BMC Genomics

BioMed Central

Methodology article

Open Access

Comparison of array-based comparative genomic hybridization with gene expression-based regional expression biases to identify genetic abnormalities in hepatocellular carcinoma

Kyle A Furge*[1], Karl J Dykema[1], Coral Ho[2] and Xin Chen[2,3,4]

Address: [1]Bioinformatics Special Program, Van Andel Research Institute, 333 Bostwick Ave. NE, Grand Rapids, MI 49503, USA, [2]Dept of Pharmaceutical Sciences, University of California, S816 513 Parnassus Ave., San Francisco, CA 94143-0046, USA, [3]Cancer Center, University of California, S816 513 Parnassus Ave., San Francisco, CA 94143-0046, USA and [4]Liver Center, University of California, S816 513 Parnassus Ave., San Francisco, CA 94143-0046, USA

Email: Kyle A Furge* - kyle.furge@vai.org; Karl J Dykema - karl.dykema@vai.org; Coral Ho - coralho@itsa.ucsf.edu; Xin Chen - xinchen@itsa.ucsf.edu

* Corresponding author

Published: 09 May 2005

BMC Genomics 2005, **6**:67 doi:10.1186/1471-2164-6-67

Received: 10 January 2005
Accepted: 09 May 2005

This article is available from: http://www.biomedcentral.com/1471-2164/6/67

Abstract

Background: Regional expression biases (REBs) are genetic intervals where gene expression is coordinately changed. For example, if a region of the genome is amplified, often the majority of genes that map within the amplified region show increased expression when compared to genes located in cytogenetically normal regions. As such, REBs have the potential to act as surrogates for cytogenetic data traditionally obtained using molecular technologies such as comparative genomic hybridization. However as REBs are identified using transcriptional information, detection of REBs may also identify local transcriptional abnormalities produced by both genetic and epigenetic mechanisms.

Results: REBs were identified from a set of hepatocellular carcinoma (HCC) gene expression profiles using a multiple span moving binomial test and compared to genetic abnormalities identified using array-based comparative genomic hybridization (aCGH). In the majority of cases, REBs overlapped genetic abnormalities as determined by aCGH. For example, both methods identified narrow regions of frequent amplification on chromosome 1p and narrow regions of frequent deletion on 17q. In a minority of cases, REBs were identified in regions not determined to be abnormal via other cytogenetic technologies. Specifically, expression biases reflective of cell proliferation were frequently identified on chromosome 6p21-23.

Conclusion: Identification of REBs using a multiple span moving binomial test produced reasonable approximations of underlying cytogenetic abnormalities. However, caution should be used when attributing REBs identified on chromosome 6p to cytogenetic events in rapidly proliferating cells.

Background

The parallel analysis of cytogenetic and transcriptional profiling data has revealed that changes in DNA copy number can have noticeable effects on gene expression. Studies comparing wild-type and mutant strains of yeast demonstrated that in regions of increased DNA copy

number (i.e. genomic amplifications), the vast majority of genes that mapped within the amplified region had increased expression when compared to gene expression in non-amplified regions [1]. In this context, the unidirectional change in expression of a large number of adjacent genes can be termed a regional expression bias (REB). The dependence of gene expression on DNA copy number has also been observed with human derived samples, for example in a variety of aneuploid tumors and tumor derived cell lines, and in tissues obtained from patients with inherited trisomy disorders [2-13]. In these samples, ~40–70% of the genes that map to a cytogenetically abnormal region show corresponding expression changes; other genes within the region either do not change expression or, occasionally, change expression in the opposite direction of the cytogenetic abnormality. Nevertheless, as described in yeast, the majority of detectable regional gene expression biases in these mammalian tissues also coincide with chromosomal amplifications or deletions. As such, it is feasible to infer cytogenetic abnormalities by examination of high-density gene expression data. While the majority of REBs correspond to cytogenetic abnormalities, several groups have also identified a subset of regional gene expression biases that do not coincide with detectable DNA copy number changes [2,5,8,9,12,14]. While technical errors between DNA and expression-based approaches may account from some of these differences, it is also possible that other epigenetic factors could produce and regulate the appearance of REBs.

Partitioning gene expression data into subsets of adjacent genes and applying a summary function to each subset is a common method to identify REBs [1,5,11,15-19]. For example, a chromosome can be broken into consecutive, non-overlapping, 100 megabase (Mb) intervals and gene expression values that map to each interval tested for an expression biases using a variety of statistical/computational approaches. While partitioning approaches have been effective in identifying REBs, these approaches may be inherently limited due to the static nature of the partition span. Other, more dynamic, approaches to identify REBs utilizing run and scan statistics have also been reported [20]. However, the utility of these approaches for genome-wide scanning of expression biases is not well described.

Traditional data smoothing approaches ranging from simple moving averages to variable span local regression are common and straightforward methods that can be used to dampen variance and extract trends and patterns from ordered data series. For example, array comparative genomic hybridization (aCGH) data can be smoothed using an exponentially smoothed moving average to more easily identify abnormal chromosomal features [21]. While other approaches, such as hidden Markov models can also be utilized to analyze ordered genomic data [22], the complex nature of gene expression data may prevent the direct application of a subset of these types of analysis techniques. In this report, we outline an approach to identify REBs that summarizes the likelihood that each gene expression value measured lies within an regional expression bias using a multiple span moving binomial test. We use this approach to identify REBs in a set of hepatocellular carcinoma samples and compare the results to high resolution cytogenetic data obtained by aCGH. We also evaluated this approach using a set of clear cell renal cell carcinoma (ccRCC) gene expression profiles. In the majority of cases, dynamically determined REBs coincide with regions of DNA copy number change as determined by other molecular technologies. Interestingly, we identified a region on chromosome 6p where REBs are identified independent of apparent cytogenetic abnormalities. We show that the REBs in this region are produced in the most part by transcriptional responses to cellular proliferation.

Results

Identification of regional expression biases

To identify REBs, a modified version of a moving average is applied to two-color gene expression data obtained from the comparison of tumor HCC tissue to adjacent non-cancerous tissue (Figure 1a). Briefly, to calculate a moving average given a series of gene expression values ordered by genomic location and a window span that consists of five data points, the first five gene expression values would be collected, the average of this set determined, and the result stored as the first element of the moving average. The next span would include the second through the sixth gene expression values and the average of this span stored as the second element of the moving average. This process would continue until the end of the data series and the results of the moving average could be examined to identify trends. To identify REBs from ordered gene expression data, rather then a use an averaging function to evaluate each window span, an approximated binomial test is used to estimate of the probability, in terms of a *z*-score, that a gene expression bias exists within each span (see Materials and Methods). In this case, a positive *z*-score would indicate a disproportionate number of genes within the span show increased expression in the tumor profile when compared to the non-cancerous sample. Analogously, a negative *z*-score would indicate a disproportionate number of genes within the span show decreased expression in the tumor profile when compared to the non-cancerous samples. In addition, rather then collect data from a single window span, a data from a range of spans is collected and summarized (Figure 1b). In this case, the smallest window span used is 25, while the largest window size used is $n/3 = 93$. A minimum span of 25 assures the estimated *z*-scores are

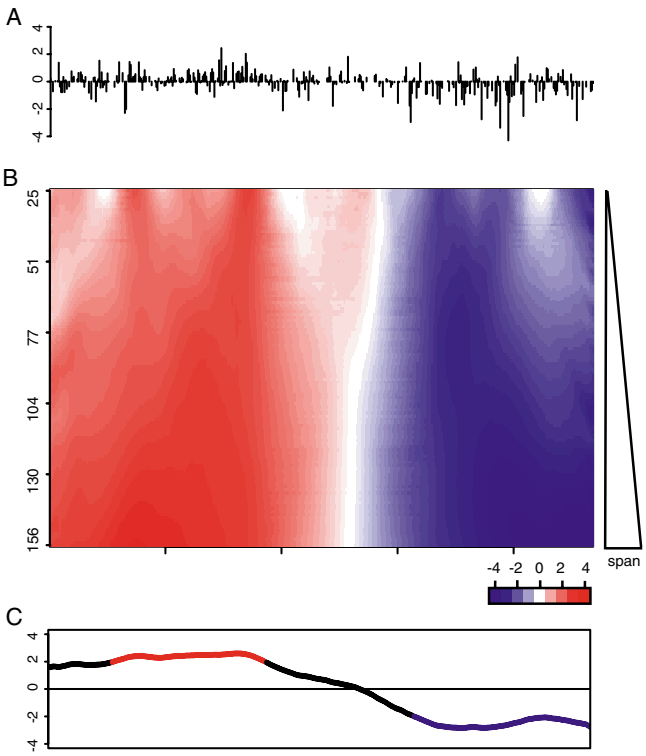

Figure 1
Identification of regional expression biases. A multiple span moving binomial test was applied to gene expression data to identify regional expression biases. **A**. Plot of $\log_2$-transformed tumor verses non-tumor expression ratios (m = 468) that map to chromosome 6 of sample HK1 organized from the p-arm telomere (left) to the q-arm telomere (right). **B**. Heatmap of the set of estimations generated by applying an approximated binomial function (see Materials and Methods) to the gene expression data using window spans of **i** = [25,...,m/3]. Genomic regions that contains a disproportionate number of relatively decreased expression values are shown in blue while genomic regions that show a disproportionate number of relatively increased expression values are shown in red. The color intensity indicates the significance of the expression bias. The highest intensity blue color indicates a z-score ≤ -4 while highest intensity red indicates a z-score ≥ 4 **C**. At each measured loci, an average z-score was computed from the set of estimations from each window span shown in **B** and plotted. Significantly down-regulated regional expression bias estimations are highlighted in blue ($z \leq -1.96$, $p \approx 0.05$) and up-regulated bias estimations highlighted in red ($z \geq 1.96$, $p \approx 0.05$).

reasonably accurate (see Material and Methods) and a maximum span of $n/3$ prevents the generation of largely redundant data. Typical of many types of data smoothers, relatively small spans produce more variable REBs estimations while larger spans produce broader, more diffuse, REB estimations (Figure 1b). To estimate REB boundaries, for each gene loci the mean z-score derived from the range of window sizes is computed (Figure 1c). In addition, for plotting, the final REBs is masked so that only significant regions of bias are displayed. For simplicity, we term this approach IR-CGMA, for Improved Resolution-Comparative Genomic Microarray Analysis keeping in mind we have essentially described the application of an unweighted, multiple span, moving binomial test to identify REBs.

Validation of IR-CGMA

To test the effectiveness of this method, we compared REBs identified by IR-CGMA to aCGH data derived from the same set of samples (Figure 2a, b). Both IR-CGMA and aCGH identified abnormalities that are commonly attributed to HCC such as +1q, -4q, -8p, +8q, -13q, -16q, -17p, and +17q [6]. To summarize the similarities and differences between IR.CGMA and aCGH, the predicted fractional allelic gain or loss was computed at each measured locus (Figure 3a). In the majority of cases, IR.CGMA identified frequent regional expression biases that corresponded to cytogenetic abnormalities as identified by aCGH. For example, on chromosome 1 both approaches identified a narrow region on the q-arm proximal to the centromere (1q21-23) that is frequently amplified. In addition, both approaches identified a region of frequent deletion on the distal tip of chromosome 17 (17p13). While in general REBs corresponded to features identified by aCGH there are regions of discrepancy. The most striking discrepancy between REBs and aCGH/CGH is located on chromosome 6p. Gain of chromosome 6p21-23 is not a frequently reported cytogenetic event in HCC either in this study or in other cytogenetic studies of HCC. However, chromosome 6p was frequently identified to be transcriptionally abnormal via REB scanning. Additionally, while gain of chromosome 17q frequently occurs in HCC, there is some discrepancy between the fraction of samples reported by IR-CGMA and aCGH.

Given these discrepancies, to provide additional validation for the use of a multiple span binomial test to identify regional expression biases, REBs were also identified from a set of gene expression data derived from clear cell renal cell carcinomas (ccRCC). Like HCC, ccRCC presents with a consistent set of cytogenetic abnormalities including loss of 3p and gain of 5p [23]. Frequent gain of chromosome 12p has also been reported in some CGH studies of ccRCC [24]. While in this study, we do not have corresponding cytogenetic data for these specific samples to perform direct comparisons, IR-CGMA did identify abnormalities that overlap genetic abnormalities frequently identified in ccRCC, including loss of 3p and gain of 5p (Figure 4, 5). Interestingly, gain of chromosome 6p is not a frequent cytogenetic abnormality associated with

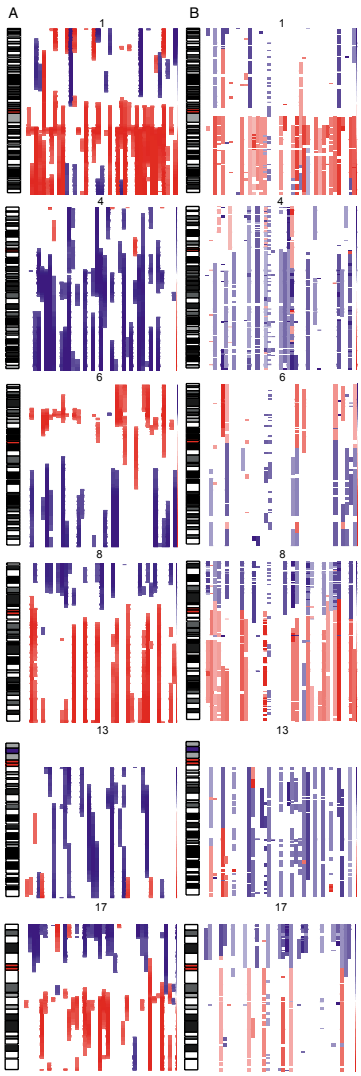

Figure 2
Identification of REBs and DNA copy number abnormalities from individual HCC samples. 39 HCC samples were analyzed for REBs from gene expression data using IR-CGMA and for DNA copy number abnormalities from aCGH data using hidden Markov modelling. Corresponding chromosome ideograms for chromosomes 1, 4, 6, 8, 13, and 17 are also shown to scale. The red bars in the ideogram highlight the centromere. **A**. IR-CGMA estimations were plotted as a heatmap to indicate significant expression biases as described in Figure 1. For consistent plotting, z-scores > 4 and z-scores < -4 were set to 4 and -4 respectively. Scales ranging from 4 to -4 are shown adjacent to each graph. Data for all autosomal chromosomes for all samples was also generated [see Additional file 1]. **B**. aCGH predictions of genomic deletions ($s \leq -0.225$, blue) and amplifications ($s \geq 0.225$, red). The highest intensity blue color indicates a $s \leq -1$ while highest intensity red indicates $s \geq 1$. Scales ranging from 1 to -1 are shown adjacent to each graph. Data for all autosomal chromosomes was also generated [see Additional file 2].

ccRCC, however, like the HCC samples, this region was frequently identified as being abnormal via REB scanning. While technical effects associated with either aCGH, traditional, CGH, or IR-CGMA may be responsible a subset of these discrepancies, it is also possible that epigenetic transcriptional regulation could contributes to the REBs. Therefore, to determine if the transcriptional abnormalities reflected certain types of epigenetic effects, we examined the gene expression data in more detail.

Examination of chromosome 6p and 17q REBs

To evaluate the nature of the REBs on chromosomes 6p and 17q in HCC, misregulated genes within these regions were identified and partitioned based on Gene Ontology (Figure 6). Only two significantly enriched ontology's were identified ($p < 0.005$) from the upregulated genes in these regions: nucleic acid metabolism (GO:0006139) and cell proliferation (GO:0009607) [25,26]. While a small number of transcripts that had relatively increased expression in the tumor samples were identified as negative regulators of cell proliferation (GO:0008283), overall these results suggest that pronounced REBs on chromosome 6p and chromosome 17q reflect the active cell division of the tumor cells compared to non-cancerous cells. To test this hypothesis, up-regulated genes mapping to these ontologies were removed from the HCC gene expression dataset (154 of 8128 genes, 1.9%) and REBs recomputed (Figure 3b). The REBs on chromosome 6p were considerably diminished and the discrepancy on chromosome 17q was partially diminished. In contrast, REBs on chromosome 1q and 8q were not appreciably changed after removing the cell proliferation associated genes. Taken together, these results suggest that the transcriptional effects of active cell proliferation participate in the production of the REBs of 6p and 17q.

Discussion

In this paper, we describe the construction and application of a straightforward data smoothing approach to identify REBs from gene expression data. As evidence for the validity of this approach, we demonstrate that REBs overlap cytogenetic abnormalities as determined using other cytogenetic profiling methods in the majority of cases. Due to the dependence of gene expression on chromosome dosage, identification of REBs can often assist in the interpretation of gene expression data. For example, detection of REBs can rapidly determine if a potential cytogenetic abnormality associates with particular sample classification, for example a more aggressive tumor subtype [27]. Perhaps more importantly, the prevalent overlap of transcriptional and cytogenetic abnormalities support HCC tumorigenesis models that advocate that recurrent cytogenetic aberrations, via their significant influences on gene expression, play important roles in HCC pathogenesis. In addition, the correlation between

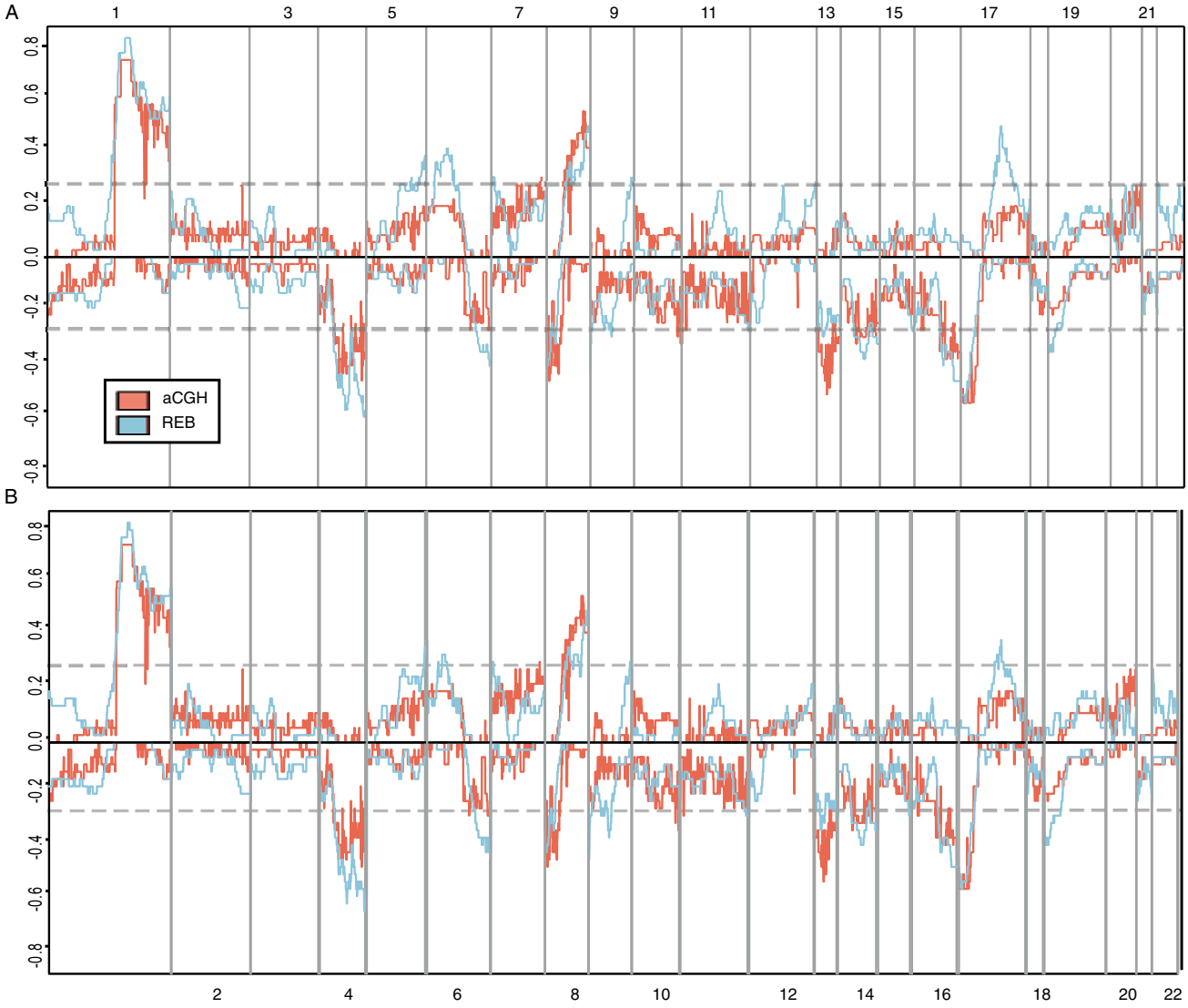

Figure 3

Summary of REBs and DNA copy number changes in HCC. A summary of the data generated as described in Figure 2. **A**. For each genetic loci on the autosomal chromosomes, the fraction of HCC samples that contained significant upwards expression biases were plotted as a positive fraction and the fraction of samples that contained significant downwards expression bias were plotted as a negative fraction. DNA copy number data determined by aCGH was plotted in a similar manner. **B**. Data was plotted as in **A**., with the exception that genes involved in cell proliferation and nucleic acid metabolism were removed, as described in the text, before REBs were identified.

REBs and specific DNA copy number variation can assist in identification of candidate genes that have important function during tumorigenesis in a specific chromosomal regions. For example, a narrow region on the q-arm of chromosome 1 proximal to the centromere (1q21-23) is predicted to be frequently amplified both by IR.CGMA and aCGH, suggesting that this region may harbour candidate oncogenes. Inspection of genes that are highly expressed in 1q21-23 included several signalling molecules (MDUSP12, SHC1) and transcriptional factors (MEF2D, ILF2, TCFL1). Particular interesting is ephrin-A1, the ligand of Eph receptor tyrosine kinase. Ephrin-A1 has been implicated in angiogenesis and therefore may contribute to HCC development [28]. Clearly, it is impor-

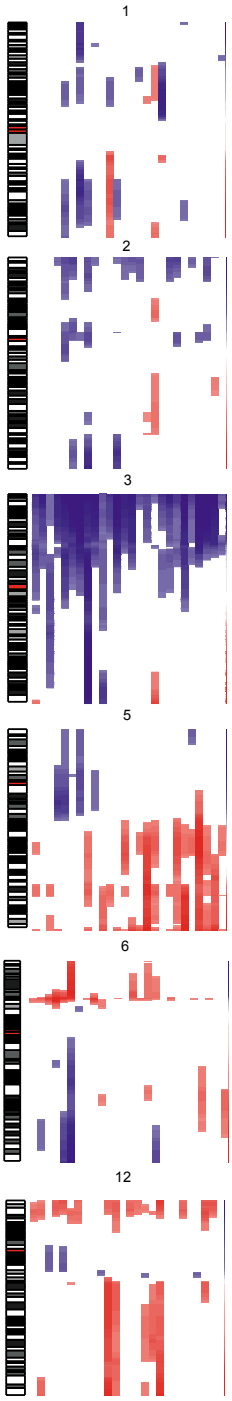

Figure 4
Identification of REBs from individual ccRCC samples. 27 ccRCC samples were analyzed for REBs and plotted as described in Figure 2a with the exception that chromosomes 1, 2, 3, 5, 6, and 12 are shown. Chromosomes 1 and 2 are shown as representative regions that do not frequent REBs. Data for all autosomal chromosomes was also generated [see Additional file 3].

tant to evaluate the functions of these genes in HCC and determine the extent in which their gene expression is regulated by DNA amplification.

We also demonstrate in this study that not all REBs corresponded with detectable cytogenetic abnormalities, particularly in the region of chromosome 6p. Therefore, it is appropriate to apply alternative molecular approaches before attributing cytogenetic abnormalities to regional expression biases located in this region. Classification of the differentially expressed genes in this region into Gene Ontologies suggests that the regional expression changes reflect aspects of tumor cell proliferation as evidenced by an enrichment of features classified in nucleic acid metabolism and cell proliferation GO categories. Another notable feature of chromosome 6p, particularly 6p21-23, is that the gene density in this region is unusually high and harbors gene clusters of several protein families [29]. The unusually high gene density may also contribute to the identification of this region as frequently abnormal by REB scanning. It has been suggested that regions of high gene density correlates with open chromatin fibers. This open chromatin structure may facilitate transcriptional activation if appropriate transcriptional signals are present [30]. Other possible explanations of the REBs include regional methylation or Histone deacetylation.

While the high variability of gene expression data may prevent the direct application of several data modelling approaches, this study suggests that application of traditional data smoothing methods are appropriate to infer cytogenetic abnormalities from gene expression data and are worth investigating further. One potential disadvantage of smoothing approaches can be difficulty determining an appropriate window span that balances overall smoothness with optimal feature identification. While cross-validation using training and test data sets could theoretically identify an optimal window span for regional expression bias identification, we could not derive a span that was appropriate for all chromosomal regions across multiple data sets (data not shown). However, the increase in computer processing power allows the utilization of more computational intensive multiple span approaches to partially compensate for single span effects.

Unlike traditional cytogenetic analysis approaches, the resolution of this technique has a complex dependency on gene density, gene coverage on the array platform used, and tissue-dependent expression patterns. On average, the genome contains about 10 genes per Mb and varies between regions that have gene densities of ~6 genes per Mb (chromosome 13) to regions that have gene densities of ~26 genes per Mb (chromosome 19) [31]. As the smoothing approach presented requires at least 25 gene

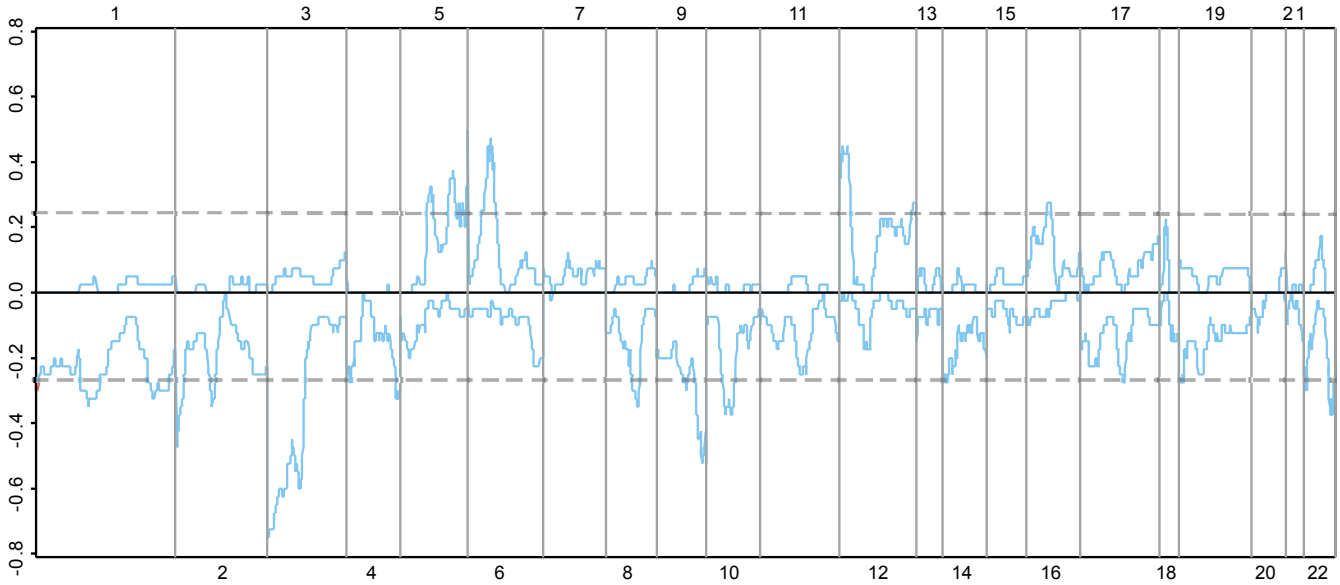

Figure 5
Summary of REBs in ccRCC. 27 ccRCC samples were analyzed for REBs and plotted as in Figure 3a.

expression values to make a prediction, theoretically, the resolution of a REB could average ~2.5 Mb across the genome and range between ~1 to 4 Mb. However, for this analysis the cDNA arrays used contained ~8500 features could be confidently mapped to predicted genes. Of these features ~6000 genes (70%) where expressed at measurable levels in the liver tissue. Assuming ~30,000 human genes, the resolution for this study would be about 5-fold lower then the theoretically limits or average ~12.5 Mb across the genome and range from ~5 Mb to ~20 Mb.

While not reported here, this approach is suitable for single channel gene expression data provided appropriate reference and test expression profiling data can be converted to log-transformed expression ratios. We have also successfully used this approach to infer cytogenetic abnormalities from other species, such as mice and rats.

Conclusion

In this report, we describe a method to identify regional expression biases using a multiple span moving binomial test. As evidence for the validity of this approach, we demonstrate that this methods identifies REBs that associate with cytogenetic abnormalities as determined by array CGH and traditional CGH in both hepatocellular carcinoma and clear cell renal cell carcinoma.

Methods

Pre-processing of gene expression data sets

Two-color gene expression profiles derived from 39 HCC tumor samples and corresponding non-cancerous liver samples [32], and 33 RCC and adjacent non-cancerous kidney tissue samples [33], were obtained from the Stanford Microarray Database [34]. In all cases, gene expression values were normalized using the within-print tip group normalization method as implemented in the BioConductor packages for the R environment [35,36]. Prior to normalization, R and G values were threshold such that R or G values <150 were set to 150. In these data sets, the cancerous and non-cancerous samples were compared to a pooled cell-line reference. To allow direct comparison of tumor to non-cancerous expression values, new gene expression ratios (R) were generated from tumor tissue ratio (T/U) and corresponding adjacent non-cancerous tissue ratios (N/U) such that $R = \log_2(T/U) - \log_2(N/U)$ [2]. Sequence comparisons were used to map microarray probe sequences to predicted Ensembl transcripts (Ensembl version 19) [29]. Included in the Ensembl transcript annotations are chromosomal mapping locations at base-pair resolution. If multiple probes mapped to the same locus a mean gene expression value was utilized.

Pre-processing of array comparative genomic hybridization data sets

Two-color array CGH data for the HCC samples was generated essentially as described [37]. A manuscript describing the details of the HCC copy number data and initial analysis is in preparation. In all cases, copy number values were transformed into copy number states using an unsupervised hidden Markov model as implemented in the BioConductor packages for the R environment [22,36]. States in which the median copy number change

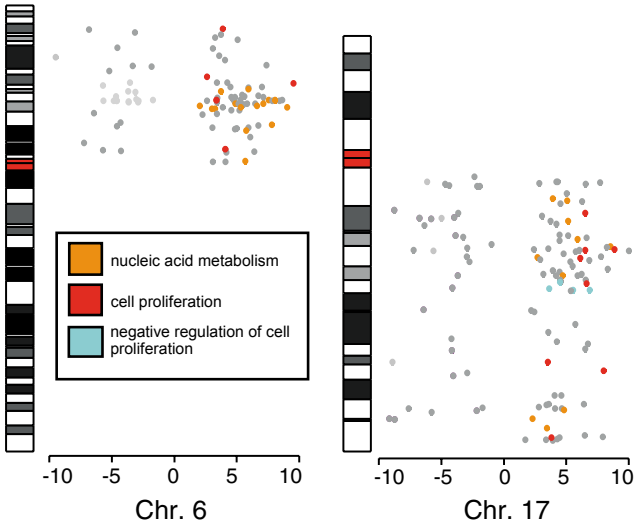

Figure 6
Functional classification of differentially expressed genes on 6p and 17q. Genes on chromosome 6p and chromosome 17q that are differentially expressed in HCC compared to adjacent non-cancerous tissue were identified as described in the Material and Methods. The *t*-statistic corresponding to each misregulated gene ($p < 0.05$) was plotted with respect to gene location. For consistent plotting, *t*-statistics > 10 and *t*-statistics < -10 were set to 10 and -10 respectively. Genes classified as nucleic acid metabolism, cell proliferation, and negative regulation of cell proliferation are highlighted orange, red, and cyan, respectively.

was ≥ 0.225 were defined as region of DNA gains and states in which the median copy number change ≤ -0.225 were defined as regions of DNA loss [37].

Identification of regional expression biases (IR-CGMA method)

Gene expression values were separated into chromosome subsets and ordered by gene mapping location. A sliding window algorithm was applied to each ordered gene expression subset such that within each window span a binomial test was applied under the assumption that the probability (p) of the appearance of a positive relative gene expression value equals the probability (q) of the appearance of a negative relative expression value, $p = q = 0.5$, and a z-score for the span is computed using the normal approximation to the binomial distribution. The z-score can be converted to an approximate significance values based on the two-tailed z-statistic ($z_{a/2}$) critical values. Data was generated using multiple window spans and an average z-score at each gene location was computed. More formally, given a set of ordered gene expression values g_j for genes $j = 1, 2, \ldots m$, let x_{ij} denote expression bias approximations for genes $j = 1, 2, \ldots m$ using window spans $i = 25, 26, 27, \ldots m/3$ where n denotes the number of window spans examined. An empty matrix $\mathbf{X}_{[n\#m]}$ is populated such that $x_{ij} = \frac{1}{i}\sum_{k=j-i+1}^{j}\frac{2r-t}{\sqrt{t}}$ for $m\text{-}i\text{+}2 > j \geq i$ where t denotes the number of non-zero and r the number of positive values within the span $\{g_k, g_{k+1}, \ldots g_{k+i-1}\}$. To not discard regions, x_{ij} is tapered when $j < i$ such that $x_{ij} = \frac{1}{j}\sum_{k=1}^{j}\frac{2r-t}{\sqrt{t}}$ and analogously tapered when $j \geq m\text{-}i\text{+}2$. Final regional expression bias estimates (b_j) are computed such that $b_j = \frac{1}{n}\sum_i x_{ij}$. Performing IR-CGMA on the 39 HCC gene expression profiles took approximately five minutes on a 2.6 GHz Intel Pentium IV with 1 GB of RAM.

Identification of misregulated genes

Identification of misregulated genes from the HCC gene expression profiles occurred in two-steps. First, genes were filtered to ensure each gene was well measured across the data set using an exact binomial test ($p < 0.05$). In this case, data was required in 24 of 39 (64%) of samples. Next, a one-sample t-test assuming unequal variance was applied to determine if expression values were significantly misregulated ($p < 0.05$).

Authors' contributions

KF and KD designed and implemented the data analysis algorithms and performed the data analysis. XC and CH obtained the HCC samples and generated both the gene expression and array CGH data. KF and XC collaborated to write the manuscript. All authors read and approved the final manuscript.

Additional material

Additional File 1

Regional expression biases for all chromosomes in the HCC samples
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-67-S1.pdf]

Additional File 2

aCGH states for all chromosomes in the HCC samples
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-67-S2.pdf]

Additional File 3

Regional expression biases for all chromosomes in the RCC samples
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-6-67-S3.pdf]

Acknowledgements

This work was supported by NIH grant R33-CA10113-01 to K.A.F, and NIH grant K01-CA096774 to X. C.

References

1. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH, Marton MJ: **Widespread aneuploidy revealed by DNA microarray expression profiling.** *Nat Genet* 2000, **25:**333-337.
2. Phillips JL, Hayward SW, Wang Y, Vasselli J, Pavlovich C, Padilla-Nash H, Pezullo JR, Ghadimi BM, Grossfeld GD, Rivera A, Linehan WM, Cunha GR, Ried T: **The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis.** *Cancer Res* 2001, **61:**8143-8149.
3. Xu XR, Huang J, Xu ZG, Qian BZ, Zhu ZD, Yan Q, Cai T, Zhang X, Xiao HS, Qu J, Liu F, Huang QH, Cheng ZH, Li NG, Du JJ, Hu W, Shen KT, Lu G, Fu G, Zhong M, Xu SH, Gu WY, Huang W, Zhao XT, Hu GX, Gu JR, Chen Z, Han ZG: **Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver.** *Proc Natl Acad Sci USA* 2001, **98:**15089-15094.
4. Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, Bloomfield CD, de La Chapelle A, Krahe R: **Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics.** *Proc Natl Acad Sci USA* 2001, **98:**1124-1129.
5. Harding MA, Arden KC, Gildea JJ, Perlman EJ, Viars C, Theodorescu D: **Functional genomic comparison of lineage-related human bladder cancer cell lines with differing tumorigenic and metastatic potentials by spectral karyotyping, comparative genomic hybridization, and a novel method of positional expression profiling.** *Cancer Res* 2002, **62:**6981-6989.
6. Crawley JJ, Furge KA: **Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene expression data.** *Genome Biol* 2002, **3:**RESEARCH0075.
7. Shaughnessy JD, Barlogie B: **Integrating cytogenetics and gene expression profiling in the molecular analysis of multiple myeloma.** *Int J Hematol* 2002, **76:**59-64.
8. Haddad R, Furge KA, Miller J, Schoumans J, Haab B, Teh B, Barr L, Webb C: **Genomic profiling and cDNA microarray analysis of human colon adenocarcinoma and associated peritoneal metastasis reveals consistant cytogenetic and transcriptional aberrations associated with progression of multiple metastases.** *Appl Genom Proteom* 2002, **1:**51-62.
9. Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO: **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* 2002, **99:**12963-12968.
10. Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A, Willson JK, Mack D, Ried T, Markowitz S: **Silence of chromosomal amplifications in colon cancer.** *Cancer Res* 2002, **62:**1134-1138.
11. Mao R, Zielke CL, Zielke R, Pevsner J: **Global up-regulation of chromsome 21 gene expression in the developing Down syndrome brain.** *Genomics* 2003, **81:**457-467.
12. Masayesva BG, Patrick H, Mayer-Garrett E, Pilkington T, Mao R, Pevsner J, Speed T, Benoit N, Moon CS, Sidransky D, Westra WH, Califino J: **Gene expression alterations over large chromosomal regions in cancers include multiple genes unrelated to malignant progression.** *Proc Natl Acad Sci USA* 2004, **101:**8715-8720.
13. Lindvall C, Furge KA, Bjorkholm M, Guo X, Blennow E, Haab B, Nordenskjold M, Teh B: **Combined genetic- and transcriptional profiling of acute myeloid leukemia with complex and normal karyotypes.** *Haematologia* 2004, **89:**1072-1081.
14. Lu YJ, Williamson D, Clark J, Wang R, Tiffin N, Skelton L, Gordon T, Williams R, Allan B, Jackman A, Cooper C, Prichard-Jones K, Shipley J: **Comparative expressed sequence hybridization to chromosomes for tumor classification and identification of differential gene expression.** *Proc Natl Acad Sci USA* 2001, **98:**9197-9292.
15. van Eijk R, Oosting J, Sieben N, van Wezel T, Cleton-Jansen AM: **Visualization of regional gene expression biases by microarray data sorting.** *Biotechniques* 2004, **36:**592-594.
16. Fischer G, Ibrahim SM, Brockmann GA, Pahnke J, Bartocci E, Thiesen HJ, Fernadez-Serrano P, Moller S: **Expressionview: visualization of quantitative trait loci and gene-expression data in Ensembl.** *Genome Biol* 2003, **4:**R77.
17. Breitkreutz BJ, Jorgensen P, Breitkreutz A, Tyers M: **AFM 4.0: a toolbox for DNA microarray analysis.** *Genome Biol* 2001, **2:**.
18. Kim J, Chung HJ, Park CH, Park WY, Kim JH: **ChromoViz: multimodal visualization of gene expression data onto chromosomes using scalable vector graphics.** *Bioinformatics* 2004, **20:**1191-1192.
19. Midorikawa Y, Tsutsumi S, Nishimura K, Kamimura N, Kano M, Sakamoto H, Makuuchi M, Aburatani H: **Distinct chromosomal bias of gene expression signatures in the progression of hepatocellular carcinoma.** *Cancer Res* 2004, **64:**7263-7270.
20. Husing J, Zeschingk M, Boes T, Jockel KH: **Combining DNA expression with positional information to detect functional silencing of chromsomal regions.** *Bioinformatics* 2003, **19:**2335-2342.
21. Awad IA, Rees CA, Hernandez-Boussard T, Ball CA, Sherlock G: **Caryoscope: an Open Source Java application for viewing microarray data in a genomic context.** *BMC Bioinformatics* 2004, **5:**151.
22. Fridlyand J, Snijders AM, Pinkel D, Albertson DG, Jain AN: **Hidden Markov models approach to the analysis of CGH data.** *JMVA* 2004, **90:**132-153.
23. Kovacs G, Akhtar M, Beckwith BJ, Bugert P, Cooper CS, Delahunt B, Eble JN, Fleming S, Ljungberg B, Medeiros LJ, Moch H, Reuter VE, Ritz E, Roos G, Schmidt D, Srigley JR, Storkel S, van den Berg E, Zbar B: **The Heidelberg classification of renal cell tumours.** *J Pathol* 1997, **183:**131-133.
24. Verdorf I, Hobisch A, Hittmair A, Duba HC, Bartsch G, Utermann G, Erdel M: **Cytogenetic characterization of 22 human renal cell tumors in relation to a histopathological classification.** *Cancer Genet Cytogenet* 1999, **111:**61-70.
25. Consortium TGO: **Gene Ontology: tool for the unification of biology.** *Nat Genet* 2000, **25:**25-29.
26. Beissbart T, Speed TP: **GOstat: Find statistically overrepresented Gene Ontologies within a group of genes.** *Bioinformatics* 2004, **20:**1464-1465.
27. Furge KA, Lucas KA, Takahashi M, Sugimura J, Kort EJ, Kanayama HO, Kagawa S, Hoekstra P, Curry J, Yang XJ, Teh BT: **Robust classification of renal cell carcinoma based on gene expression data and predicted cytogenetic profiles.** *Cancer Res* 2004, **64:**4117-4121.
28. Brantley-Sieders DM, Chen J: **Eph receptor tyrosine kinases in angiogenesis: from development to disease.** *Angiogenesis* 2004, **7:**17-28.
29. Clamp M, Andrews D, Barker D, Bevan P, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Hubbard T, Kasprzyk A, Keefe D, Lehvaslaiho H, Iyer V, Melsopp C, Mongin E, Pettett R, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Birney E: **Ensembl 2002: accommodating comparative genomics.** *Nucleic Acids Res* 2003, **31:**38-42.
30. Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA: **Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers.** *Cell* 2004, **118:**555-566.
31. Semple C: **Deep genomics in shallow times: the finished sequence of human chromosome 13 and 19.** *European Journal of Human Genetics* 2004, **12:**875-876.
32. Chen X, Cheung ST, So S, Fan ST, Barry C, Higgins J, Lai K, Dudoit S, Ng I, vandeRijn M, Bostein D, Brown PO: **Gene expression patterns in human liver cancer.** *Mol Bio Cell* 2002, **13:**1929-1939.
33. Higgins JP, Shinghal R, Gill H, Reese JH, Terris M, Cohen RJ, Fero M, Pollack JR, Van De Rijn M, Brooks JD: **Gene expression patterns**

in renal cell carcinoma assessed by complementary DNA microarray. *Am J Pathol* 2003, **162:**925-932.
34. Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese JC, Dwight SS, Kaloper M, Weng S, Jin H, Ball CA, Eisen MB, Spellman PT, Brown PO, Botstein D, Cherry JM: **The Stanford Microarray Database.** *Nucleic Acids Res* 2001, **29:**152-155.
35. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Res* 2002, **30:**e15..
36. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5:**R80.
37. Veltman JA, Fridlyand J, Pejavar S, Olshen AB, Korkola JE, DeVries S, Carroll P, Kuo WL, Pinkel D, Albertson D, Cordon-Cardo-Carlos, Jain AN, Waldman FM: **Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors.** *Cancer Res* 2003, **63:**2872-2880.