1. (E) The genetic map positions of the LN-18 chromosome 6 DNA pieces cannot be resolved, making it unsuitable for copy number estimates.

genes ($r > +0.99$, Supporting Information Table S3A and B); this gave HK gene copy number estimates of 1,191, 1,198, 1,383, 1,391, 1,491, and 2,045 for the PC-3, RWPE-2, LN-18, DU 145, A549, and SUIT-2 cancer lines (Supporting Information Table S3A and B).

PC-3

Chromosome		# Copies	# Genes
#5	Het	1.87	13
	LOH	1.00	6
#6	Het	2.96	15
	LOH	2.03	15
#7		2.56	20
#9		1.31	13
#11		2.51	29
#13		2.03	15
#14		2.54	21
#15		1.27	18
#16		1.54	14
#17	Het	1.98	14
	LOH	1.15	13
#18		2.69	11
#19		2.00	25
#20		3.46	13
#21		3.51	7
#22		1.82	8
	Σ		270

RWPE-2

Chromosome		# Copies	# Genes
#1		2.00	56
#2		2.00	47
#3		2.00	36
#4		2.00	15
#5		3.00	31
#6		2.00	30
#7		3.00	20
#8		2.00	23
#10		2.00	21
#11		2.00	29
#13		2.00	15
#15		3.00	18
#16		2.00	14
#17		2.00	27
#19		2.00	25
#21		2.00	7
#22		2.00	8
	Σ		422

SUIT-2

Chromosome		# Copies	# Genes
#1		3.92	56
#2		4.00	47
#5		3.97	31
#7		4.07	20
#8	Het	4.72	18
	LOH	3.97	5
#9		3.85	13
#10		3.92	21
#15		3.47	18
#16		3.84	14
#18	Het	1.92	3
	LOH	1.00	8
#19	Het	3.97	20
	LOH	2.90	5
#20		3.95	13
	Σ		292

DU 145

Chromosome		# Copies	# Genes
#1	Het	2.76	23
	LOH	1.88	33
#3	Het	2.69	11
	LOH	1.88	25
#6	Het	2.75	14
	LOH	1.89	16
#7		2.89	20
#8		2.22	23
#12	Het	2.61	34
	LOH	1.83	6
#16	Het	2.77	8
	LOH	1.91	6
#17		2.89	27
#18	Het	3.46	7
	LOH	1.74	4
#22		1.74	8
	Σ		265

LN-18

Chromosome		# Copies	# Genes
#3	Het	2.90	17
	LOH	1.90	14
#5	Het	2.90	23
	LOH	1.97	8
#7		3.52	20
#8		1.99	23
#9		1.93	13
#13		1.86	15
#15	Het	2.90	13
	LOH	1.97	5
#19	Het	2.79	1
	LOH	1.93	24
#21	Het	2.69	6
	LOH	1.93	1
#22	Het	1.83	7
	LOH	0.90	1
	Σ		191

A549

Chromosome		# Copies	# Genes
#1	Het	2.96	24
	LOH	2.00	32
#3	Het	3.00	17
	LOH	2.00	8
#4		2.00	15
#5		3.00	31
#7		2.96	20
#9		2.93	13
#10		3.00	21
#11	Het	3.00	22
	LOH	2.00	7
#12		2.96	40
#13		2.00	15
#14		3.04	21
#16		2.87	14
#17		3.75	27
#18		2.00	11
#19	Het	4.00	10
	LOH	2.00	15
#20	Het	4.00	1
	LOH	3.00	12
#21		2.00	7
#22		2.00	8
	Σ		391

Legend: ▇ Haploid (green)   ▇ Diploid (yellow)   ▇ Triploid (red)

LOH = Genomic region with loss of heterozygosity
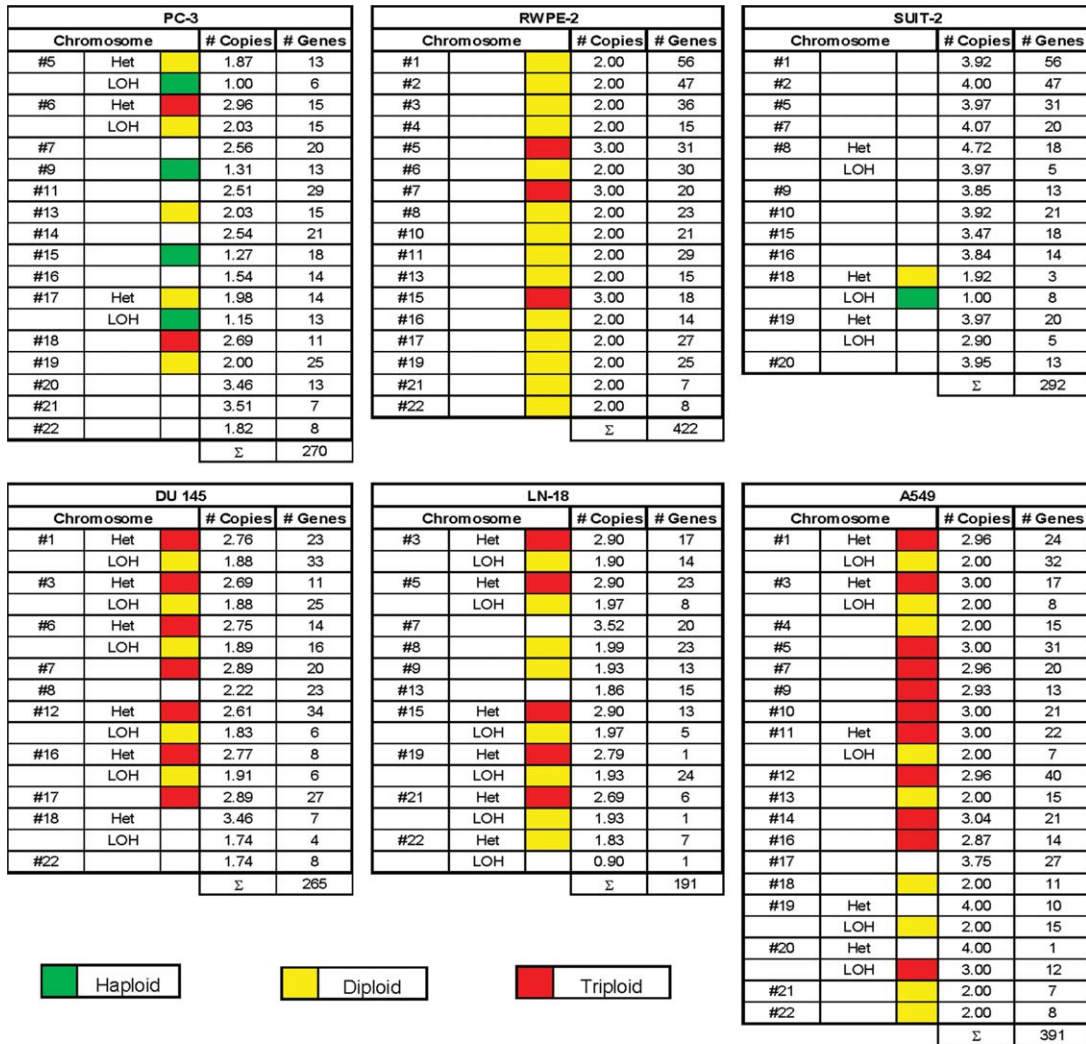
Het = Heterozygous genomic region

Figure 2.   HK gene copy number estimates for six aneuploid cell lines. The copy number estimates are based on the average numbers of normal homologs per karyotype for chromosomes without derivative chromosomes and average copy numbers of normal homologs and derivative chromosome DNA pieces for chromosomes with heterozygous and LOH SNP-maps.

Gene Expression Intensity Determinants (GEIDs)

The mRNA level of a given gene in a specific cell type may be governed by (1) copy number, (2) the global rate of mRNA production and/or degradation in the cell type the gene is expressed in, (3) possibly by intrinsic gene-specific, gene expression intensity determinants (GEIDs) of relative mRNA levels, and (4) other factors not tested in this analysis. GEIDs may produce noise in studies comparing the effects of copy number variations on the mRNA levels of different genes, as high copy-number genes with intrinsically low mRNA levels may have lower mRNA levels compared to low copy-number genes with intrinsically high mRNA levels per allele.

We tested for possible GEID measurement errors by first comparing the relative expression intensities (REIs) of different HK genes with identical copy numbers in the same cell types to eliminate mRNA variations from both copy number differences and possible global variations in overall mRNA production/degradation rates between cell types. Thus, variations in the equal copy number REIs measured in the same cell types, if found, could potentially be due to either gene-specific GEIDs producing similar relative mRNA levels for different genes across all cell types and/or to more complex transcription or mRNA degradation control mechanisms which
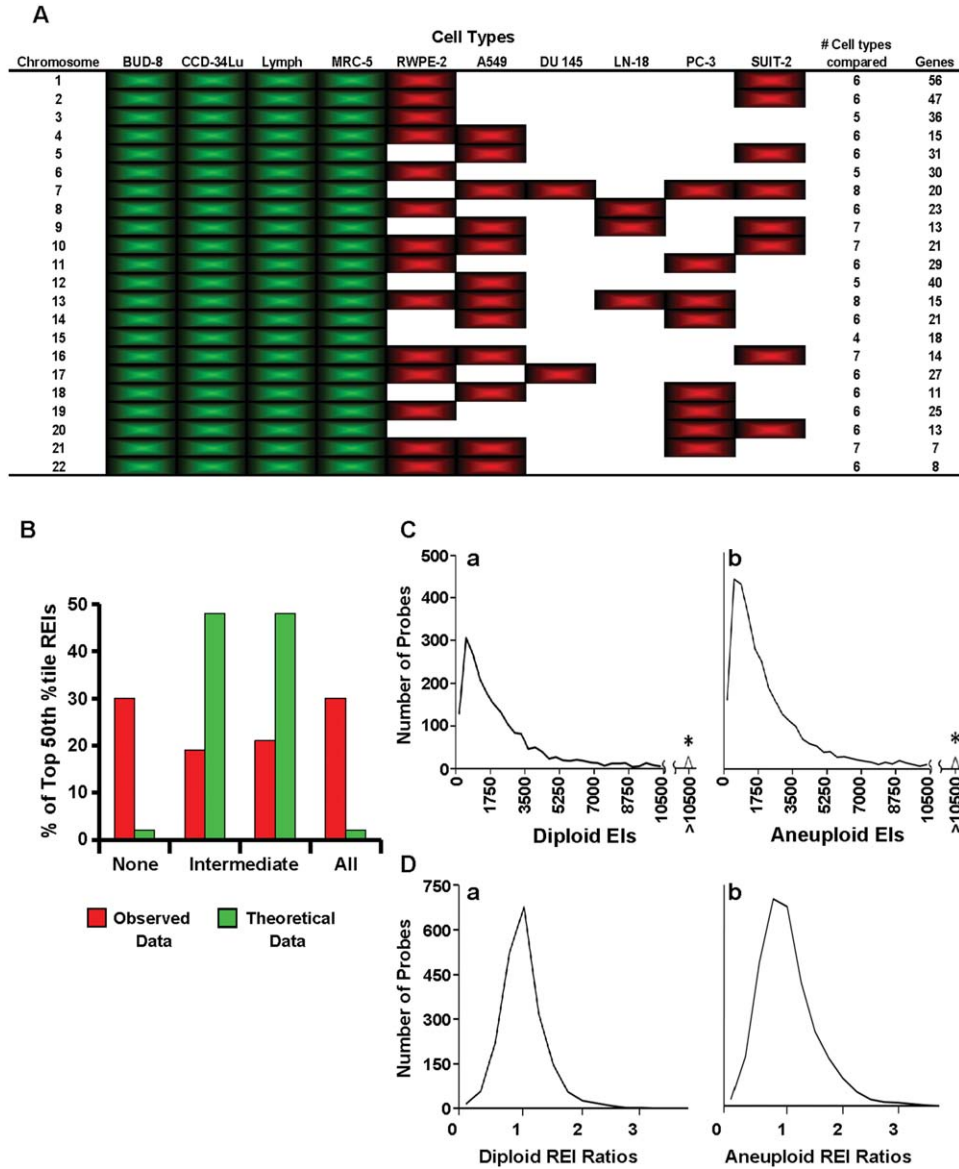
Figure 3. Gene expression intensity determinants (GEIDs) control relative mRNA levels for equal copy number HK genes in the same cell types. (A) HK gene EIs were compared for genes with similar copy numbers in diploid (green) and aneuploid (red) cells. Relative EIs (REIs) of equal copy-number genes in the same cell type were classified as top (top REI) or bottom (bottom REI) 50th%tiles of EI. The top-REI frequencies of genes in each chromosome were then determined in four to eight different chromosomes (across rows), with the numbers of chromosomes and HK genes per chromosome compared shown on the right. (B) GEID control of EIs: 60% (312/520) of HK genes had top-REIs in all, or none, of cell types tested, a significantly higher proportion than the ~4% (20/520) predicted for random REIs ($P < 0.001$, red, green, see Supporting Information Table S5). (C) Asymmetrical distributions of raw diploid (a) and aneuploid (b) HK EIs with tails of high expression-intensity genes: *, intensities >10,500. (D) Symmetrical diploid (a) and aneuploid (b) REI ratio distributions are produced by division of each HK gene EI in the ten cell types by its average diploid intensity.

could vary for the same genes in different cell types. In order to distinguish between these possibilities, the REIs for the same genes measured in each of the cell types were compared across the different cell types; if the same genes had similar REIs in all cell types, then the existence of gene-specific GEIDs regulating individual HK gene mRNA levels would be established: to do this, the EIs of HK genes mapping to similar numbers of normal homologs and without any DNA fragments in derivative chromosomes were compared in each cell type. Each EI was classified as being in the top, or bottom, 50th%tile of Relative EIs (top-REI or bottom-REI) compared to EIs of other equal copy number genes in the cell type. Since diploid cells have two gene copies, the 520 HK gene EIs in each diploid cell type were classified into top- or bottom-REIs (Fig. 3A, green). The aneuploid

gene sets were more complex. For example, the A549 cell line has two groups of equal copy-number chromosomes without derivative chromosomes: (1) diploid chromosomes 4, 13, 18, 21, and 22; and (2) triploid chromosomes 5, 7, 9, 10, 12, 14, and 16 (Fig. 2); thus, separate REI classifications were made for the 118 diploid and 235 triploid A549 HK genes (Fig. 3A, red).

Top-REI frequencies were compared one chromosome at a time, across four to eight cell types (Fig. 3A, across rows). Each HK gene was classified as (1) low intensity, no top-REIs in any cell type; (2) intermediate intensity, top- and bottom-REIs in different cell types; or (3) high intensity, top-REIs in all cell types: 60% (312/520) of genes were in low (30%, 157/520) or high (30%, 155/520) intensity classes (Fig. 3B, red), significantly more than the ~4% (20/520) estimated for these classes by chance alone ($P < 0.001$, Fig. 3B, green and Supporting Information Table S5). This finding supports HK genes producing similar relative amounts of mRNA in different cell types due to unknown control mechanism(s) which we call GEIDs. To further strengthen this conclusion, we reasoned that each diploid HK gene's EI mirrors its GEID. If true, division of each gene's EI by its average EI in the four diploid cell types would create a REI ratio without GEID noise. This was the case: the asymmetry in the raw diploid and aneuploid EI distributions of the 520 HK genes (Fig. 3Ca and b) almost completely disappeared in the diploid REI ratio distribution (Fig. 3Da) and largely disappeared in the aneuploid REI ratio distribution (Fig. 3Db).

Improved Detection of HK Gene Dosage Effects by GEID Noise Corrections

The raw EIs of 58 haploid, 737 diploid, and 527 triploid genes (Fig. 2, green, yellow, red) are widely distributed and have close haploid (1C), diploid (2C), and triploid (3C) medians (Fig. 4A, green, yellow, red; medians, horizontal lines). In contrast, the haploid, diploid and triploid REI ratios are more compactly distributed with increased separation of the (1C), (2C), and (3C) medians (Fig. 4B) after this reduction of the GEID noise.

Detection of gene dosage effects was improved by making initial EI comparisons within the same cell type followed by combined data analyses: 54% (282/527) of the raw triploid A549, DU 145, LN-18, PC-3, and RWPE-2 EIs combined as a single group were more than the (2C) median compared with 47% (348/734) of similarly combined

diploid EIs ($P = 0.035$, Fig. 4C). However, when numbers of raw diploid and triploid EIs more than the (2C) medians were determined separately in each cell type followed by combined data analysis, 59% (313/527) of triploid EIs were more than the (2C) median compared with 50% (366/734) of diploid EIs ($P = 8.9 \times 10^{-4}$, Fig. 4D).

The elimination of GEID noise by analysis of REI ratios rather than raw EIs also improved detection of gene dosage effects: 73% (387/527) of triploid REI ratios, combined after separate determinations of the numbers of REI ratios more than the (2C) medians in each cell type, were more the (2C) median ($P = 8.6 \times 10^{-18}$, Fig. 4E) compared with 59% (313/527) when raw EIs were similarly analyzed ($P = 8.9 \times 10^{-4}$, Fig. 4D). Also, 81% (69/85) of combined diploid REI ratios were more than the (1C) median compared with 48% (28/58) of combined haploid REI ratios ($P = 5.0 \times 10^{-5}$, Fig. 4F). Thus, haploid (1C), diploid (2C), and triploid (3C) HK genes show progressive increases in mRNA levels when measured in the same cell types, arguing against feedback loops producing diploid (2C) mRNA levels for HK genes with haploid or triploid copy numbers (Figs. 4E and 4F).

Standard EI Measurements Find Similar Diploid and Aneuploid HK Gene mRNA Levels

The EIs of the 520 HK genes expressed by all ten cell types were directly compared, gene by gene, to generate total expression grades (TEGs) for each cell type, with higher TEGs indicating relatively increased overall EIs for all 520 HK genes when compared to the other cell types. See Materials and Methods and Supporting Information Figure S2 for calculation of TEG scores. Since the HK gene mRNA levels increase with copy number comparisons made in the same cell types (Figs. 4E and 4F), we expected the supra-diploid, aneuploid HK gene copy numbers (Supporting Information Table S3) would lead to increased aneuploid HK gene mRNA levels compared with HK genes in the diploid cell types. This was not the case, as the TEGs were not related to the HK gene copy numbers in the ten cell types ($r = -0.15$, Fig. 5A). For example, the diploid lymphocytes had the highest TEG score of all ten cell types (Fig. 5A). This result was confirmed by calculation of the average REI ratios for all 520 HK genes in each cell type; Again, no correlation was found between the average HK gene REI ratios and gene copy numbers ($r = -0.24$, Fig. 5B). These results are in sharp contrast to those of Figure 4 in which clear HK gene dosage effects
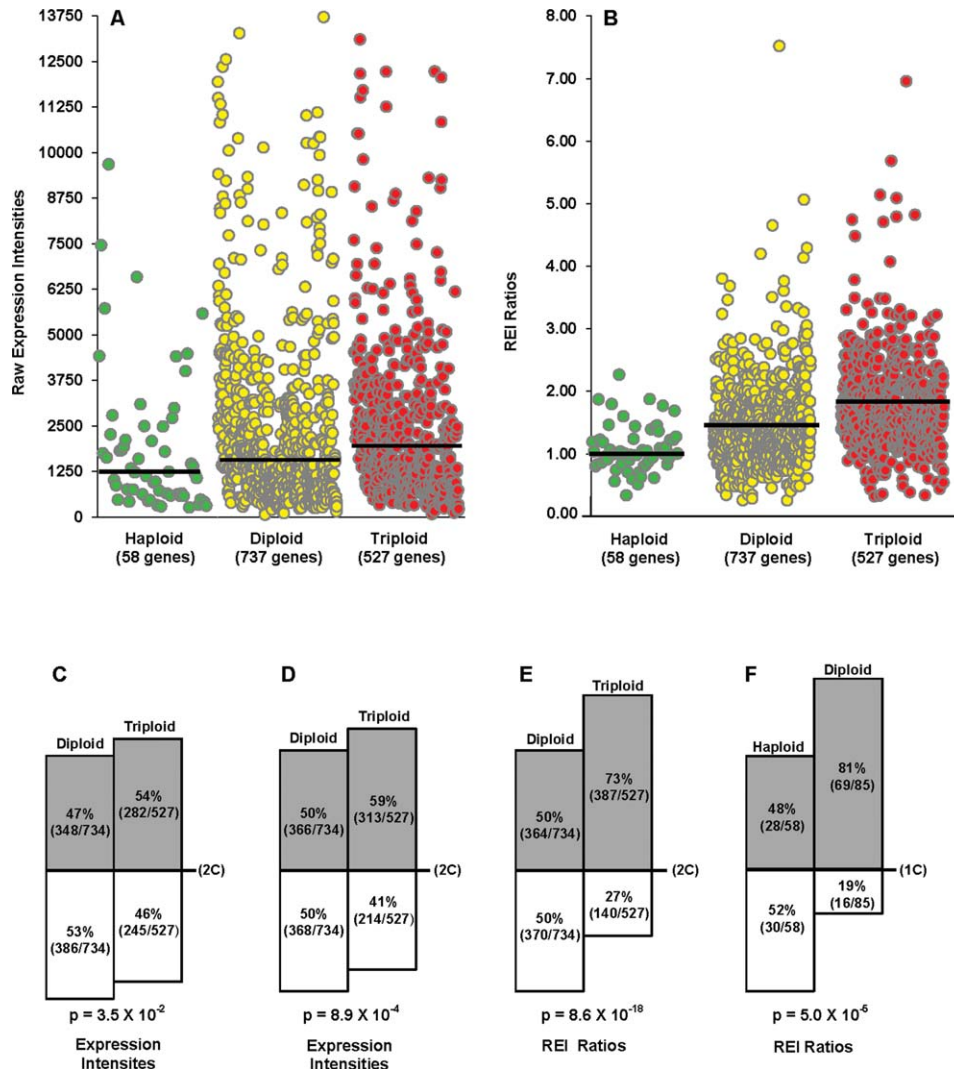
Figure 4. Corrections for GEID noise find HK gene mRNA levels are not controlled by feedback loops. (A) The combined, raw EIs of haploid (green), diploid (yellow), and triploid (red) A549, DU 145, LN-18, and RWPE-2 HK genes are widely distributed. See Figure 2 for gene locations and copy numbers. (B) Relative EI ratios (REI ratios) of the haploid, diploid, and triploid genes are more compactly distributed and have increased separation of the (1C), (2C), and (3C) medians (black lines). (C) Initial combination of all raw diploid and triploid EIs into two separate groups followed shows only a marginal difference between diploid and triploid EIs ($P = 0.035$), with the three diploid SUIT-2 HK genes excluded because SUIT-2 did not have any triploid genes (Fig. 2). (D) A significant difference between the raw diploid and triploid EIs was found when the same raw EI values underwent separate determinations of the numbers of diploid and triploid EIs more than the (2C) medians in each cell type prior data combination and analysis. (E) REI ratios eliminate GEID noise in gene dosage measurements. The numbers of diploid and triploid REI ratios more than the (2C) median show a highly significant difference compared to a similar analysis of raw diploid and triploid EIs (D), establishing that triploid HK genes do not lower mRNA levels back to diploid values by feedback loops. (F) The numbers of haploid and diploid REI ratios more than the (1C) median in the PC-3 and SUIT-2 lines also show a highly significant difference, establishing that haploid HK genes do not increase mRNA levels to diploid values by feedback loops.

were found for comparisons within the same cell type.

mRNA Extraction and Sample Normalization Errors

The paradox of finding HK gene dosage effects for comparisons made within the same cell types (Fig. 4), but not between different cell types (Fig. 5), could be explained by (1) the existence of global mRNA control mechanism(s) maintaining roughly similar overall HK gene mRNA levels in different cell types regardless of gene copy numbers; or (2) the results of Figure 5 being incorrect because of mRNA extraction and sample normalization errors.

Variable amounts of mRNA are often recovered in separate extractions of identical numbers of the same cell types. For example, the replicate mRNA extractions of equal numbers of PC-3 cancer cells without further purification steps averaged 25.72 ±
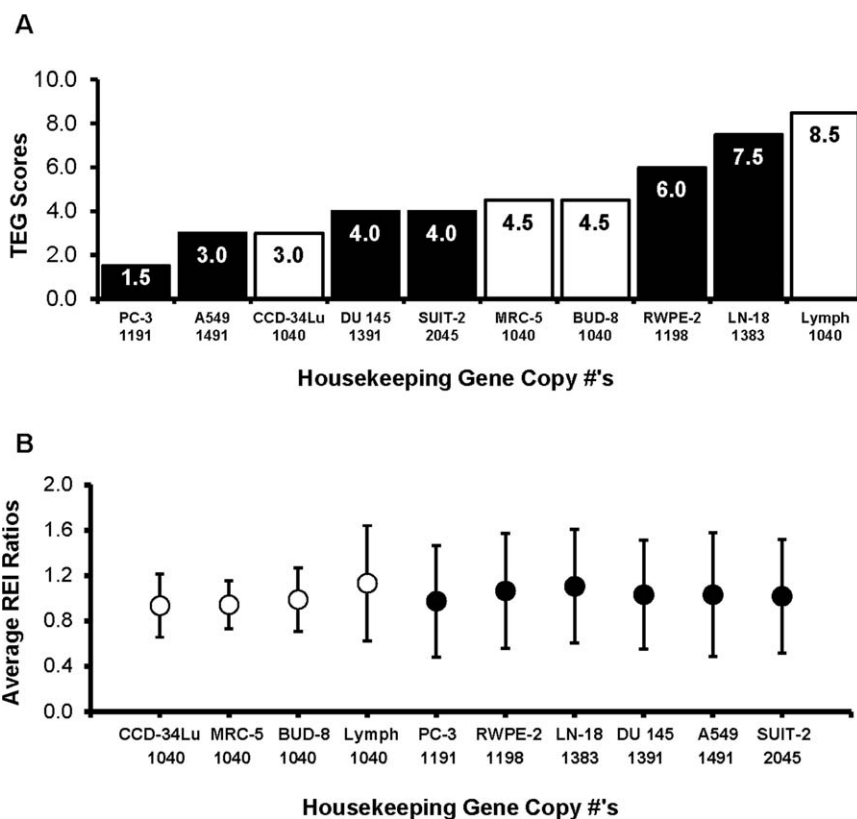
A

B

Figure 5.   Standard expression array measurements find similar diploid and aneuploid HK gene mRNA levels despite copy number differences (see Supporting Information Table S3). (A) Total expression grades (TEGs), derived from comparisons of all 520 HK gene EIs each cell type, are not related to HK gene copy numbers (see Supporting Information Fig. S2). (B) The average REI ratios ($\pm$s) for all 520 HK genes are also not related to each cell type's number of HK genes or aneuploid or diploid chromosomal compositions. White, diploid cell types; black, aneuploid cell lines.

9.4% µg, replicate mRNA extractions of the same number of lymphocytes averaged $16.12 \pm 6.9\%$ µg, and replicate mRNAs extracted from the same numbers of SUIT-2 cells in another experiment averaged $46.8 \pm 44\%$ µg. These large variations in the amounts of mRNA recovered from separate extractions of the same numbers of identical cells are compounded by subsequent mRNA sample normalization steps, such as further purification, amplification and labeling, which often lead to identical amounts of labeled mRNA from each extracted mRNA sample being submitted for analysis. It is difficult to work back through these highly variable mRNA extraction and purification/labeling steps to reconstruct the tenuous connections between the actual total amounts of mRNA in the original cells and the mRNA samples submitted for analysis.

   To illustrate these errors, assume cell A has 1,000 mRNA molecules per cell each represented by a marble. Further, these 1,000 mRNA molecules are divided into 500 HK gene transcripts (white marbles) and 500 DE transcripts (red marbles). Thus, a random sample of 100 cell A marbles following

mRNA extraction would have 50% (50/100) of white and red marbles (Fig. 6A). Now consider cell B, which has 3,000 mRNA molecules per cell, consisting of 1,500 HK gene transcripts (white marbles) and 1,500 DE transcripts (green marbles). A random sample of 100 cell B marbles following mRNA extraction would also have 50% (50/100) of white and green marbles (Fig. 6B). Finally, if the two random samples of 100 cell A marbles and 100 cell B marbles were mixed, this 50:50 mixture of equal mRNA amounts would have equal %s (25%, 50/200) of red and green marbles (Fig. 6C). Thus, sample normalization steps produce the same amount of labeled mRNA from both samples being submitted for analysis and lead to a threefold difference between the total HK and DE transcripts in cells A and B being missed (Figs. 6A–6C).

HKGM Corrections of mRNA Extraction and Sample Normalization Errors

   The HKGM assay was used to correct errors from both extraction variability and sample
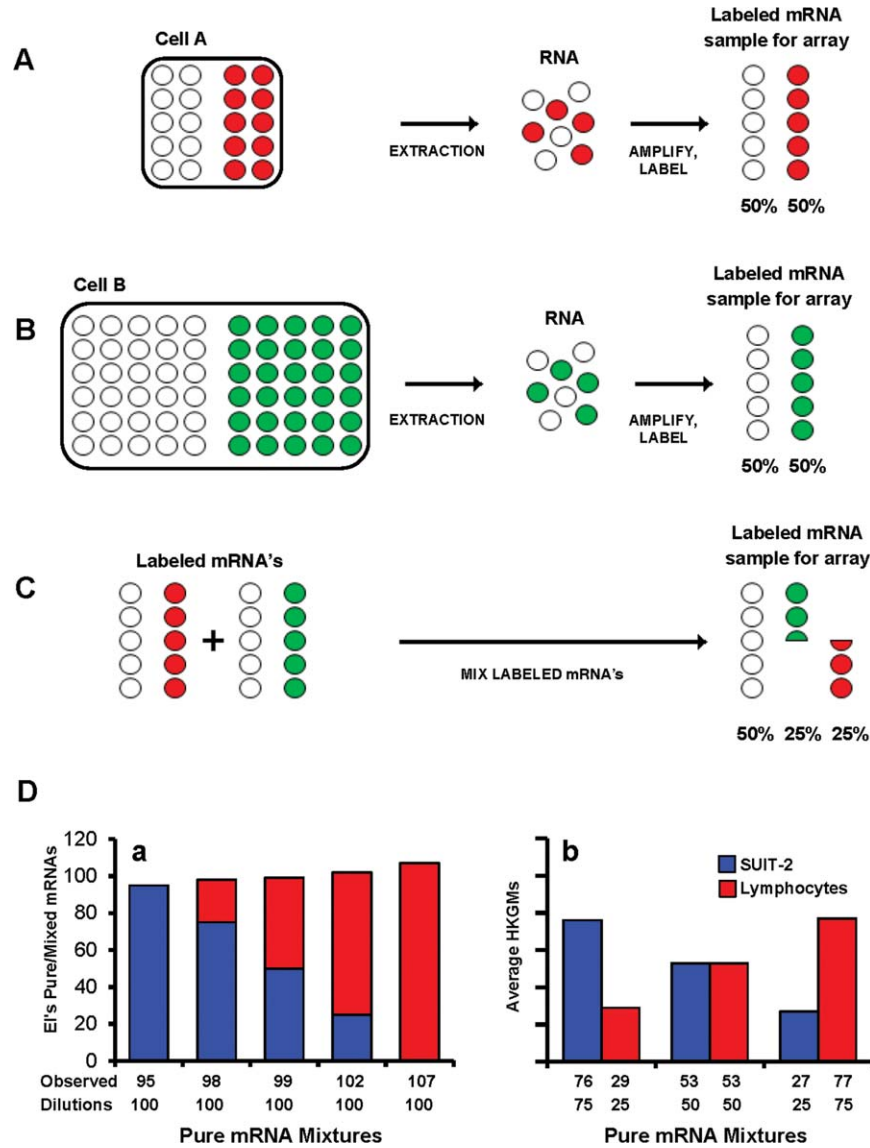
Figure 6. Basis of mRNA extraction variability and sample normalization errors in standard mRNA measurements and the HKGM assay. (A) Cell A has 1,000 mRNA molecules, divided into 500 HK gene transcripts (white) and 500 DE gene transcripts (red). Extraction produces variable amounts of cell A mRNA that are then purified, labeled, and amplified to a fixed mRNA amount for measurement, showing 50% of HK gene transcripts (white) and 50% of DE (red) gene transcripts. (B) Cell B has 3,000 mRNA molecules, consisting of 1,500 HK gene transcripts (white) and 1,500 DE gene transcripts (green). Extraction produces variable amounts of cell B mRNA that are purified, labeled, and amplified to the same, fixed mRNA amount for measurement, also showing an identical 50% of HK (white) and 50% of DE (green) gene transcripts as cell A. (C) Measurement of a mixture of pure, labeled A and B mRNAs produced in (A) and (B) finds 50% HK (white), 25% cell A DE (red), and 25% cell B DE (green) gene transcripts, indicating the normalization errors have prevented detection of the threefold difference between cell A and B mRNA levels. (D) The HK gene marker (HKGM) assay used to correct extraction variability and normalization errors requires that HK and DE gene EIs measured in pure and mRNA mixtures to give stoichiometric results. a. HK gene EIs measured for pure and 75:25, 50:50, and 25:75% mixtures of purified lymphocyte and SUIT-2 mRNAs averaged to 100% were roughly similar at all mixture compositions. b. SUIT-2 and lymphocyte HKGMs measured in pure mixtures of labeled mRNAs also follow the 75:25, 50:50, and 25:75% mRNA dilutions ($r = +0.99$), indicating that purified SUIT-2 and lymphocyte mRNAs measured in pure and mixed mRNAs behave as predicted by the mixture compositions.

normalizations and analyzes DE and HK gene EIs measured in separately extracted pure A and B cell mRNAs and mRNAs extracted from 50:50 mixtures of intact A and B cells to calculate the relative proportions of A and B HK genes. The HKGM assay makes no assumptions about the relative HK and DE gene mRNA levels expressed in

the A/B cells other than (1) all EIs must be high enough for accurate mRNA measurement (Fig. 1A), (2) Cell A DE genes are only expressed in cell A, (3) Cell B DE genes are only expressed in cell B, and (4) HK genes are expressed by both A and B cells. For example, the EIs of the HK and DE genes used to calculate the SUIT-2 and

lymphocyte HKGMs were all more than 300 in the pure (100%) SUIT-2, pure (100%) lymphocyte and the 50:50 intact lymphocyte/SUIT-2 cell mixture mRNAs for measurement accuracy (Fig. 1A); this led to the analyses of the EIs for 472 HK genes, 635 SUIT-2 DE genes, and 411 lymphocyte DE genes in the first repetition of this experiment. The average EIs of the 472 HK genes in the pure SUIT-2, pure lymphocyte, and mixed cell mRNAs were used as baseline values for the EIs measured for the individual DE genes in the pure lymphocyte, pure SUIT-2, and mixed cell mRNAs (see Materials and Methods, HKGM assay).

The accuracy of the HKGM assay depends on stoichiometric EI measurements. The average HK gene EIs measured in pure SUIT-2 (blue) and lymphocyte (red) mRNAs and in 75:25, 50:50, and 25:75 SUIT-2/lymphocyte mRNAs mixtures increased slightly with more lymphocyte mRNA, although all EIs were close to the 100% average (Fig. 6Da); a finding consistent with HK gene EIs being stable in mRNA mixtures. The HKMGs calculated from HK and DE EIs measured in the pure mRNAs and mRNAs mixed after mRNA extraction, purification, and labeling correlated with the 100:0, 75:25, 50:50, 25:75, and 0:100 SUIT-2 and lymphocyte mRNA dilutions ($r = +0.99$), establishing the HKMG assay can gauge relative HK gene mRNA proportions from two different cell types in mRNA mixtures (Fig. 6Db).

The HKGM assay gives an estimate of the relative numbers of HK gene mRNAs present in two different cell types and is based on calculating the relative proportions of DE gene mRNAs from each cell type in intact-cell mixed mRNAs, as illustrated in Figure 7A, e.g., expression array analysis of a true mixture of all marbles in single type A and B cells (Figs. 6A and 6B) finds 50% (2,000/4,000) white marbles, 37.5% (1,500/4,000) green marbles, and 12.5% (500/4,000) red marbles (Fig. 7A). Further, the mRNA expression levels of the HK and DE genes in the two cell types being compared do not influence the HKGM assay. For example, assume (1) cell type C has 20 HK mRNAs (white) and 3 DE mRNAs (red) and (2) cell type D has 60 HK mRNAs (white) and 20 DE mRNAs (green). The proportions of HK and DE gene mRNAs in cells C/D are different from those of A/B cells (Fig. 7A), yet both A/B and C/D cells have identical HK gene mRNA ratios of 0.33, calculated for the C/D cells as follows: The C/D

mRNA mixture has 80 white, 3 red, and 20 green marbles: 12 red marbles (0.15×80) can be estimated for this C/D mixture if no white marble dilution had occurred. White marble dilution from cell D leads to 25% (3/12) of the expected red marbles in the mixture, or (# C white marbles)/(# C + D white marbles) = 0.25 (3/12), or # C white marbles = $(0.25) \times$ (# C + D white marbles) (Equation #1). Similarly, 75% (20/26.66) of expected 26.66 green marbles (0.333×80) are in the C/D mixture because of white marble dilution from cell C, or (# D white marbles)/(# C + D white marbles) = 0.75 (20/26.66), or # D white marbles = $(0.75) \times$ (# C + D white marbles) [Eq. (2)]. Dividing Eq. (1) by (2) shows the ratio of C/D white marbles = (0.25)/(0.75) or 0.33, the same value obtained by analyzing the Figure 7A data (not shown).

Aneuploid Cancer Lines Have Relatively Increased HK Gene mRNA Levels

The HKGMs of (1) lymphocytes and SUIT-2 cells and (2) lymphocytes and PC-3 cells were compared in replicate mRNAs extracted from 50:50 intact cell mixtures (Figs. 7B and 7C) to test whether mRNA extraction/normalization errors explained the relatively low HK gene mRNA levels found for the aneuploid SUIT-2 and PC-3 lines by standard expression array analysis (Fig. 5): Significantly more of the SUIT-2 and PC-3 cancer cell HKGMs were above the diploid (2C) medians than diploid lymphocyte HKGMs ($P < 0.001$, Figs. 7B and 7C). Division of the average PC-3 and SUIT-2 cancer HKGMs by the average diploid lymphocyte HKGMs produces ratios proportional to the aneuploid/diploid HK gene mRNA levels (Fig. 7A). The average SUIT-2/lymphocyte HKGM mRNA ratio was ~3.16 (Expt. #1 ratio = 2.85 [77/27]; Expt. #2 ratio = 3.48 [87/25]), indicating a ~3.16 increase in HK gene transcripts in SUIT-2 cells compared with diploid lymphocytes (Fig. 7D). The average PC-3/lymphocyte HKGM mRNA ratio was ~1.75 (Expt. #1 ratio = 2.19 [86/39]; Expt. #2 ratio = 1.32 [89/67]), consistent with a ~1.75 increase in PC-3 HK gene transcripts compared to lymphocytes (Fig. 7E). Thus, the relatively low aneuploid HK gene mRNA levels measured by standard expression array analysis for aneuploid cells with high HK gene copy numbers (Fig. 5) were due to mRNA extraction variability and sample normalization errors (Figs. 7D and 7E).
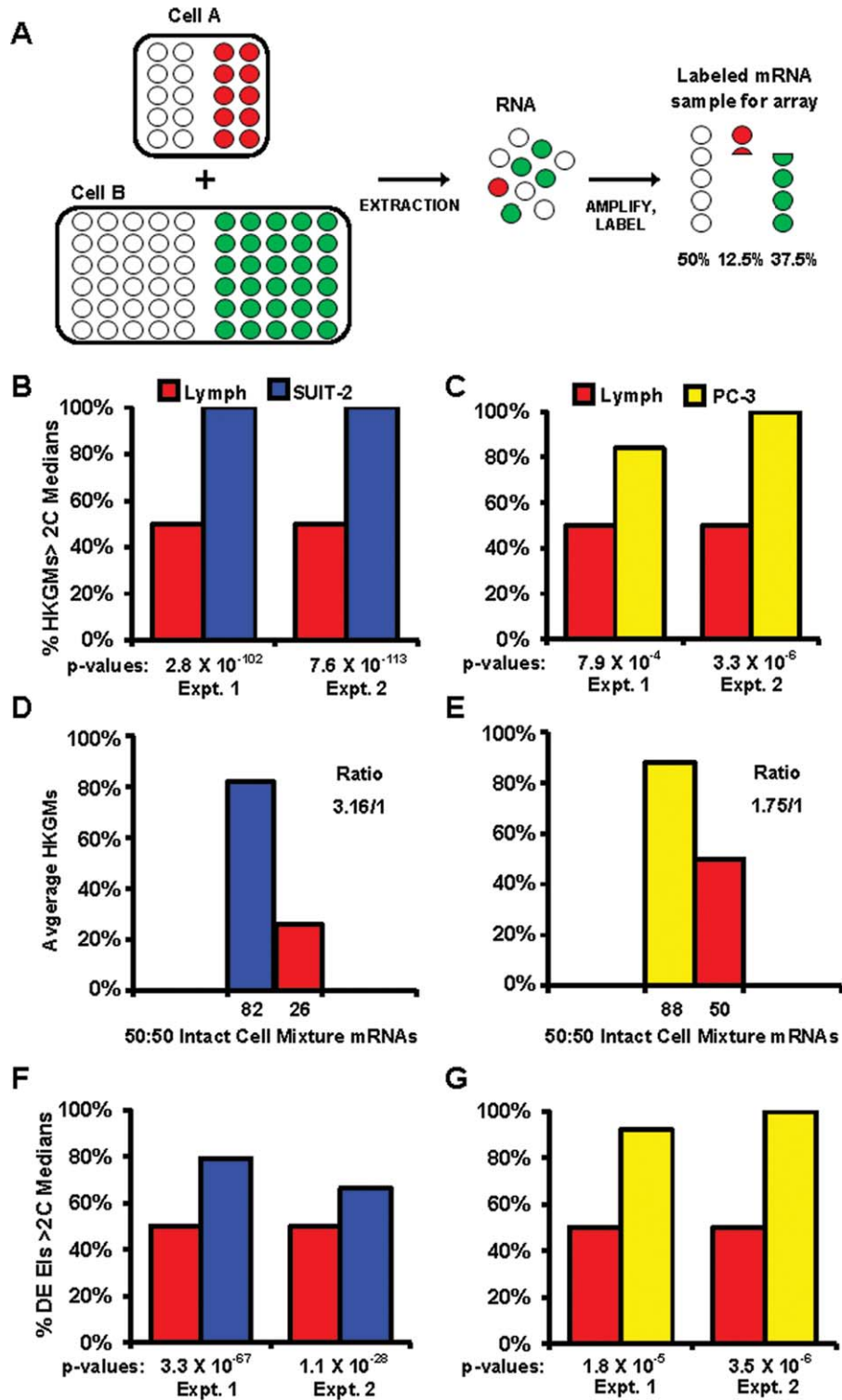
Figure 7. HKGM correction of mRNA extraction and sample normalization errors. (A) Measurements of mRNAs extracted from a 50:50 mixture of A and B cells have 50% of HK gene mRNAs (white), 12.5% of cell A DE gene mRNAs (red), and 37.5% of cell B DE gene mRNAs (green). The 0.33 (12.5/37.5) cell A/B DE mRNA ratio is lower than the 1/1 ratio found when equal amounts of separately prepared A/B mRNAs are combined and measured together (Fig. 6C). This decrease of cell A DE mRNA (red) in the A/B intact-cell mixture mRNA is due to dilution of the cell-type A DE genes by the more numerous cell-type B HK gene mRNAs (white). (B) More of the individual SUIT-2 HKGMs measured in mRNAs extracted from 50:50 mixtures of SUIT-2 cells and lymphocytes were above the (2C) medians than lymphocyte HKGMs in two experimental repetitions ($P < 0.001$). (C) More of the individual PC-3 HKGMs measured in mRNAs extracted from 50:50 mixtures of PC-3 cells and lymphocytes were

above the (2C) medians than lymphocyte HKGMs in two experimental repetitions ($P < 0.001$). (D) The average SUIT-2/lymphocyte HKGM ratios measured for mRNAs extracted from the 50:50 intact SUIT-2cell/lymphocyte mixtures were 77/27 (replicate 1) and 87/25 (replicate 2), averaging to 82% and 26%; giving an average ratio of ~3.16/1.00 (82/26) for SUIT-2/lymphocyte HK gene transcripts. (E) The PC-3/lymphocyte HKGM ratios were 86/39 (replicate 1) and 89/67 (replicate 2) in 50:50 intact PC-3cell/lymphocyte mixture mRNAs; giving an average ratio of ~1.75/1.00 (88/50) for PC-3/lymphocyte HK gene transcripts. (F) More of the individual SUIT-2 DE gene EIs measured in 50:50 intact-cell mixture mRNAs were above the (2C) medians than lymphocyte DE gene EIs in two experimental repetitions ($P < 0.001$). (G) More of the individual PC-3 DE gene EIs measured in 50:50 intact-cell mixture mRNAs were above the (2C) medians than lymphocyte DE gene EIs in two experimental repetitions ($P < 0.001$).

Aneuploid Cancer Lines also have Relatively Increased DE Gene mRNA Levels

Since both HK and DE gene copy numbers are elevated in the SUIT-2 and PC-3 lines, mRNA extraction and sample normalization errors may also have led to underestimates of aneuploid DE gene mRNA levels. A variation of the HKGM assay was used to test for this by selecting (1) cancer DE genes with + expression in cancer cell mRNAs, the 50:50 intact cell mixture mRNAs, and − expressions in lymphocytes and (2) lymphocyte DE genes with + expressions in lymphocyte mRNAs, the 50:50 intact cell mixture mRNAs, and − expressions in cancer cells. These EIs were analyzed as follows: the average EIs of DE genes in lymphocyte (L_{100}) and aneuploid cancer (A_{100}) mRNAs were normalized by $A_{100} = c$ (constant) $\times$ (L_{100}), or $c = A_{100}/L_{100}$. The EIs of the same genes measured in 50:50 intact-cell mixture mRNAs were set to identity by $A_{50:50} = (c) \times (b) \times (L_{50:50})$, with b being the A/L transcript level ratio after correction for mRNA extraction variability and pre-measurement, sample normalization errors, or $b = (A_{50:50}/A_{100}) \times (L_{100}/L_{50:50})$.

More EIs of the DE cancer genes in the 50:50 intact cell mixture mRNAs were above the diploid (2C) medians than the lymphocyte DE gene EIs ($P < 0.001$, Figs. 7F and 7G). The ratios of the average cancer DE gene EIs to average lymphocyte gene EIs in the 50:50 intact cell mixture mRNAs were ~2.54 (Expt. #1, $b = 2.58$; Expt. #2, $b = 2.48$) and ~1.78 (Expt. #1, $b = 1.31$; Expt. #2, $b = 2.26$), respectively, for the SUIT-2 and PC-3 DE genes. Thus, similar to HK genes (Figs. 7B and 7E), cancer DE genes also have elevated mRNA levels compared with lymphocyte DE genes after correcting mRNA extraction and sample normalization errors (Figs. 7F and 7G).

High Gene Copy Numbers and Global Increases in Single-Allelic mRNA Levels both Elevate Aneuploid Cancer mRNAs

The increases in mRNA levels of PC-3 and SUIT-2 HK and DE cancer genes compared to diploid lymphocyte HK and DE genes (Figs. 7B–7G) could be solely from high cancer gene copy numbers, or to a combination of increased gene copy numbers and global mRNA increases from transcription up-regulation and/or decreased mRNA degradation rates for all cancer genes. The mRNA level ratios of single cancer genes to single lymphocyte genes, or mRNA indices, were calculated by dividing the (1) SUIT-2 and PC-3 HKGM and (2) DE/lymphocyte ratios (Figs. 7D and 7E) by each cancer lines' HK gene copy numbers, or 2,045 for SUIT-2 and 1,191 for PC-3, followed by multiplication by the 1,040 lymphocyte HK gene copies (Supporting Information Table S3). The same transform was also used to estimate the relative numbers of SUIT-2 and PC-3 DE genes. If the increased mRNA levels of the cancer genes compared to lymphocyte genes (Figs. 7B–7G) were solely attributable to increased gene copy numbers, then the DE and HK mRNA indices should approach a value of one after this correction for the gene copy number differences.

This was not the case: The PC-3 and SUIT-2 mRNA indices for HK genes were ~1.53 (~1.75 $\times$ 1,040/1,191) and ~1.61 (~3.16 $\times$ 1,040/2,045) and the PC-3 and SUIT-2 mRNA indices for DE genes were ~1.55 (~1.78 $\times$ 1,040/1,191) and ~1.29 (~2.54 $\times$ 1,040/2,045). The SUIT-2/lymphocyte and PC-3/lymphocyte mRNA ratios still had ~50% increases in mRNA levels per allele compared with single lymphocyte alleles even after correction for gene copy number differences. Thus, both gene dosage effects (Fig. 4) and global mRNA increases for single cancer HK and DE gene alleles are responsible for the increases found for cancer cell mRNAs when compared with diploid lymphocytes (Figs. 7B–7G). This conclusion must be tempered, however, because many DE genes were not analyzed in these comparisons because of insufficient EIs for accurate mRNA measurements (Fig. 1A).

DISCUSSION

The effects of gene copy number changes on mRNA levels are not clear (Galitski et al., 1999). Increased mRNA levels are found for some, but not all trisomic genes in aneuploidies such as Down syndrome and following single chromosome transfers into cell lines (Lyle et al., 2004; Upender et al., 2004; Prandini et al., 2007). HER2 gene copy numbers correlate with HER2 protein levels in breast cancer cells (Slamon et al., 1987) and mRNA levels parallel comparative genomic hybridization (CGH) detected CNVs of breast cancer genes in some (Hyman et al., 2002; Pollack et al., 2002), but not all studies (Marella et al., 2009). A poor correlation was found between CNVs and mRNA levels in prostate cancer lines (Jiang et al., 2008). Also, gene copy numbers and EIs strongly correlated in one colon cancer study (Tsafrir et al., 2006) but not in another (Platzer et al., 2002).

There are several possible explanations for these varying results. Certain genes are regulated by transcriptional feedback loops that may dampen CNV effects (Dominguez, 2006; Pasquinelli, 2012). It is possible that some CNVs may perturb broad transcriptome networks and unpredictably alter the mRNA levels of many genes (Galitski et al., 1999; Upender et al., 2004; de Nadal and Posas, 2010). Also, epigenetic changes and random genetic damage can alter mRNA levels without obvious CNVs (Nestor et al., 2007; Jones, 2012; Sproul et al., 2012; Baer et al., 2013). However, technical problems in measuring mRNA levels may contribute to these varying results.

We found accurate mRNA measurements are possible by following several principles. The error rates for assay intensity signals can be determined by calibration against mRNAs of known concentrations (Fig. 1A). If error rates of the genes being studied are too high because of insufficient signal strength, an error signal intensity correction algorithm can be applied (Workman et al., 2002). Stepwise RT-PCR, proteomic assays, or indirectly measuring the amount of gene product present may be needed to quantitate the expression levels of some genes with very low mRNA levels.

We also found that GEIDs control the relative mRNA levels of many HK and possibly other genes (Fig. 3B). Therefore, studies comparing the mRNA levels of different genes in the same cell type should correct for GEID noise, as high copy-number genes with low GEIDs can have decreased mRNA levels compared with low copy-number genes with high GEIDs. The division of each EI by its diploid average to correct GEID noise allowed significant differences to be found between haploid, diploid, and triploid HK gene mRNA levels, establishing that aneuploid HK gene mRNA levels are not controlled by feedback loops which abolish gene dosage effects (Fig. 4).

Errors from mRNA extraction variability and pre-measurement, sample normalizations prior to EI measurements can cause large artifacts in comparisons of DE and HK gene mRNA levels between different cell types: Extractions of mRNAs from identical numbers of the same cell types produce variable amounts of mRNA in each sample. Normalization of the extracted mRNAs to submit a preselected amount of purified and labeled mRNA for analysis further separates the amounts of mRNA present in the original cell type from the measured amounts of mRNA (Figs. 6A–6C): For example, the REI ratios and TEG scores of the HK genes were not related to HK gene copy numbers for the ten cell types measured by standard expression array analyses (Fig. 5). However, approximately two to threefold differences between the HK/DE gene mRNA levels were found between the aneuploid cancer cells and the benign lymphocytes after correction of mRNA extraction and sample normalization errors (Figs. 7D and 7E).

Although some premeasurement errors may be minimized by improved amplification, and labeling of mRNAs from known cell numbers combined with adding exogenous RNA spikes (Benes and Muckenthaler, 2003; Jiang et al., 2011; Loven et al., 2012), these steps are generally performed post-mRNA extraction and cannot completely correct for the variable amounts of mRNA recovered from different extraction procedures. The use of DE and HK gene mRNAs as natural internal spikes to gauge relative transcript levels in mRNAs extracted from 50:50 intact-cell mixtures exposes the mRNAs of both cell types to the same extraction and mRNA normalization assay conditions and is a simple way to minimize these errors in gene expression level comparisons between cell types (Fig. 7). This assay makes no assumptions about HK/DE gene mRNA levels in either cell type, as the HK and DE mRNAs are simply internal markers for overall gene expression levels.

Most HK genes have relatively high mRNA levels (Fig.1A, red) and may be haplo-sufficient in that single alleles can produce enough mRNA for cell growth. Thus, the many extra aneuploid HK genes (Supporting Information Table S3) are probably not selected to produce additional mRNA, but may be remnants of tetraploid intermediates formed in the pathway to aneuploidy (Storchova and Kuffer, 2008; Carter et al., 2012; Davoli and de Lange, 2012). Some of the extra cancer cell HK genes may also be selected to complement defective alleles (Nestor et al., 2007) or retained because of linkage to haplo-insufficient genes (Birchler et al., 2007; Torres et al., 2008), oncogenes, or other non-genic DNA elements under selection in the cancer genome.

Increased cancer cell mRNA levels may be due elevated MYC levels (Lin et al., 2012; Nie et al., 2012) or other mRNA amplification mechanisms. High cancer mRNA levels may be selected by a need to restore the mRNA levels of haplo-insufficient genes created by the genetic instability of cancer (Bazeley et al., 2011) and/or to support oncogene and/or non-oncogene addictions (Luo et al., 2009): There are probably many haploid genes in every cancer genome acquired

during pre-malignant stem cell and tumor cell divisions from random, loss of function mutations, small-scale gene losses, epigenetic changes, and large-scale haploid regions resulting from chromosomal instability (Fig. 2, green and Supporting Information Fig. S1B) (Salk et al., 2010). Some of these haploid genes may be haplo-insufficient and unable to produce enough mRNA for optimal gene function (Birchler et al., 2007). Thus, the ~50% increase in mRNA levels found for the PC-3 and SUIT-2 gene cancer genes (Fig. 7) may allow haplo-insufficient genes to produce enough mRNA for normal cell growth (Bazeley et al., 2011) or to raise the mRNAs of genes selected in oncogene and non-oncogene addictions in order to increase the numbers of rate-limiting proteins and thereby promote tumor growth (Dai and Lu, 2008; Luo et al., 2009; Ruggero, 2009) . In either case, molecular attacks against the transcription amplification mechanism(s) selected in a given cancer may have therapeutic benefits either by allowing latent, growth retarding effects of haplo-insufficient genes to emerge and/or by decreasing the numbers of other crucial mRNA transcripts needed for tumor cell proliferation.

REFERENCES

Albertson DG, Collins C, McCormick F, Gray JW. 2003. Chromosome aberrations in solid tumors. Nat Genet 34:369–376.

Allison DC, Nestor AL. 1999. Evidence for a relatively random array of human chromosomes on the mitotic ring. J Cell Biol 145:1–14.

Allison DC, Meier P, Majeune M, Cohen EP. 1975. Antigen expression of mouse spleen x thymocyte heterokaryons. Cell 6: 521–527.

Baer C, Claus R, Plass C. 2013. Genome-wide epigenetic regulation of mirnas in cancer. Cancer Res 73:473–477.

Barnes M, Freudenberg J, Thompson S, Aronow B, Pavlidis P. 2005. Experimental comparison and cross-validation of the affymetrix and illumina gene expression analysis platforms. Nucleic Acids Res 33:5914–5923.

Bazeley PS, Kalinoski ALN, Ways JA, Liu ST, Ramdath RS, Matsui SI, Allison DC. 2011. A model for random genetic damage directing selection of diploid or aneuploid tumours. Cell Proliferat 44:212–223.

Benes V, Muckenthaler M. 2003. Standardization of protocols in cdna microarray analysis. Trends Biochem Sci 28:244–249.

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M. 2010. The landscape of somatic copy-number alteration across human cancers. Nature 463:899–905.

Birchler JA, Yao H, Chudalayandi S. 2007. Biological consequences of dosage dependent gene regulatory systems. BBA-Gene Struct Expr 1769:422–428.

Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhim R, Pellman D, Levine DA, Lander ES, Meyerson M, Getz G. 2012. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 30:413–421.

Chudin E, Walker R, Kosaka A, Wu SX, Rabert D, Chang TK, Kreder DE. 2001. Assessment of the relationship between signal intensities and transcript concentration for affymetrix gene-chip (r) arrays. Genome Biol 3:research0005-research0005.0010.

Dai MS, Lu H. 2008. Crosstalk between c-myc and ribosome in ribosomal biogenesis and cancer. J Cell Biochem 105:670–677.

Davoli T, de Lange T. 2012. Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse cells. Cancer Cell 21:765–776.

de Nadal E, Posas F. 2010. Multilayered control of gene expression by stress-activated protein kinases. EMBO J 29:4–13.

Dominguez G. 2006. The cart gene: Structure and regulation. Peptides 27:1913–1918.

Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR. 1999. Ploidy regulation of gene expression. Science 285:251–254.

Greshock J, Feng B, Nogueira C, Ivanova E, Perna I, Nathanson K, Protopopov A, Weber BL, Chin L. 2007. A comparison of DNA copy number profiling platforms. Cancer Res 67:10173–10180.

Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi OP, Kallioniemi A. 2002. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 62:6240–6245.

Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B. 2011. Synthetic spike-in standards for rna-seq experiments. Genome Res 21:1543–1551.

Jiang M, Li M, Fu XP, Huang Y, Qian H, Sun RP, Mao YM, Xie Y, Li Y. 2008. Simultaneously detection of genomic and expression alterations in prostate cancer using cdna microarray. Prostate 68:1496–1509.

Jones PA. 2012. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat Rev Genet 13:484–492.

Kushnaryov VM, Sedmak JJ, Bendler JW, Grossberg SE. 1982. Ultrastructural-localization of interferon receptors on the surfaces of cultured-cells and erythrocytes. Infect Immun 36:811–821.

Lengauer C, Kinzler KW, Vogelstein B. 1998. Genetic instabilities in human cancers. Nature 396:643–649.

Lin CY, Loven J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA. 2012. Transcriptional amplification in tumor cells with elevated c-myc. Cell 151:56–67.

Loven J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, Levens DL, Lee TI, Young RA. 2012. Revisiting global gene expression analysis. Cell 151:476–482.

Luo J, Solimini NL, Elledge SJ. 2009. Principles of cancer therapy: Oncogene and non-oncogene addiction. Cell 136:823–837.

Lyle R, Gehrig C, Neergaard-Henrichsen C, Deutsch S, Antonarakis SE. 2004. Gene expression from the aneuploid chromosome in a trisomy mouse model of down syndrome. Genome Res 14:1268–1274.

Marella NV, Malyavantham KS, Wang J, Matsui S-I, Liang P, Berezney R. 2009. Cytogenetic and cdna microarray expression analysis of mcf10 human breast cancer progression cell lines. Cancer Res 69:5946–5953.

Nestor AL, Hollopeter SL, Matsui SI, Allison D. 2007. A model for genetic complementation controlling the chromosomal abnormalities and loss of heterozygosity formation in cancer. Cytogenet Genome Res 116:235–247.

Nie ZQ, Hu GQ, Wei G, Cui KR, Yamane A, Resch W, Wang RN, Green DR, Tessarollo L, Casellas R, Zhao KJ, Levens D. 2012. C-myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. Cell 151:68–79.

Pasquinelli AE. 2012. Non-coding rna micrornas and their targets: Recognition, regulation and an emerging reciprocal relationship. Nat Rev Genet 13:271–282.

Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A, Willson JK, Mack D, Ried T, Markowitz S. 2002. Silence of chromosomal amplifications in colon cancer. Cancer Res 62:1134–1138.

Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO. 2002. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci USA 99:12963–12968.

Prandini P, Deutsch S, Lyle R, Gagnebin M, Vivier CD, Delorenzi M, Gehrig C, Descombes P, Sherman S, Bricarelli FD, Baldo C, Novelli A, Dallapiccola B, Antonarakis SE. 2007. Natural gene-expression variation in down syndrome modulates the outcome of gene-dosage imbalance. Am J Hum Genet 81: 252–263.

Ruggero D. 2009. The role of myc-induced protein synthesis in cancer. Cancer Res 69:8839–8843.

Salk JJ, Fox EJ, Loeb LA. 2010. Mutational heterogeneity in human cancers: Origin and consequences. Annu Rev Pathol Mech 5:51–75.

Singer MJ, Mesner LD, Friedman CL, Trask BJ, Hamlin JL. 2000. Amplification of the human dihydrofolate reductase gene via double minutes is initiated by chromosome breaks. Proc Natl Acad Sci USA 97:7921–7926.

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. 1987. Human breast cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene. Science 235:177–182.

Sproul D, Kitchen RR, Nestor CE, Dixon JM, Sims AH, Harrison DJ, Ramsahoye BH, Meehan RR. 2012. Tissue of origin determines cancer-associated cpg island promoter hypermethylation patterns. Genome Biol 13:R84.

Storchova Z, Kuffer C. 2008. The consequences of tetraploidy and aneuploidy. J Cell Sci 121:3859–3866.

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA 101:6062–6067.

Torres EM, Williams BR, Amon A. 2008. Aneuploidy: Cells losing their balance. Genetics 179:737–746.

Tsafrir D, Bacolod M, Selvanayagam Z, Tsafrir I, Shia J, Zeng Z, Liu H, Krier C, Stengel RF, Barany F, Gerald WL, Paty PB, Domany E, Notterman DA. 2006. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res 66:2129–2137.

Upender MB, Habermann JK, McShane LM, Korn EL, Barrett JC, Difilippantonio MJ, Ried T. 2004. Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. Cancer Res 64:6941–6949.

Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, Saxild HH, Nielsen C, Brunak S, Knudsen S. 2002. A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol 3:research0048. 0041-0048.0016.