AD_____

Award Number: W81XWH-04-1-0045

TITLE: Identification of Prostate Cancer-Related Genes Using Inhibition of NMD in Prostate Cancer Cell Lines

PRINCIPAL INVESTIGATOR: Yurij Ionov Ph.D.

CONTRACTING ORGANIZATION: Health Research, Incorporated
Buffalo, New York 14263

REPORT DATE: January 2007

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-01-2007	Final	1 Jan 2004 – 31 Dec 2006

4. TITLE AND SUBTITLE

Identification of Prostate Cancer-Related Genes Using Inhibition of NMD in Prostate Cancer Cell Lines

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0045

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Yurij Ionov Ph.D.

E-Mail: yurij.ionov@roswellpark.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Health Research, Incorporated
Buffalo, New York 14263

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A strategy to identify mutant genes using inhibition of nonsense mediated decay (NMD) in cell lines has been proposed by others. Blocking translation with antibiotic emetine has been shown to inhibit the NMD. Stabilization of mutant mRNA following the inhibition of NMD with emetine can be detected using microarray technology, such as Affymetrix genechips, for example. Unfortunately, too many genes that do not contain any mutations show mRNA increase following emetine treatment due to stress response to the inhibition of translation or due to being a natural substrate for NMD, thus complicating the identification of mutant genes. We have developed a modification of the method which includes inhibition of NMD using two alternative methods. First is to inhibit translation with emetine. Second is to block NMD by inhibiting SMG-1 kinase with caffeine and than to block transcription with actinomycin D and either to continue blocking NMD with caffeine or to activate NMD by caffeine withdrawal. Analyzing the mRNA accumulation following emetine treatment as well as mRNA degradation following blocking of transcription in the presence or absence of NMD with the Affymetrix analysis using simple analytical algorithm allows selection of candidate genes for sequencing with high efficiency. Using our method of NMD inhibition we have identified several genes containing bi-allelic inactivating mutation in prostate cancer cell lines.

15. SUBJECT TERMS
No subject terms provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	30	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

INTRODUCTION

It is widely accepted that cancer is a genetic disease and that inactivation of tumor suppressor genes as well as activation of oncogenes is the causes of cancer. Finding genes responsible for the development of cancer has a diagnostic and prognostic significance and would help to create drugs against the disease. Although for some cancer types a number of genes and pathways for carcinogenesis have been already identified, the mechanisms of prostate cancer development remain poorly understood. Loss of heterozygosity (LOH) studies indicate that there are many yet unknown prostate cancer related tumor suppressor genes. As the title of the review by Ostrander EA and Stanford JL. in the American Journal of Human genetics (2000 pp 1367-75) states: Genetics of Prostate Cancer: Too Many Loci, Too Few Genes. The goal of this grant proposal is using a novel gene identification strategy, called GINI analysis, to identify genes that contain inactivating mutations in both alleles or contain mutation in one and have lost the other allele in prostate cancer cell lines. Such genes, if also mutated in primary tumours, can be putative prostate tumor suppressor genes. The idea of GINI is based on the phenomenon of Nonsense Mediated mRNA Decay (NMD), which means a fast degradation of mRNA transcripts containing premature translation termination codons (PTC) more than 50 bp upstream of the last exon/exon junction. Blocking translation with emetine has been shown to result in inhibition of NMD and accumulation of mutant transcripts which can be detected using microarray technology, such as Affymetrix genechips, for example. Unfortunately, too many genes that do not contain any mutations show mRNA increase following emetine treatment due to stress response to the inhibition of translation or due to being natural substrate for NMD, which complicates the identification of mutant genes. The goal of this grant proposal was first to develop a strategy to distinguish the mutant mRNA stabilization due to inhibition of NMD from the mRNA increase of non-mutated genes due to stress response to emetine and second to identify mutant genes in prostate cancer cell lines using the inhibition of NMD strategy.

This is the final report for this grant covering the 36 month period from January 2004 till December 2006. Although funding provided for this work allowed to hire a postdoctoral fellow for only 2 years, the tasks of the proposal have been completed.

BODY

Specific Aim1. Identification of the candidate genes for sequencing analysis in prostate cancer cell lines.

Task1. Application of different methods of NMD inhibition to select candidate genes for sequencing analysis.

The low efficiency of finding mutant genes using GINI analysis results from the high number of false positives, which are non-mutated genes showing mRNA increase following NMD inhibition. Inhibition of NMD by blocking protein translation with emetine results in increased amount of mRNA for approximately 10% of all genes of the genome We suggested that genes induced by stress responses to different treatments used to inhibit NMD can be different; therefore genes induced by one treatment but not by the other can be considered as false positives even if they induced only in the analyzed but not in the control cell line. Candidates for sequencing than should be the genes that show mRNA increase following all methods of NMD inhibition in tested but in control cell lines.

To evaluate different approaches to inhibit NMD for the efficiency of identifying mutant genes using GINI we used a ‘cancer chip’- a cDNA microarray with representatives for 5,000 genes known to be related in some way to cancer and the RKO colon cancer cell line, where two previously identified heterozygous frameshift mutations in P300 gene result in nonsense mediated decay of P300 mRNA. We used as a control HT29 colon cancer cells assuming that mutant genes in different cell lines are different, while genes that show mRNA increase due to stress response are common. We tried several approaches to inhibit NMD assuming that different methods of NMD inhibition produce different false positives created by stress response to the treatment and that selecting only those genes that show mRNA increase following each method of NMD

inhibition would maximize the chances to find mutant genes. These included the inhibition of NMD with emetine, the expression of dominant negative form of the major regulator of NMD hUpf1 as well as the downregulation of hUpf1 expression using siRNA.

A stable expression of the dominant negative form of hUpf1 gene did not produce any changes in gene expression in RKO cells. We attempted to express the dominant negative form of hUpf1 gene in the adenoviral vector but we did not have a success in the production of a viral stock which could efficiently overexpress the gene in the infected cells.

a) Comparing emetine and siRNA approach

When we tested the inhibition of the expression of the endogenous hUpf1 using siRNA we found that this approach did not work efficiently for all cell lines. However, since this approach finally did work for the RKO cells we were able to compare the list of genes that show mRNA increase produced by emetine treatment in RKO but not in the control HT29 cells with the list of genes induced by siRNA against the hUpf1 in RKO cells. We found that the lists of candidate genes for sequencing selected by analysing mRNA changes produced by both approach essentially overlap. Table1 shows the first 30 genes that had the highest mRNA increase following emetine treatment in RKO cells and had less than 2-fold increase in control HT29 cells detected by hybridization to cDNA microarray. Table 1 also shows genes that in addition to be induced by emetine in RKO but not in HT29 cells are also induced by siRNA in against hUpf1 in RKO cells. The list of overlapping genes contained P300 genes indicating that the approach of analyzing mRNA changes produced by two alternative methods of NMD inhibition can be used for identifying mutant genes. Indeed, sequencing the first 10 genes from the list of genes that show mRNA increase following both methods of NMD inhibition has identified except previously identified in P300 gene bi-allelic frameshift mutations in RHAMM and MSH3 genes in RKO cells. However, the remaining sequenced genes did not contain the mutations implying that the efficiency of finding mutations using combination of emetine and siRNA is still low and that combining siRNA and emetine treatments did not improved significantly the efficiency of GINI. Besides, siRNA approach in contrast to pharmacological inhibition of NMD does not work equally well for all cell lines. In LS180 colon cancer cells for example the siRNA treatment did not produce significant downregulation of hUpf1 expression. Therefore, although it was not in the statement of work we also tested alternative to emetine pharmacological methods to inhibit NMD.

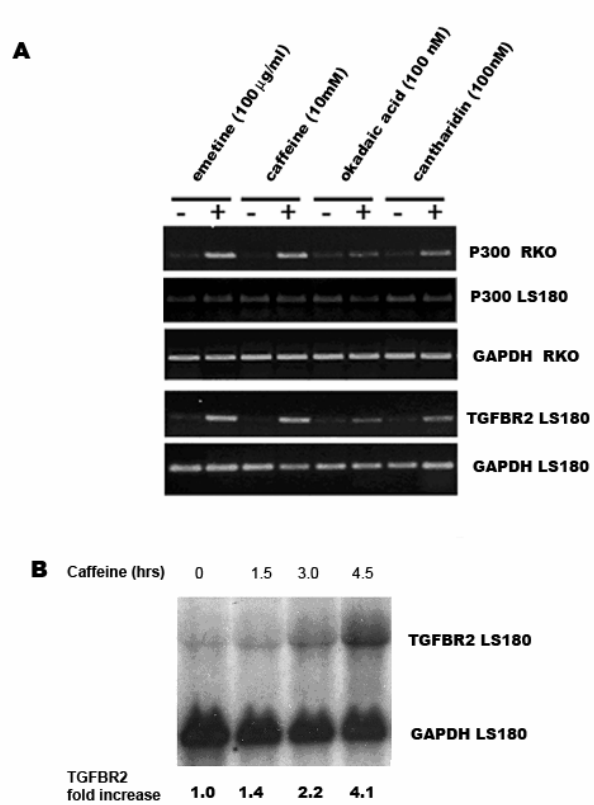

Figure 1.

Figure 1. Accumulation of mutant P300 and TGFBR2 mRNAs in RKO and LS180 cells respectively following different approaches to inhibit NMD. A. RT-PCR analysis demonstrates that similar to incubation with emetine the incubation with 10 mM caffeine, the inhibitor of the SMG-1 kinase as well as with 100nM okadaic acid or 100 nM cantharidin, the inhibitors of PP2A phosphatase result in increased amounts of mutant (RKO cells) but not wild type(LS180 cells) P300 mRNA as well as mutant TGFBR2 mRNA (LS180 cells). Total RNAs for the RT-PCR analysis were prepared following 3 hrs incubation with indicated drugs. **B.** Quantitative analysis of Northern blot using a PhosphorImager (Molecular Dynamics) demonstrates significant accumulation of mutant TGFBR2 mRNA in LS180 cells following 4.5 hrs of incubation with 10mM caffeine.

b) Interfering with hUpf1 phosphorylation can be used for identifying mutant genes.

It has been shown that sequential phosphorylation and dephoshorylation of hUpf1 protein by PI3-like kinase hSMG-1 and PP2A phosphatase respectively is required for the initiation of the NMD mediated mRNA degradation (Brumbaugh et al., 2004; Ohnishi et al., 2003). Moreover, a pharmacological block of NMD using inhibition of hSMG-1 phosphorylation with caffeine or wortmannin has been previously reported (Usuki et al., 2004). To test the possibility of using pharmacological interference with hUpf1 phosphorylation/dephosphorylation for genomewide identifying mutant genes we analyzed the effect of inhibitors of hSMG-1 kinase or PP2A phosphatase on the stability of mutant P300 gene in RKO cells. RT-PCR analysis on Figure 1 shows that incubation of cells with caffeine, the inhibitor of hSMG-1 kinase, as well as with okadaic acid or cantharidin, the inhibitors of PP2A phosphatase, like the incubation of cells with emetine result in increased amount of mutant but not wild type P300 and TGFBR2 mRNAs analyzed in RKO and LS180 cells respectively (Ionov et al., 2004). Since caffeine produced minimal cell death in colon cancer cells as compared with other reagents we further used caffeine as an alternative to emetine. Northern Blot on Figure 1B demonstrates the significant increase of mutant TGFBR2 mRNA following 4.5 hrs of incubation of LS180 cells with caffeine.

c) Blocking transcription following NMD inhibition helps to minimize number of false positives.

In our previous work in order to eliminate the false positives produced by differential de novo mRNA synthesis due to stress response to emetine we blocked transcription together with blocking NMD (Ionov et al., 2004). Despite successful identification of two mutant genes blocking transcription together with NMD caused decreased amounts of mRNA which resulted in the absence of signal for the halve of the probes presented on the microarray. To avoid this problem we tested a modified approach which uses blocking transcription not

simultaneously with but after NMD inhibition. The rational behind the modified approach is that blocking transcription after the accumulation of mutant transcripts following NMD inhibition will enable to detect different speed of mutant mRNA degradation when NMD is blocked or released in the absence of de novo transcription. When we blocked transcription with actinomycin D four hours after blocking NMD with emetine we could not, however, detect any significant difference in the mRNA levels of known mutant genes. It is possible that if some 'short-living' protein is required for the NMD to work, than blocking translation with emetine results in the absence of this protein and inability to activate NMD following inhibition of transcription. Since caffeine does not block translation we tested the effect of blocking NMD using caffeine on the mRNA degradation of mutant transcripts in the absence of de novo transcription. Northern blot on the Figure 2 shows that pretreatment with caffeine but not with emetine allows to detect significant difference in the amount of mutant TGFBR2 mRNA in LS180 cells following blocking transcription in the context of released or blocked NMD. This implies that pretreatment of cells with caffeine will increase the efficiency of detection of the differences in the amount of mutant mRNA after NMD is blocked or released when the transcription is inhibited with actinomycin D.

Using 'cancer chip' we compared mRNA expression profiles in RKO cells after pretreatment with caffeine for 4 hrs followed by incubation in the presence of actinomycin D plus caffeine or actinomycin D only. We will further call this version of GINI as C(A+C)/CA, where C stands for caffeine and A for actinomycin D. We selected genes that after incubation with caffeine and following incubation with actinomicine D and caffeine had more than two-fold mRNA amounts than after incubation with caffeine following incubation with actinomycin D only. Six genes, which include P300, RHAMM and MSH3 (Table 1) have been selected as candidates for sequencing when we combined the results of inhibiting NMD using C(C+A)/CA treatment with that of emetine treatment. The remaining three genes did not contain any mutations. These false positives could be caused by microarray hybridization errors. Despite this the 50% efficiency of identifying mutant genes using combination of two different approaches to inhibit NMD suggests that combining the analysis of mRNA accumulation following emetine treatment with the analysis of mRNA degradation during C(C+A)/CA treatment can be the optimal strategy to identify mutated genes in prostate cancer cell lines when using Affymetrix HGU133Plus2.0 platform. The details of the procedure are described in the appended manuscript published in oncogene (Ivanov et al., 2006).

d) Using a simple analytical algorithm to analyse mRNA changes produced by emetine treatment for identification of mutant genes.

In addition we have developed a simple analytical algorithm to process the data of the Affymetrix Genechip hybridization showing mRNA profile alteration following emetine treatment. The details of the algorithm are in the appended manuscript published in Cancer Genetics and Cytogenetics in 2005. Briefly, the algorithm is based on the assumption that different genes are mutated in different prostate cancer cell lines due to the heterogeneity of prostate cancer. The method utilizes comparing mRNA profiles of two cancer cell lines before emetine treatment and also comparing between cell lines mRNA changes produced by emetine treatment. Genes selected for sequencing analysis should satisfy the following criteria: 1) the gene should have low expression in one cell lines and the high expression in the other, 2) the genes should show a high mRNA increase following emetine treatment only in cell line with low expression before treatment.

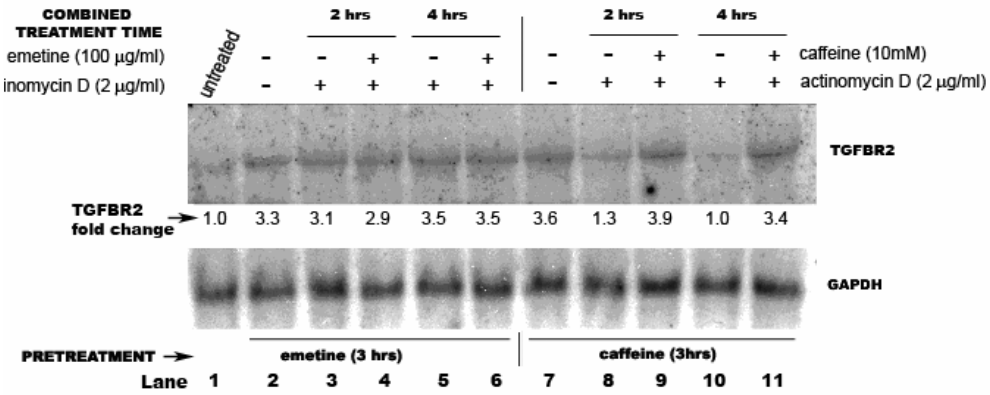

Figure 2. Blocking mRNA synthesis with actinomycin D following inhibition of NMD with caffeine but not with emetine helps to detect NMD dependent degradation of mutant TGFBR2 mRNA in LS180 cells. Northern blot demonstrates that incubation with emetine (100 µg/ml) as well as with caffeine (10mM) for 3 hrs result in increased amounts of mutant TGFBR2 mRNA in LS180 cells (lanes 2 and 7) as compared with untreated cells (lane 1). Blocking transcription with actinomycin D following emetine treatment results in similar amounts of TGFBR2 mRNA when NMD is inhibited (lanes 4 and 6) or not (lanes 3 and 5) with emetine. Blocking transcription with actinomycin D following caffeine treatment results in significantly different

Specific aim 2. Identification of mutations in genes selected from the Affymetrix Genechip.

Task 2. Mutation analysis of the genes selected by using NMD inhibition and Affymetrix Genechip.

To identify genes mutated in prostate cancer cell lines we applied the strategy developed working on Task 1 of the proposal. We treated LNCaP, DU145, 22Rv-1 and PC3 cells with emetine, caffeine as well as with caffeine followed by combined caffeine and actinomycin D or actinomycin D only treatments. Using our analytical algorithm that takes into consideration not only the changes in mRNA expression profiles produced by different treatments but comparing the gene expression levels in untreated cells we selected candidates for sequencing analysis out of more than 50,000 genes represented on the Affymetrix U133Plus2.0 GeneChip. Sequencing has identified bi-allelic inactivating mutations in 10 genes in the cell lines analyzed shown in Table 1. Figure 3 illustrates the example of identified mutations in the PARD3 gene in LNCaP cells.

Table 1. Genes mutated in prostate cancer cell lines identified using GINI

Gene	Gene symbol	Mutation	Cell line with gene mutated	Function
Cleft Lip and Palate Transmembrane Protein 1	CLPTM1	DelC in $(C)_2$ nt 858-9	LNCaP	RNA binding
myristoylated alanine-rich protein kinase C substrate	MARCKS	DelA in $(A)_{11}$ nt 454-64	DU145	cell motility
Janus Kinase 1	JAK1	Del A in (A)8 nt 2573-80, DelC in (C)7 nt 1283-89	LNCaP	Signal transduction

synaptojanin 2	SYNJ2	DelC in $(C)_8$ nt 3331-8	LNCaP	RNA binding phosphoinositide 5-phosphatase activity
mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	MLH1*	C39X (Huusko et al., 2004)	DU145	DNA mismatch repair
androgen-induced proliferation inhibitor	APRIN	DelA in $(A)_9$ nt 4161-9, DelC in $(C)_7$ nt 2164-70	22Rv-1	negative regulation of cell proliferation
par-3 partitioning defective 3 homolog (C. elegans)	PARD3	DelT in (T)5 nt 950-4, C>T nt 1921	LNCaP	establishment and/or maintenance of cell polarity
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	SMARCA4*	85nt deletion in exon 10 (Wong et al., 2000)	Du145	regulation of transcription
PHD finger protein 3	PHF3	DelA in $(A)_7$ nt 3402-8 nt 449-59	DU145	regulation of transcription
transforming growth factor, beta receptor II (70/80kDa)	TGFBR2	DelA in (A)10 nt 1141-8	22rv-1	protein serine/threonine kinase signaling pathway

* These mutations have been independently reported by other groups

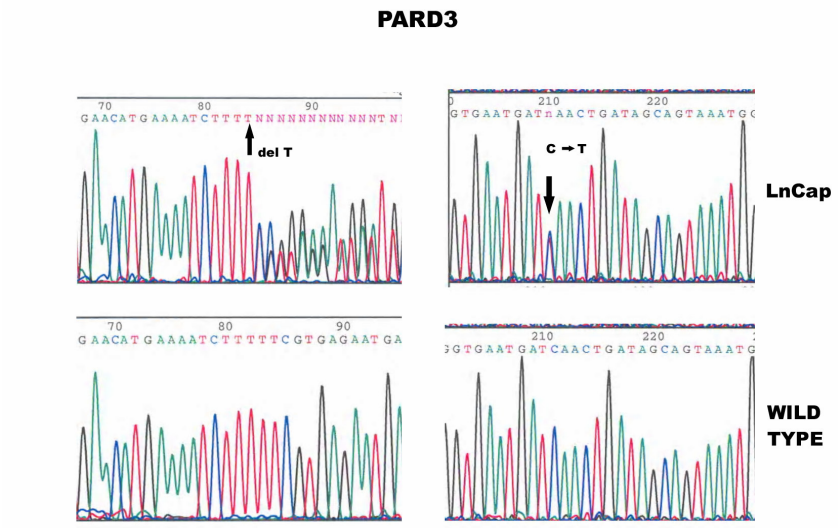

Figure 3. *Sequencing chromatograms show 2 heterozygous inactivating mutations in PARD3 gene in LnCAP cells.*

An additional work performed besides the scope of the proposal

Several of the identified inactivated mutant genes can be related to the development of prostate cancer. Functional studies are required to prove the tumor suppressor activities of the genes mutated in prostate cancer cell lines. Unfortunately, the specific aim related to the functional studies had been cut from my original proposal. However, we performed an additional work to assess the significance of the identified mutations for prostate cancer. We sequenced the Pard 3 gene exons in 17 DNA samples of prostate tumor DNA. No mutations have been identified. This implies that the *pard3* gene mutation is a rare event in prostate carcinogenesis. However, this does not preclude the importance of the pathway containing the *pard3* gene for

prostate cancer. This gene is implicated in asymmetrical cell division suggesting that inactivation of pard3 pathway could affect the replication of prostate stem cells.

We compared the mutation spectra between 22Rv1 prostate cancer cell line and its parental CWR22 xenograft.(22Rv1 is a human prostate carcinoma epithelial cell line derived from a xenograft that was serially propagated in mice after castration-induced regression and relapse of the parental, androgen-dependent CWR22 xenograft). We have found that mutations in *AS3, JAK1, TGFBR2* genes identified in 22Rv1 cell lines were also present in parental CWR22 cells. This implies that mutations identified using NMD inhibition occurred not during in vitro cultivation of 22Rv1 cells but most likely during in vivo tumor development.

In order to identify novel genes involved in prostate tumor progression we used GINI analysis to identify additional mutations in the C-4-2 cells, the metastatic derivative of the LNCaP cell line. We hypothesized that progression to metastatic stage in C-4-4 cells could be driven through the acquiring the additional mutations in genes involved in metastatic development in LnCAP cells. Although the task was simplified by the existence of the perfect isogenic control to identify easily nonspecific mRNA increases due to stress response, no additional mutations have been identified in the metastatic derivative of the LnCAP cells. This implies that despite high mutation rate mediated by the inactivation of mismatch repair system in LnCAP cells the progression to metastatic stage in this model system was driven most likely by epigenetic rather than genetic alterations.

SUMMARY STATUS OF TASKS OUTLINED IN THE STATEMENT OF WORK

Task1. Completed.

Task2. Completed.

KEY RESEARCH ACCOMPLISHMENTS

1) An efficient method of selection of the candidate genes for sequencing analysis from the genes that show mRNA increase following inhibition of NMD has been developed.

2) An alternative version of GINI analysis which uses analysis of NMD mediated mRNA degradation in the absence of transcription following inhibition of NMD with caffeine has been developed.

3) Ten genes containing inactivating biallelic mutations in prostate cancer cell lines have been identified using our version of GINI and analytical algorithm to select genes for sequencing analysis.

REPORTABLE OUTCOMES

Publications

M. Rossi, L. Hawthorn, J. Platt, T. Burkhardt, J. Cowell, Y. Ionov. Identification of inactivating mutations in the JAK1, SYNJ2, and CLPTM1 genes in prostate cancer cells using inhibition of nonsense-mediated decay and microarray analysis. Cancer Genetics and Cytogenetics 2005, September 161(2):97-103.

I Ivanov, K C Lo, L Hawthorn, J K Cowell and Y Ionov. Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. Oncogene, 2006 November 6, 1-12, Epub ahead of print.

A third paper on identifying PARD3 and AS3 gene mutations using GINI in prostate cancer cells is currently being prepared.

CONCLUSIONS

Through the funded period we have developed a reliable method to identify mutant genes in prostate cancer cell lines using inhibition of NMD and Affymetrix GeneChip analysis. The efficiency of the method is close to 50%, which is the highest reported for the identifying mutant genes using GINI analysis. The approach which includes combining inhibition of NMD and comparative genomic hybridization, described by other group (Huusko et all (2004) is much less efficient than the one developed in our lab. Analyzing several cell lines they were able to identify mutation only in one gene, EPHB2 in one cell line. Ten genes have been identified to contain biallelic inactivating mutations in prostate cancer cell lines using our approach. The reason for this is that in prostate cancer cell lines with microsatellite instability, such as in LNCaP, DU145 and 22Rv1 inactivation of both alleles by two independent mutations is more likely than mutation in one allele and loss of the other. Therefore the addition of CGH analysis to GINI provides little advantage. Several genes previously not known to be mutated, such as CLPTM1, AS3, PARD3, are potential prostate tumor suppressor genes. I am planning to apply for additional funding to study functional significance of mutations for prostate tumor development.

References

Brumbaugh, K.M., Otterness, D.M., Geisen, C., Oliveira, V., Brognard, J., Li, X., Lejeune, F., Tibbetts, R.S., Maquat, L.E. & Abraham, R.T. (2004). *Mol Cell*, **14,** 585-98.

Huusko, P., Ponciano-Jackson, D., Wolf, M., Kiefer, J.A., Azorsa, D.O., Tuzmen, S., Weaver, D., Robbins, C., Moses, T., Allinen, M., Hautaniemi, S., Chen, Y., Elkahloun, A., Basik, M., Bova, G.S., Bubendorf, L., Lugli, A., Sauter, G., Schleutker, J., Ozcelik, H., Elowe, S., Pawson, T., Trent, J.M., Carpten, J.D., Kallioniemi, O.P. & Mousses, S. (2004). *Nat Genet*, **36,** 979-83.

Ionov, Y., Nowak, N., Perucho, M., Markowitz, S. & Cowell, J.K. (2004). *Oncogene*, **23,** 639-45.

Ivanov, I., Lo, K.C., Hawthorn, L., Cowell, J.K. & Ionov, Y. (2006). *Oncogene*.

Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P. & Ohno, S. (2003). *Mol Cell*, **12,** 1187-200.

Usuki, F., Yamashita, A., Higuchi, I., Ohnishi, T., Shiraishi, T., Osame, M. & Ohno, S. (2004). *Ann Neurol*, **55,** 740-4.

Wong, A.K., Shanahan, F., Chen, Y., Lian, L., Ha, P., Hendricks, K., Ghaffari, S., Iliev, D., Penn, B., Woodland, A.M., Smith, R., Salada, G., Carillo, A., Laity, K., Gupte, J., Swedlund, B., Tavtigian, S.V., Teng, D.H. & Lees, E. (2000). *Cancer Res*, **60,** 6171-7.

ELSEVIER

Cancer Genetics and Cytogenetics 161 (2005) 97–103

Cancer Genetics and Cytogenetics

Identification of inactivating mutations in the *JAK1*, *SYNJ2*, and *CLPTM1* genes in prostate cancer cells using inhibition of nonsense-mediated decay and microarray analysis

Michael R. Rossi, Lesleyann Hawthorn, Julie Platt, Tania Burkhardt, John K. Cowell, Yurij Ionov*

Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263

Received 24 November 2004; received in revised form 1 February 2005; accepted 2 February 2005

Abstract We have developed a simple analytical method that increases the efficiency of identifying mutant genes in cell lines after the inhibition of nonsense-mediated decay (NMD). The approach assumes that the spectra of mutant genes differ between cell lines of the same tumor origin. Thus, by analyzing more than one cell line in parallel and taking into account not only changes in mRNA levels after the inhibition of NMD, but also comparing mRNA levels between cell lines before the inhibition of NMD, the vast majority of false positives were eliminated from the analysis. In this study, we used Affymetrix oligonucleotide arrays to compare mRNA profiles of two prostate cancer cell lines, PC3 and LNCaP, before and after emetine treatment. As a result of our modified approach, from the 14,500 genes present on the array, 7 were identified as candidates from LNCaP cells and 1 was identified from PC3 cells. Sequence analysis of five of these candidate genes identified gene-inactivating mutations in four of them. Homozygous mutations were found in the synaptojanin 2 (*SYNJ2*) and the cleft lip and palate *CLPTM1* genes. Two different heterozygous mutations in the Janus kinase 1 (*JAK1*) gene result in complete loss of the protein in several different prostate cancer cell lines.

1. Introduction

Nonsense mutations that occur in coding DNA sequences upstream of the last exon of a gene elicit a rapid degradation of mutant mRNA through the nonsense-mediated mRNA decay (NMD) pathway [1,2]. Since initiating the degradation of mutant mRNA requires an initial round of translation, blocking translation with specific drugs, such as emetine, has been shown to abrogate the NMD-mediated degradation of mutant mRNAs [3]. This effectively results in an increased amount of mutant cellular mRNA from genes containing nonsense or frameshift mutations, which can be detected using expression microarrays. A strategy for using microarray analysis of mRNA profile alterations resulting from inhibiting NMD in cell lines (GINI) has been proposed for the identification of genes containing nonsense mutations [4]. Unfortunately, inhibition of NMD by blocking translation with emetine is accompanied by the upregulation of a large number of genes that do not contain nonsense mutations due to a stress response to the drug, which complicates choosing candidates for sequence analysis. One approach to minimize the number of false positives in the pool of the candidate genes selected for sequence analysis was to combine inhibition of NMD with a block of transcription using actinomycin D. This approach largely prevents de novo synthesis of mRNA during treatment, thereby reducing the upregulation of stress response genes. Selecting only genes that show mRNA increases after treatment with both emetine and actinomycin D streamlines the mutation analysis. Although this approach has already proven successful in identifying inactivating mutations in colorectal [5] and prostate [6] cancer cell lines, the analysis is compromised by the overall reduction in mRNA transcripts, which results in loss of hybridization signals for almost half of the genes present on the microarray. Thus, genes that may be mutated but are expressed at moderate to low levels are excluded from detection. Although the number of false-positive candidate genes could be minimized by using nonmalignant control cell lines to identify those involved in the stress response, the differences in response to drug treatments between normal

* Corresponding author. Tel.: 716-845-5935; fax: 716-845-1698.
E-mail address: Yurij.Ionov@RoswellPark.org (Y. Ionov).

0165-4608/05/$ – see front matter
doi:10.1016/j.cancergencyto.2005.02.006

and cancer cells [7] still produces a large number of false positives.

Using cancer cell lines, we have now developed and tested a simple algorithm to more accurately identify candidate genes for sequence analysis from the large number of genes that show mRNA increases after inhibition of NMD with emetine.

We used PC3 and LNCaP, both of which reportedly have microsatellite instability [8], to maximize the possibility of identifying mutant genes because of their inability to repair replication defects in the coding mononucleotide repeats [9,10]. In this study four of the five candidate genes identified demonstrated either homozygous or heterozygous mutations in a range of prostate cancer cell lines. These findings provide a valuable proof of principle that our analytical algorithm can streamline the GINI analysis of cancer cells and identifies novel genes that appear to be related to prostate cancer development and progression.

2. Materials and methods

2.1. Cell culture and emetine treatment

Prostate cancer cell lines LNCaP and PC3 were maintained in RPMI 1640 supplemented with 10% fetal bovine serum and antibiotics. The emetine treatment protocol was as described previously [5].

2.2. Oligonucleotide array analysis

mRNA levels in emetine-treated and untreated cells were measured using the Affymetrix U133A array. The protocol for this analysis has been outlined [11]. GeneChip expression array analysis was performed using Affymetrix Microarray Suite software (GeneChip Analysis Suite). This software uses two independent sets of algorithms, a quantitative algorithm (robust estimator of the mean difference in probe intensities) that computes raw signal and signal log ratio (SLR) directly from the hybridization intensities of probes, as well as a confidence algorithm (nonparametric) that provides *P* values that measure the confidence in detection (absent or present) or change (increase, decrease, or no change) of a specific target.

2.3. Mutation analysis

One microgram of total RNA from emetine-treated LNCaP and PC3 cells was reverse transcribed using the SuperScript II protocol (Invitrogen, Carlsbad, CA). Three overlapping polymerase chain reaction (PCR) primer sets were used to generate products spanning the entire *JAK1* open reading frame for both cell types. Genomic DNA isolated from 6 prostate cell lines, 4 colon cell lines, and 10 primary colon tumors were also PCR amplified using the following *JAK1* exon primers: Ex4F 5′-TTCATTTTCCTGCCTTCCAG-3′, Ex4R 5′-CCACAAACTCCAGCTTCTCC-3′, Ex8F 5′-CTGAAGCTCTCTTCCCACGA-3′, Ex8R 5′-CTAAAACACGGGCTCTCTGC-3′. The primers that amplify the mononucleotide repeat within exon 23 of *SYNJ2* in genomic DNA were *SYNJ2* Ex23F 5′-GCC TCC TGT GCTCAGATCC-3′ and *SYNJ2* Ex23R 5′-GGAGCCGTGTTTTTCAGTAGC-3′. The primers to amplify the cDNA containing two short mononucleotide repeats $(C)_5$ of the *CLPTM1* gene were *CLPTM1* Ex8-11F 5′-TCACCATCAACATCGTGGAC-3′ and *CLPTM1* Ex. 8-11R 5′-CGAGGACTCGATATACGTGGA-3′.

PCR analyses for both cDNA and genomic DNA were performed using the Phusion high-fidelity DNA polymerase protocol (Finnzymes OY, Espoo, Finland). The PCR products were gel purified and sequenced using the PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA).

2.4. Protein analysis

Western blotting was carried out as described previously [5] using a polyclonal *JAK1* antibody from Cell Signaling Technology (Beverly, MA). The second horseradish peroxidase–anti-rabbit antibody was obtained from Santa Cruz Biotechnology (Santa Cruz, CA).

3. Results

3.1. Expression array analysis

To identify candidate genes related to prostate cancer, we analyzed the mRNA levels between the LNCaP and PC3 prostate cancer cell lines before and after treatment with the NMD-inhibiting drug, emetine. First, we excluded genes that were not expressed in either cell line regardless of treatment, as well as genes that were induced by emetine treatment in both cell lines. In the latter case we assumed that the majority of these mRNA increases were most likely caused by a cellular stress response related to emetine rather than to inhibition of NMD (see below). Next, we rationalized that if we assume different genes will carry nonsense mutations in different cell lines, then in untreated cells, some genes (those with mutations) would show lower mRNA levels in one cell line (A) and higher mRNA levels (those without mutation) in the other cell lines (B). At the same time, only the genes that showed low levels of mRNA should then show increased levels after treatment with emetine. Since it was not possible to predict exactly how a nonsense mutation would affect the RNA levels for any particular gene, we established several arbitrary cut-offs for these mRNA level changes to maximize the chances of identifying mutated genes. Thus, for comparing mRNA levels in untreated cells we used the formula $B/A \geqslant fourfold$, and in the treated cells $A^{treated}/A \geqslant fourfold$, while for the same gene $B^{treated}/B < twofold$. Using Affymetrix analysis software (http://www.affymetrix.com/support/technical/manual/expression_manual.affx), fold changes are represented by the SLR, and a fourfold increase in the absolute signal intensity of

emetine-treated cells versus untreated cells is equivalent to an SLR of 2.

More than 800 of the 22,000 probes sets, or approximately 3.5% of the total number of genes (14,500) represented on the U133A platform, had increased levels of expression after emetine treatment in both LNCaP and PC3 cells. A number of these genes, including TXNIP and GADD45B (data not shown), are known stress response genes and were excluded from further analysis. Using the approach described above, we were able to narrow our search down to seven candidate genes in LNCaP cells (Table 1) as potential carriers of nonsense mutations. *RIN2, JAK1, NRP1, CLPTM1, SYNJ2, DATF1*, and *RANBP2* each had an SLR $\geq$ 2 ($A^{treated}/A \geq$ *fourfold*) for LNCaP cells and an SLR $<$ 1 for PC3 cells ($B^{treated}/B <$ *twofold*). There was a significant increase (I) in the comparison between emetine-treated and untreated LNCaP cells, and the "*P* detection" values were less than 0.00007 for all these genes. Although *RIN2, JAK1, DATF1*, and *RANBP2* appear to have significant increases ($P <$ 0.004) in expression in PC3 cells after emetine treatment, these increases are below the SLR $<$ 1 threshold and are therefore assumed to be within the range of normal variability.

The same approach was used to identify candidate genes in PC3 cells (Table 1). In this analysis we identified *p53* as the only candidate mutated gene ($P = 0.003$). In fact, there have been a number of reports already demonstrating that PC3 cells carry inactivating truncating mutations in *p53*, whereas LNCaP cells express wild-type *p53* [12,13]. During our analysis, we also identified genes whose mRNA were significantly increased after emetine treatment in both PC3 and LNCaP cells. While these could represent mutant genes in both cell lines, they are equally likely to be stress response genes induced in both cells. As such, we have not pursued these genes further at this time.

3.2. Mutation analysis of candidate genes

From the genes shown in Table 1, *SYNJ2, CLPTM1*, and *JAK1* were chosen as candidates for mutation analysis because they contain mononucleotide repeat regions within the open reading frame of the genes. The death-associated transcription factor-1 gene (*DATF1*) was also chosen for mutation analysis because its predicted function implicated it as a potential tumor suppressor gene. A homozygous deletion of one "C" nucleotide in a $(C)_8$ mononucleotide repeat in exon 23 (Fig. 1A) was identified in the *SYNJ2* gene in LNCaP cells. PC3 cells did not show this mutation. Extending the mutation analysis to other prostate cancer cell lines identified a heterozygous mutation in the same region in exon 23 in LAPC-4 cells. We also sequenced exon 23 in the SW48, HCT116, RKO, LoVo, and LS180 colon cancer cell lines, which are known to have microsatellite instability (MSI). A heterozygous deletion of one C in the same mononucleotide repeat was found in LS180 and LoVo cells.

Table 1
Candidate genes for mutation analysis in LNCaP and PC3 cells

				LNCaP				PC3			
Gene symbol	Gene title	Probe set ID	Cyto	Untreated	Emetine	SLR	*P*	Untreated	Emetine	SLR	*P*
RIN2	Ras and Rab interactor 2	209684_at	20p11.22	207 (A)	1022 (P)	3.2 (I)	0.000027	3776 (P)	5003 (P)	0.3 (I)	0.004073
JAK1	Janus kinase 1	201648_at	1p31.3	477 (P)	4125 (P)	2.8 (I)	0.00002	9086 (P)	11753 (P)	0.5 (I)	0.00002
NRP1	Neuropilin 1	212298_at	10p11.2	543 (P)	4479 (P)	2.6 (I)	0.00002	10577 (P)	9798 (P)	−0.2 (NC)	0.942324
CLPTM1	Cleft lip and palate transmembrane protein 1	211136_s_at	19q13.2	338 (P)	2066 (P)	2.4 (I)	0.00002	4083 (P)	4839 (P)	0.1 (NC)	0.5
SYNJ2	Synaptojanin 2	212828_at	6q25.3	340 (A)	1554 (P)	2.2 (I)	0.000068	6414 (P)	5296 (P)	−0.4 (D)	0.999448
DATF1	death associated transcription factor 1	218325_s_at	20q13.33	498 (P)	2256 (P)	2.1 (I)	0.00002	2169 (P)	3165 (P)	0.7 (I)	0.00002
RANBP2	RAN binding protein 2	201711_x_at	2q12.3	517 (P)	2173 (P)	2.1 (I)	0.00002	4063 (P)	6082 (P)	0.5 (I)	0.00006
TP53	Tumor protein p53	201746_at	17p13.1	13656 (P)	12041 (P)	−0.2 (NC)	0.846768	298.6 (A)	2275 (P)	2.8 (I)	0.000241

LNCaP and PC3 cells treated with emetine were analyzed using the Affymetrix GeneChip Human Genome U133A Array. Eight candidate genes met our gene expression criteria (i.e. $A^{treated}/A \geq 4$ fold; $B^{treated}/B < 2$ fold). With the exception of *TP53*, all of the genes have a significant increase and SLR $\geq$ 2 in LNCaP cells following emetine treatment, whereas in PC3 cells, gene expression did not exceed the SLR $<$ 1 threshold following emetine treatment. The signal log$_2$ ratio (SLR) is a comparison between signal intensities for emetine treated and untreated cells and the absolute signal intensities with their corresponding absent (A) or present (P) calls are shown to give context to the SLR values. The change in signal intensities for a specific probe between emetine treated and untreated cells are shown as either increased (I), decreased (D) or not changed (NC). The P-values for these calls are also provided for assigning statistical significance.

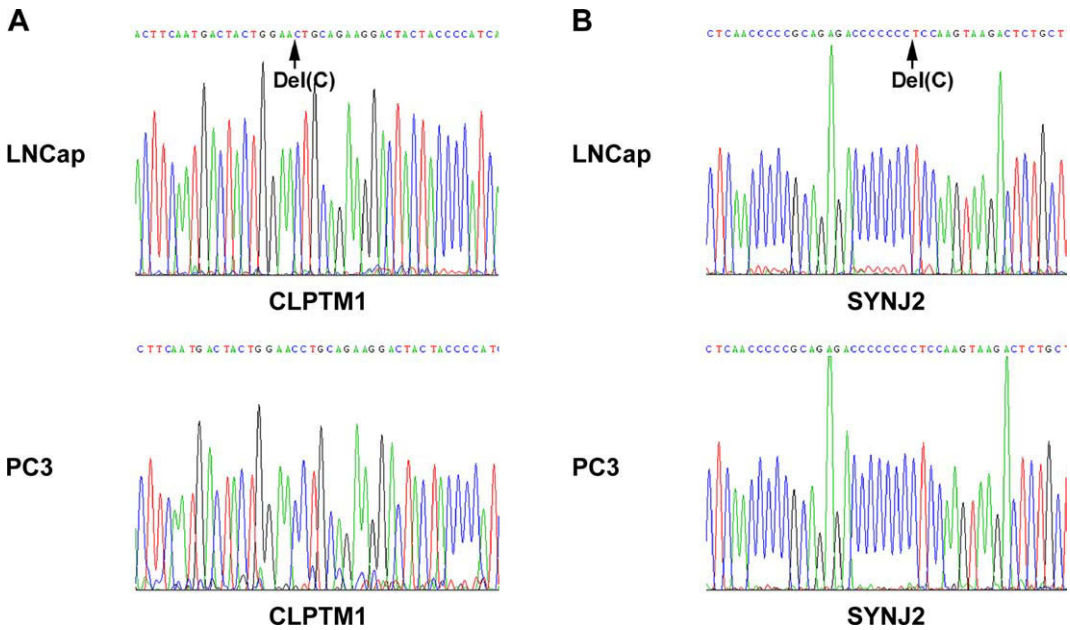

Fig. 1. Mutation analysis in prostate cancer cells. Disruption of NMD suggested that the *CLPTM1* (A) and *SYNJ2* (B) genes carried nonsense mutations in LNCaP cells. Sequence analysis identifies homozygous deletions within mononucleotide repeats in the coding regions of both of these genes.

The *CLPTM1* gene contains two $(C)_5$ repeats in its open reading frame, but sequencing of the entire cDNA in LNCaP cells identified a homozygous deletion of the 857C nucleotide that lies outside the $(C)_5$ repeats (Fig. 1B). No mutations in the *CLPTM1* gene were found in the PC3 cells. Sequencing the cDNA for the *DATF1* gene from emetine-treated LNCaP cells did not identify any mutations.

Sequence analysis of the *JAK1* cDNA from the emetine-treated LNCaP cells identified two different heterozygous mutations, one in exon 4, which contains a $(A)_8$ mononucleotide repeat (Fig. 2A), and the other in exon 8, which contains a $(C)_7$ tract (Fig. 2B). Both are predicted to be inactivating mutations, indicating a loss of function for *JAK1* in these cells (see below). PC3 carried only the wild-type sequence (data not shown). The *JAK1* mutations were confirmed by sequence analysis of the same exons from genomic DNA. Sequence analysis of exon-specific PCR products from all 24 exons of *JAK1* did not identify mutations in PC3 as expected. Next, we extended the analysis of *JAK1* to additional prostate cancer cell lines. DU145 cells did not carry mutations in *JAK1*, but 22Rv-1 and C42 (a derivative of LNCaP) cells both contained heterozygous frameshift mutations of a single A in exon 4 and a single C in exon 8, which were identical to those seen in LNCaP cells. LAPC-4 cells also carried a heterozygous mutation in exon 8, but in this case the frameshift mutation was the insertion of a C nucleotide. In addition to prostate cancer cell lines, we examined genomic DNA from 4 MSI (+) colon cell lines as well as 10 primary colon tumors with MSI. No mutations were found in *JAK1* in either the colon cell lines or in the colon tumors (data not shown). These results suggest that mutation of *JAK1* is more related to the malignant phenotype in prostate cancer cells rather than a being a general consequence of MSI.

3.3. JAK1 *expression in prostate cancer cells*

Western blot analysis was performed to investigate the presence or absence of the *JAK1* protein in prostate cancer cells. As shown in Fig. 3, *JAK1* is highly expressed in DU145 cells and present at detectable levels in PC3 cells. Neither of these cell lines carries *JAK1* mutations. The *JAK1* protein was not detectable in 22Rv-1, LNCaP, or C42 cells, and there was only minimal expression in LAPC-4. These data are consistent with the mutation analysis that predicted bi-allelic inactivating mutations in 22Rv-1, LNCaP, and C42, as well as heterozygous inactivation in LAPC-4.

4. Discussion

In our previous study [5] we demonstrated the usefulness of manipulating the NMD pathway for the identification of potential tumor suppressor genes. The major problem with using emetine to block NMD, however, was the concomitant upregulation of stress response genes which, despite using actinomycin D to prevent de novo transcription, still led to the identification of a large number of candidate genes, many of which were not found to be mutated. In the present study we designed and tested an algorithm for selecting candidate genes for sequence analysis from among the hundreds of genes that show increased mRNA levels after microarray analysis. Using this approach, we greatly reduced the number of genes that invited sequence analysis and, of the

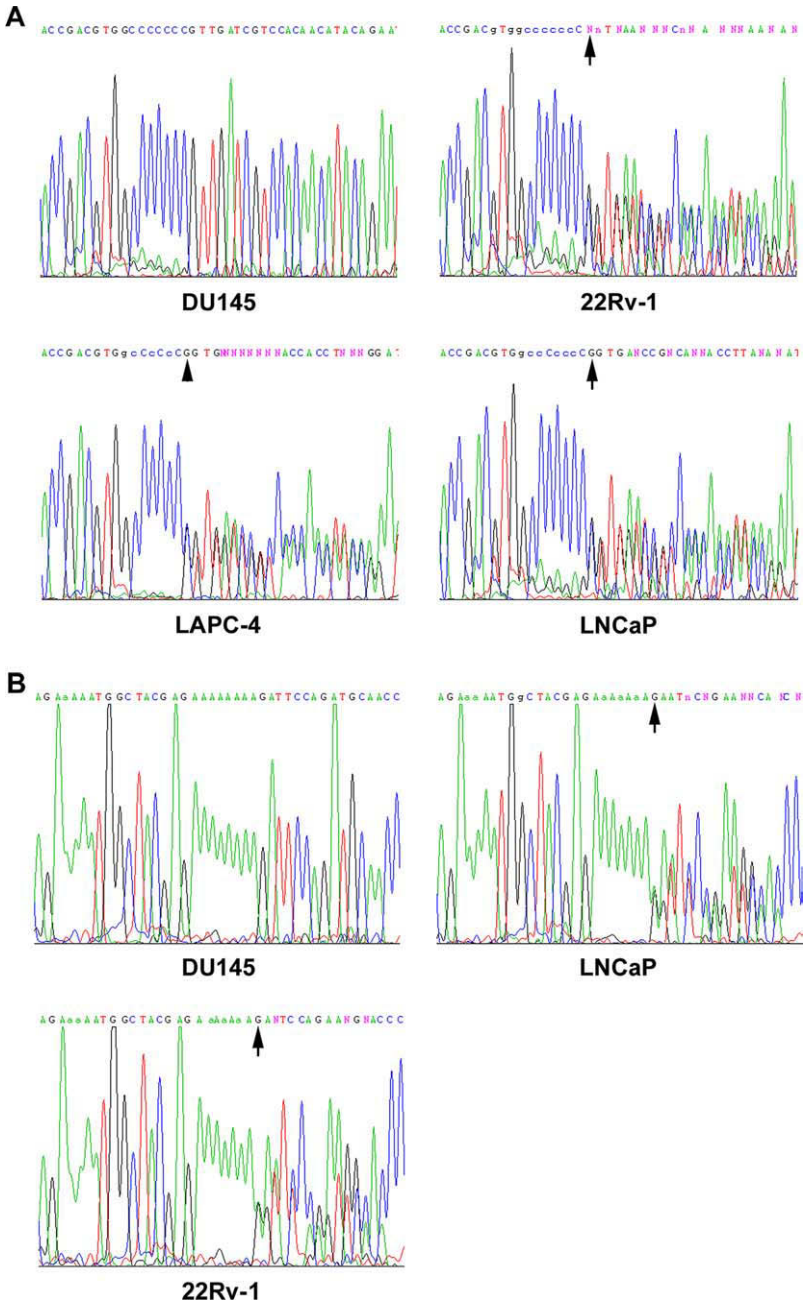

Fig. 2. JAK1 gene mutations in prostate cancer cells. Genomic DNA isolated from LNCaP, PC3, 22Rv-1, DU145, LAPC-4, and C42 cells were used to confirm gene mutations in the two exons of *JAK1* containing microsatellite regions. (A) DU145 is representative of the wild-type chromatograph for exon 8, which contains a $(C)_7$ tract. Heterozygous frameshift mutations (arrows) were observed in both 22Rv-1 and LNCaP cells, and an insertion of a single C (arrowhead) was identified in one allele of LAPC-4. (B) Heterozygous frameshift mutations (arrows) were also observed in exon 4 in 22Rv-1 and LNCaP cells. Loss of a single A in a region containing $(A)_8$ was seen in alleles of 22Rv-1 and LNCaP when compared to DU145 cells containing wild-type *JAK1*.

five genes selected (including *p53*), truncating mutations were identified in four of them. This success rate greatly improves the usefulness of the GINI approach to identify candidate tumor suppressor genes. It is possible that the criteria we established to filter the data were too stringent, so we may have missed some genes that carry truncating mutations in the LNCaP and PC3 cells. It is therefore possible that lowering the stringency of the filters will result in the identification of a higher number of candidate mutant genes, and such analyses are currently underway. It is also possible that we have missed genes that are mutated in both LNCaP and PC3 cells because they would have the same profile as stress response genes. By including additional cell lines in this analysis in the future, it may be possible to identify these mutant genes because it is unlikely that the same genes will be mutated in all cell lines tested.

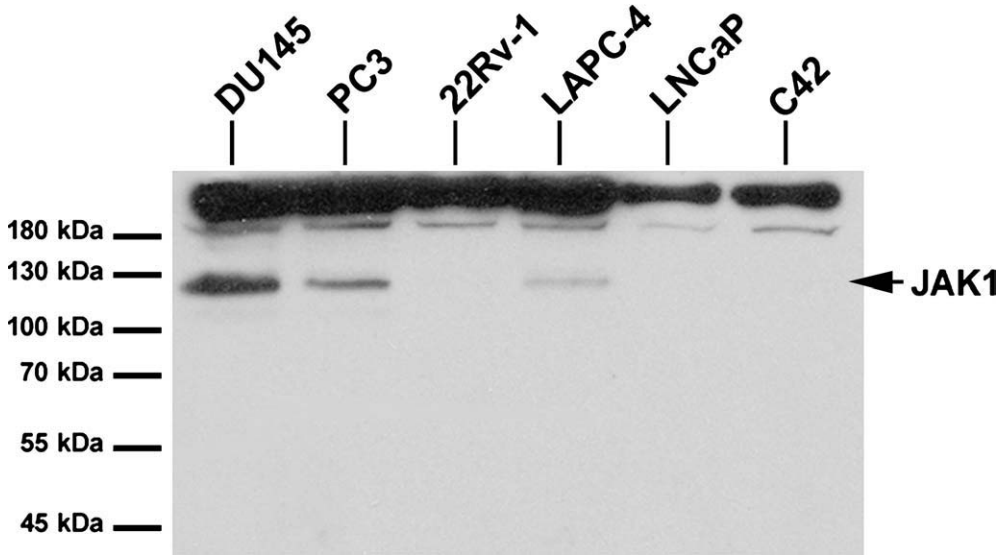

Fig. 3. *JAK1* protein in prostate cancer cells. Western blot analysis using 50 μg of total cell lysates for each of the prostate cancer cell lines shown demonstrated the presence of the 130-kD *JAK1* protein in DU145 and PC3 cells. There was a faint band for LAPC-4 cells but the *JAK1* protein was below detectable levels in 22Rv-1, LNCaP, and C42 cells. The nonspecific, approximately 250-kD band acts as a protein-loading control in this analysis.

Surprisingly, the number of candidate tumor suppressor genes identified in the LNCaP cells was much higher than in PC3 cells. One possible explanation for this is the MSI status of these cells. While the MSI+ status of the LNCaP cells has been proven both by our identification of frameshift mutations in microsatellite repeats in *JAK1* and *TGF beta receptor type II* genes (data not shown), as well as the reported mutation in the *MSH2* mismatch repair gene [14], the MSI status in the PC3 cells has been controversial [15]. If, in fact, PC3 cells do not have MSI, then it may not be surprising that the number of frameshift mutations identified using GINI analysis in these cells might be less than in LNCaP cells. Another possibility is that the mutation in the *p53* gene in PC3 cells could negate the need for mutations in other weaker tumor suppressor genes.

In this study, we identified inactivating mutations in the *JAK1* gene, which appear to be specific to prostate cancer cells rather than being a general consequence of microsatellite instability because MSI+ colon cancer cells did not show mutations in this gene. There are at least four members of the Janus kinase family of tyrosine kinases (*TYK2*, *JAK1*, *JAK2*, and *JAK3*) that are, with the exception of *JAK3*, expressed in a variety of different cell types [16]. The role of *JAK1* in interleukin signaling, as well as its role in malignant cell growth and survival, is well documented [17]. *JAK* kinases are involved in a number of signaling pathways that mediate malignant transformation through activating *STAT*, *Bcl-2*, *PI3*, and *Src* kinases, as well as serving to inhibit *p53*-dependent cell cycle arrest and apoptosis [16–18]. There is considerable evidence that *JAK1* is generally activated in cancer cells and contributes to driving tumor cell growth in both leukemias and solid tumors [16,17]. As a result, cell growth can be suppressed by treatment with the *STAT3/JAK2* inhibitor tyrphostin AG490 in prostate cancer cells [19,20]. It is unclear, therefore, why *JAK1* would be inactivated in a subset of prostate cell lines as opposed to being overexpressed. At this time, we can only speculate that *JAK1* inactivation confers some selective advantage for certain prostate cancer cells, possibly by allowing for *JAK*-independent signaling. A more detailed examination of primary prostatic tumors will be necessary to determine if *JAK1* mutations have clinical consequences. It is also of note that the cell lines that are null for *JAK1* expression provide a valuable resource for studying the function of this gene.

We have now been able to extend our previous studies [5] to demonstrate that the manipulation of NMD in conjunction with expression array analysis is a powerful tool for identifying novel gene mutations in cancer cells. We have also demonstrated that comparing the consequences of emetine treatment between cell lines from the same tissue provides a more streamlined approach toward the identification of mutant genes because it reduces the background of candidate genes that result from stress responses. We are currently extending our sequence analysis of other candidate genes described in this report to determine their involvement in prostate cancer and the effectiveness of our data analysis approach to define genes related to the malignant phenotype.

Acknowledgments

This work was supported by the Prostate Program of the Department of Defense (grant PC030076), an ACS Institutional Review grant, and in part by the National Cancer Institute Roswell Park Cancer Center Support (grant CA 16056).

References

[1] Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. Nonsense-mediated decay approaches the clinic. Nat Genet 2004;36:801–8.

[2] Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 2004;5:89–99.

[3] Ishigaki Y, Li X, Serin G, Maquat LE. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell 2001; 106:607–17.

[4] Noensie EN, Dietz HC. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. Nat Biotechnol 2001;19:434–9.

[5] Ionov Y, Nowak N, Perucho M, Markowitz S, Cowell JK. Manipulation of nonsense mediated decay identifies gene mutations in colon cancer cells with microsatellite instability. Oncogene 2004;23:639–45.

[6] Huusko P, Ponciano-Jackson D, Wolf M, Kiefer JA, Azorsa DO, Tuzmen S, Weaver D, Robbins C, Moses T, Allinen M, Hautaniemi S, Chen Y, Elkahloun A, Basik M, Bova GS, Bubendorf L, Lugli A, Sauter G, Schleutker J, Ozcelik H, Elowe S, Pawson T, Trent JM, Carpten JD, Kallioniemi OP, Mousses S. Nonsense-mediated decay microarray analysis identifies mutations of EPHB2 in human prostate cancer. Nat Genet 2004;36:979–83.

[7] Scotto KW. Transcriptional regulation of ABC drug transporters. Oncogene 2003;22:7496–511.

[8] Chen Y, Wang J, Fraig MM, Henderson K, Bissada NK, Watson DK, Schweinfest CW. Alterations in PMS2, MSH2 and MLH1 expression in human prostate cancer. Int J Oncol 2003;22:1033–43.

[9] Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 1995;268:1336–8.

[10] Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, Perucho M. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science 1997;275:967–9.

[11] Kunapuli P, Kasyapa CS, Hawthorn L, Cowell JK. LGI1, a putative tumor metastasis suppressor gene, controls in vitro invasiveness and expression of matrix metalloproteinases in glioma cells through the ERK1/2 pathway. J Biol Chem 2004;279:23151–7.

[12] Isaacs WB, Carter BS, Ewing CM. Wild-type p53 suppresses growth of human prostate cancer cells containing mutant p53 alleles. Cancer Res 1991;51:4716–20.

[13] van Bokhoven A, Varella-Garcia M, Korch C, Johannes WU, Smith EE, Miller HL, Nordeen SK, Miller GJ, Lucia MS. Spectral karyotype (SKY) analysis of human prostate carcinoma cell lines. Prostate 2003;57:205–25.

[14] Leach FS, Velasco A, Hsieh JT, Sagalowsky AI, McConnell JD. The mismatch repair gene hMSH2 is mutated in the prostate cancer cell line LNCaP. J Urol 2000;164:1830–3.

[15] Karan D, Schmied BM, Dave BJ, Wittel UA, Lin MF, Batra SK. Decreased androgen-responsive growth of human prostate cancer is associated with increased genetic alterations. Clin Cancer Res 2001; 34:72–80.

[16] Verma A, Kambhampati S, Parmar S, Platanias LC. Jak family of kinases in cancer. Cancer Metastasis Rev 2003;22:423–34.

[17] Leonard WJ, O'Shea JJ. Jaks and STATs: biological implications. Annu Rev Immunol 1998;16:293–322.

[18] Quelle FW, Wang J, Feng J, Wang D, Cleveland JL, Ihle JN, Zambetti GP. Cytokine rescue of p53-dependent apoptosis and cell cycle arrest is mediated by distinct Jak kinase signaling pathways. Genes Dev 1998;12: 1099–107.

[19] Jacobberger JW, Sramkoski RM, Wormsley SB, Bolton WE. Estimation of kinetic cell-cycle-related gene expression in G1 and G2 phases from immunofluorescence flow cytometry data. Cytometry 1999;35: 284–9.

[20] Ni Z, Lou W, Leman ES, Gao AC. Inhibition of constitutively activated Stat3 signaling pathway suppresses growth of prostate cancer cells. Cancer Res 2000;60:1225–8.

Oncogene (2006), 1–12

www.nature.com/onc

ORIGINAL ARTICLE

Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells

I Ivanov, KC Lo, L Hawthorn, JK Cowell and Y Ionov

Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY, USA

Inhibition of the nonsense-mediated decay (NMD) mechanism in cells results in stabilization of transcripts carrying premature translation termination codons. A strategy referred to as gene identification by NMD inhibition (GINI) has been proposed to identify genes carrying nonsense mutations. Genes containing frameshift mutations in colon cancer cell line have been identified using a modified version of GINI. To increase the efficiency of identifying mutant genes using GINI, we have now further improved the strategy. In this approach, inhibition of NMD with emetine is complemented with inhibiting NMD by blocking the phosphorylation of the hUpf1 protein with caffeine. In addition, to enhance the GINI strategy, comparing mRNA level alterations produced by inhibiting transcription alone or inhibiting transcription together with NMD following caffeine pretreatment were used for the efficient identification of false positives produced as a result of stress response to NMD inhibition. To demonstrate the improved efficiency of this approach, we analysed colon cancer cell lines showing microsatellite instability. Bi-allelic inactivating mutations were found in the *FXR1*, *SEC31L1*, *NCOR1*, *BAT3*, *PHF14*, *ZNF294*, *C19ORF5* genes as well as genes coding for proteins with yet unknown functions.
Oncogene advance online publication, 6 November 2006; doi:10.1038/sj.onc.1210098

Keywords: nonsense-mediated decay; mutations; caffeine; GINI

Introduction

It is widely accepted that the transformation of colon epithelial cells results from the activation of oncogenes as well as the inactivation of tumor suppressor genes controlling cell proliferation and survival (Vogelstein and Kinzler, 2004). In colon cancer cell lines with microsatellite instability (MSI) (Aaltonen *et al*., 1993; Ionov *et al*., 1993; Thibodeau *et al*., 1993), the high overall mutation frequency caused by the inactivation of DNA mismatch repair (MMR) genes (Fishel *et al*., 1993; Leach *et al*., 1993) results in the mutational inactivation of tumor suppressor genes. Consequently, these cell lines provide a valuable tool for identifying colon cancer related tumor suppressor genes. According to the two-hit hypothesis of tumorigenesis (Knudson, 1971), inactivating mutations affecting both alleles of a gene suggest a tumor suppressor function for the mutated gene.

Even though the genes mutated in colon tumors with MSI are often different from those mutated in MSI-negative tumors, the tumorigenic pathways affected by these mutations are often the same. For example, the absence of p53 mutations in colon cancers with MSI is often accompanied by mutations in the *BAX* (Rampino *et al*., 1997) or *P300* genes (Ionov *et al*., 2004a), both of which are involved in the p53 signaling pathway. Another example involves the transforming growth factor (TGF)β signaling pathway. In contrast to MSI-positive tumors, colon cancers without MSI rarely have inactivating mutations in the *TGFβ* receptor type II gene, although they frequently show mutations in *SMAD4*, a downstream effector of TGFβ signaling (Salovaara *et al*., 2002). This observation implies that identifying genes containing bi-allelic inactivating mutations in colon cancer cell lines with MSI will identify potential tumor suppressor genes or tumorigenic pathways related to malignant transformation. This observation is reinforced if the expression of these genes is also lost in primary tumor samples.

Gene inactivating events frequently result from nonsense or frameshift mutations, which create premature translation termination codons (PTC). Such mutations frequently elicit a rapid degradation of the mutant mRNA through the nonsense-mediated mRNA decay (NMD) pathway (Holbrook *et al*., 2004; Maquat, 2005; Weischenfeldt *et al*., 2005). NMD is an mRNA surveillance pathway, which prevents the production of truncated proteins that result mainly from the products of aberrant splicing. NMD is activated when multi-protein exon junction complexes (EJC) assemble on the sites of exon/exon junctions that are not removed from the spliced mRNA after the pioneer round of translation (Ishigaki *et al*., 2001). These EJCs are removed from the spliced mRNA during the movement of translated mRNA through the ribosome. When the translation

Correspondence: Dr Y Ionov, Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA.
E-mail: Yurij.Ionov@RoswellPark.org
Received 15 June 2006; revised 21 August 2006; accepted 27 September 2006

machinery stalls on a PTC located more than ~50 nucleotides upstream of the last exon/exon junction, the unremoved EJCs downstream of the PTC recruit the Upf protein complex, which initiates the degradation of the PTC-containing mRNA transcripts (Holbrook *et al*., 2004).

Inhibition of translation has been shown to inhibit NMD and result in the accumulation of PTC-containing mRNA transcripts (Ishigaki *et al*., 2001). The increase in mRNA levels produced by inhibition of NMD can be detected using gene expression microarrays (Noensie and Dietz, 2001). The strategy of gene identification by NMD inhibition (GINI) provides an opportunity for the genome-wide identification of genes containing inactivating mutations (Noensie and Dietz, 2001). Using cDNA microarrays, Noensie and Dietz (2001) have demonstrated that blocking translation with emetine, for example, results in an elevated mRNA content for genes known to contain bi-allelic inactivating mutations. Although several genes mutated in colon and prostate cancer cell lines have already been identified (Huusko *et al*., 2004; Ionov *et al*., 2004b; Rossi *et al*., 2005), the efficiency of identifying mutations using the GINI strategy is still low, because too many genes that do not contain mutations also show mRNA increases following inhibition of NMD as a result of a stress response. This complication often makes it difficult to select candidate genes for sequencing analysis.

To improve the efficiency of identifying mutant genes using GINI, we evaluated different methods of NMD inhibition in this study. As a result, we have developed a modified version of GINI analysis (termed GINI2) that is more efficient in distinguishing those genes containing PTC from false positives that show mRNA increases due to a stress response to NMD inhibition. When we applied this GINI2 version to the analysis of colon cancer cell lines, a significantly increased number of genes containing bi-allelic inactivating mutations were identified. Some of these genes were also frequently mutated in primary colon tumors.

Results

Inhibition of NMD with caffeine can be used to detect NMD-mediated degradation of PTC-containing mRNA transcripts

We have shown previously (Ionov *et al*., 2004b) that increases of mRNA levels following inhibition of NMD can result from increased transcription of genes owing to a stress response to emetine. To overcome this problem, we developed a strategy to detect stabilization of mutant mRNA by combining the inhibition of NMD with the inhibition of transcription using actinomycin D. Although this approach has proved successful in identifying inactivating mutations in colorectal cancer cell lines (Ionov *et al*., 2004b), the analysis was compromised by the overall reduction in the levels of mRNA transcripts because of actinomycin D treatment. As a result, detectable hybridization signal for ~50% of the genes present on the microarray was lost. Thus, genes that might have been mutated, but which were expressed at only moderate to low levels, were excluded from detection.

To overcome this limitation, we have modified our method of simultaneous inhibition of NMD and transcription (Ionov *et al*., 2004b). An additional step has been introduced into our original protocol (Ionov *et al*., 2004b), which involves inhibiting NMD for several hours before blocking either transcription alone or transcription together with NMD. Inhibiting NMD before blocking both transcription and NMD should result in higher basal amounts of mutant mRNA, and consequently should result in stronger microarray hybridization signal for mutant genes. The predicted changes in mRNA levels for mutant and stress response genes following these different protocols of NMD inhibition are shown schematically in Figure 1.

This strategy was tested using the LS180 and RKO colon cancer cell lines, which we have previously shown to carry mutations in the *TGFBR2* and *P300* genes, respectively. These mRNAs are degraded by NMD (Ionov *et al*., 2004b) in these cell lines. Surprisingly, no significant difference in mRNA levels from the mutant genes were seen after incubation in the presence of actinomycin D, with or without emetine, following the initial pretreatment with emetine (data not shown). This was possibly due to the fact that emetine levels remain high within the cells, even after an extensive washing procedure. We, therefore, investigated using a different pharmacological inhibitor of NMD.

Sequential phosphorylation and dephosphorylation of the hUpf1 protein by the human suppressor with morphogenetic effect on genitalia-1 (hSMG-1) phosphatidylinositol 3-like kinase and protein phosphatase 2A (PP2A), respectively, is required for the initiation of NMD-induced mRNA degradation (Ohnishi *et al*., 2003; Brumbaugh *et al*., 2004). Moreover, a pharmacological block of NMD using caffeine or wortmannin, which inhibit hSMG-1 phosphorylation, has been reported previously (Usuki *et al*., 2004). We therefore analysed whether inhibition of either hSMG-1 kinase or PP2A phosphatase affects the stability of the mutant *P300* and *TGFBR2* genes in RKO and LS180 cells. Incubation with either caffeine (an inhibitor of hSMG-1 kinase), or okadaic acid or cantharidin (inhibitors of PP2A phosphatase) were shown using reverse transcription–polymerase chain reaction (RT–PCR) analysis (Figure 2a) to increase the amounts of the mutant *P300* and *TGFBR2* mRNAs in RKO and LS180 cells, respectively. Importantly, at the doses required to inhibit NMD, incubation with caffeine resulted in only minimal cell death, compared with okadaic acid and cantharidin. We therefore replaced emetine with caffeine in our GINI strategy. Northern blot analysis (Figure 2b) confirms that a significant increase of mutant *TGFBR2* mRNA was present following incubation of LS180 cells with caffeine for 4.5 h. Next, we tested whether pretreatment with caffeine would affect NMD activation following caffeine withdrawal. Northern blot analysis (Figure 3) demonstrates that pretreatment with caffeine, but not with emetine, results in significant

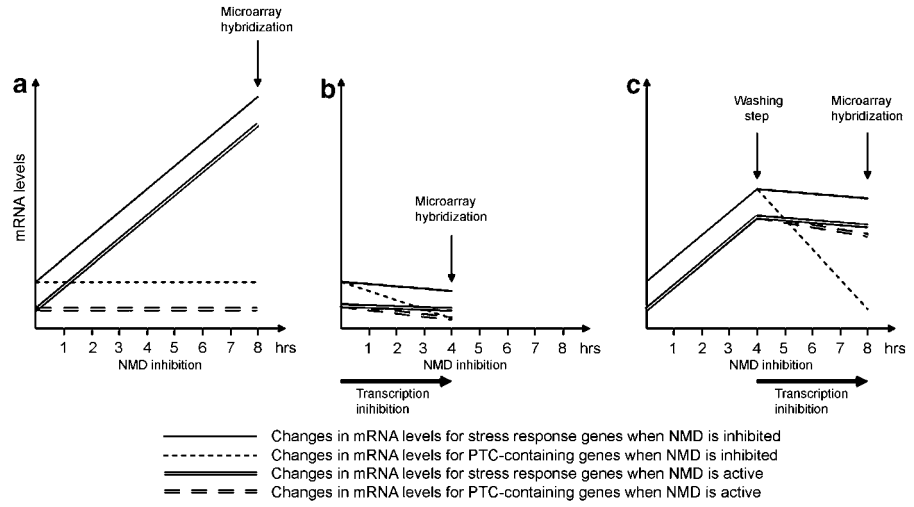

Figure 1 Theoretical depiction of changes in mRNA levels for stress response and PTC-containing genes following different protocols to inhibit NMD. (**a**) Continuous treatment of cells with an inhibitor of NMD results in increase of mRNA levels for PTC-containing genes, as well as for stress response genes. Microarray hybridization analysis cannot distinguish between the two causes of these increases: the enhanced mRNA synthesis due to stress response or the decreased mRNA degradation in case of PTC-containing genes. (**b**) When transcription is prevented using actinomycin D, even though NMD is blocked, the levels of PTC-containing transcripts and stress response genes are too low to be efficiently detected on microarrays. Changes in mRNA levels induced by NMD inhibition detected by hybridization analysis for many genes are within the 'hybridization noise', and are not reliable. (**c**) Blocking NMD for 4 h allows the accumulation of both PTC-containing transcripts and stress response gene transcripts to levels where they can be detected on the expression arrays. When transcription is blocked after this initial accumulation, the degradation of PTC-containing transcripts will occur quickly if the NMD blocking agent is removed, or will be delayed when NMD is sustained. The degradation of stress response transcripts after blocking transcription, on the other hand, does not depend on NMD. As a result, significant mRNA differences will be detected on the expression array for PTC-containing genes but not for the stress response genes.

differences in mutant *TGFBR2* mRNA levels in LS180 cells following either inhibition of transcription together with NMD, or inhibition of transcription only. This result illustrates that caffeine pretreatment can be used as an alternative to emetine to enhance the detectability of the microarray hybridization signals for mutant transcripts following inhibition of transcription with actinomycin D.

Identifying genes containing inactivating mutations in colon cancer cell lines using inhibition of NMD and Affymetrix genechip analysis

We used Affymetrix Genechip hybridization to measure the mRNA profile alterations produced by two alternative strategies of NMD inhibition, which are illustrated in Figure 1a and c. In the first method (Figure 1a), emetine was used to inhibit NMD as described in the original GINI protocol (Noensie and Dietz, 2001). The second method is a modification of our previously described (Ionov *et al*., 2004b) protocol for simultaneous inhibition of NMD and transcription, which is illustrated in Figure 1b. For convenience, we refer to the original emetine treatment as GINI1 and the modified caffeine protocol as GINI2. In the GINI2 protocol, caffeine was used to first inhibit NMD, thus allowing for an accumulation of mutant mRNAs in the cells (Figure 1c). The caffeine was then removed and the cells were treated further with either actinomycin D alone or actinomycin D and caffeine together. RNA from these experiments was then analysed using the U133Plus2.0 arrays.

It has been shown that the genes mutated in MSI-positive cells are generally different from those mutated in MSI-negative cells (Ionov *et al*., 1993; Konishi *et al*., 1996; Lengauer *et al*., 1998). Thus, to identify mutated genes in MSI (+) LS180 cells, we compared the GINI1 and GINI2 results from this cell line with those obtained using MSI (−) SW480 cells, assuming that different genes would be mutated in these two cell lines. Total RNA was isolated from both LS180 and SW480 cells treated according to either GINI1 or GINI2 protocols, as well as from untreated cells, and the mRNA levels were analysed using Affymetrix U133Plus2.0 GeneChip hybridization. The hybridization data were then compared between treated and untreated cells and the signal log ratios (SLRs), which equal Log_2 of fold changes of hybridization signal intensities produced by drug treatments, were calculated using the Affymetrix Microarray Suite version 5.0. The row data represented as .CEL files for GINI1 and GINI2 analyses for SW480 and LS180 cells have been submitted to GEO databases (accession number GSE5486).

Candidate genes for subsequent sequence analysis were selected according to three arbitrarily chosen parameters: (1) show a $SLR = Log_2$ (fold change) $\geqslant 1.7$ in LS180 cells and an $SLR \leqslant 1$ in SW480 following emetine treatment, (2) show $SLR \geqslant 1$ in LS180 cells following the GINI2 protocol and (3) show the intensity of hybridization signal for untreated LS180 cells lower

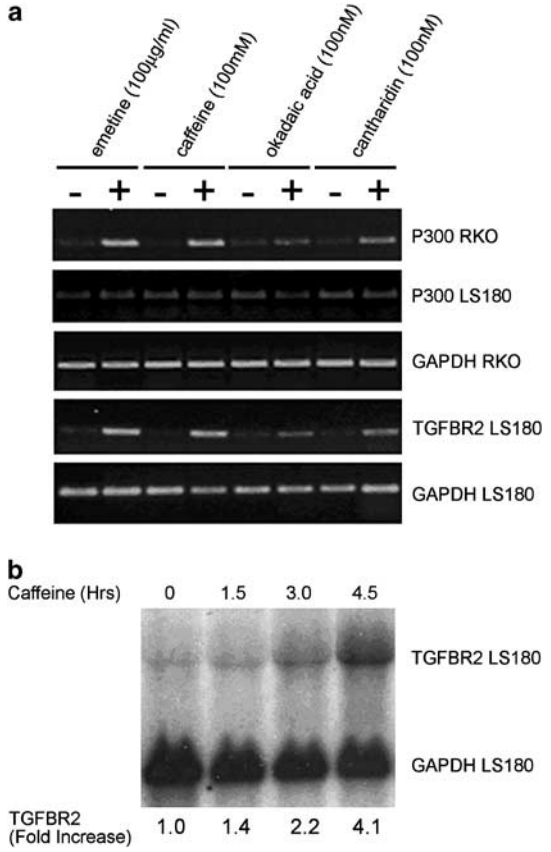

Figure 2 Accumulation of mutant *P300* and *TGFBR2* mRNAs in RKO and LS180 cells, respectively, following different approaches of NMD inhibition. (**a**) RT–PCR analysis demonstrates that, similar to incubation with emetine, incubation with 10 mM caffeine as well as with 100 nM okadaic acid or 100 nM cantharidin results in increased amounts of mutant (in RKO cells) but not wild-type (in LS180 cells) *P300* mRNA, as well as mutant *TGFBR2* mRNA (in LS180 cells). Total RNAs for the RT–PCR analysis were prepared following 3-h incubation with the indicated drugs. (**b**) Quantitative analysis of *TGFBR2* mRNA levels using Northern blotting. PhosphorImaging demonstrates significant accumulation of mutant *TGFBR2* mRNA in LS180 cells following 4.5-h incubation with 10 mM caffeine.

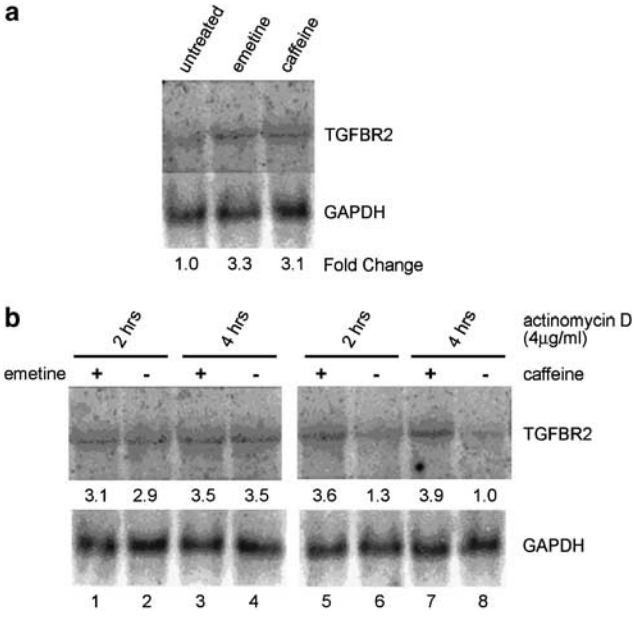

Figure 3 Blocking mRNA synthesis with actinomycin D, following inhibition of NMD with caffeine, but not with emetine enhances detection of NMD-dependent degradation of mutant *TGFBR2* mRNA in LS180 cells. (**a**) Northern blot analysis demonstrates that after 3-h incubation with either emetine (100 μg/ml) or caffeine (10 mM) increased amount of mutant *TGFBR2* mRNA in LS180 cells compared with untreated cells is observed. (**b**) (Left) Blocking transcription with actinomycin D following emetine pretreatment results in similar amounts of *TGFBR2* mRNA when NMD is either inhibited with emetine (lanes 2 and 4) or not (lanes 1 and 3). Blocking transcription with actinomycin D (Right) following caffeine treatment results in significantly different amounts of *TGFBR2* mRNA when NMD is either inhibited with caffeine (lanes 6 and 8) or not (lanes 5 and 7). Quantification (fold change) was performed using PhosphorImaging.

than that for the untreated SW480 cells. This follows as we expected that the levels of mutant transcripts in LS180 cells should be lower than corresponding wild-type levels in SW480 cells as a result of NMD degradation of mutant transcripts. This approach has proved effective in the past in our analysis of prostate cancer cell lines (Rossi *et al*., 2005). To further minimize the false discovery rate produced by hybridization 'noise', we excluded those genes which were scored as 'absent' by the Affymetrix analysis tool, either in the emetine treated LS180 cells or the untreated control SW480 cells. Of the approximately 55 000 probe sets representing approximately 20 000 unique genes on the Affymetrix U133Plus2 gene expression array, 15 candidate genes were identified (Table 1a) using these criteria. Sequencing the cDNA from these genes in LS180 cells has identified bi-allelic inactivating mutations in 7 genes (Table 1a).

To select candidates for sequencing we used crude differences in intensity to calculate the changes in transcript levels. As the number of false positives generated by this analysis was still relatively high (eight out of 15 candidates), we tested whether using additional statistical analysis could improve the efficiency of candidate selection. To do this, we used an adaptation of Robust Multi-Chip Average (RMA) (Irizarry *et al*., 2003), with specific correction for GC biases known as GC-RMA (Wu *et al*., 2004). When this statistical analysis was applied to the GINI1 and GINI2 row data, four genes from the original list were eliminated as well as one gene (CTNND1), which had been shown to contain bi-allelic mutations. This gene was located at the limit of the threshold cutoff, which supports the idea of reanalysing the data with more liberal thresholds. As a result of GC-RMA analysis, therefore, the efficiency of identifying mutant genes was greater than 50% (six mutant genes out of 10 candidates as seen in Table 1b). We also used GC-RMA analysis to test the efficiency of selecting candidate genes following GINI1 or GINI2 protocols separately. Neither GINI1 nor GINI2 alone achieve the same efficiency of analysis of combined GINI1 and GINI2 protocols. A possible explanation

Table 1a Candidate genes for sequencing in LS180 cells

Probe set ID	*Gene title*	*Gene symbol*	*LS180 emetine SLR (change)*	*LS180 C(A+C)/CA SLR (change)*	*LS180 untreated signal (detection)*	*SW480 untreated signal (detection)*	*SW480-emetine SLR (change)*
219121_s_at[a]	**RNA-binding motif protein 35A**	***RBM35A***	**3.6(I)**	**1.1(I)**	**143.8(A)**	**727.4(P)**	**0.8(I)**
225897_at	**Myristoylated alanine-rich protein kinase C substrate**	***MARCKS***	**3.5(I)**	**6.7(I)**	**182.4(A)**	**540(P)**	**0.6(NC)**
201669_s_at	**Myristoylated alanine-rich protein kinase C substrate**	***MARCKS***	**3.1(I)**	**5.2(I)**	**158.6(P)**	**1162.6(P)**	**0.7(I)**
222810_s_at	RAS protein activator like 2	*RASAL2*	3.1(I)	1.5(I)	33.4(A)	854.5(P)	0.5(NC)
208944_at	**Transforming growth factor, beta receptor II (70/80 kDa)**	***TGFBR2***	**2.5(I)**	**2.2(I)**	**1974.7(P)**	**2132.9(P)**	**−1.2(D)**
218522_s_at	**BPY2 interacting protein 1**	***BPY2IP1***	**2.2(I)**	**3.8(I)**	**339.6(P)**	**548(P)**	**−0.2(NC)**
201637_s_at	**Fragile X mental retardation, autosomal homolog 1**	***FXR1***	**2.1(I)**	**2.1(I)**	**368(P)**	**6638.8(P)**	**0.1(NC)**
206777_s_at	Crystallin, beta B2	*CRYBB2*	2(I)	1.3(I)	267.2(A)	1080(P)	0.2(NC)
211240_x_at	**Catenin (cadherin-associated protein), delta 1**	***CTNND1***	**1.9(I)**	**1(I)**	**832.4(P)**	**1730.1(P)**	**0.4(NC)**
222760_at	Zinc finger protein 703	*ZNF703*	1.9(I)	1.3(I)	1856.6(P)	2655.4(P)	0.4(NC)
224326_s_at	Polycomb group ring finger 6	*PCGF6*	1.9(I)	1.1(I)	302.8(P)	698.5(P)	0.1(NC)
225548_at	Shroom	*SHRM*	1.9(I)	1.5(I)	573.7(P)	1764(P)	0.8(I)
235521_at	Homeo box A3	*HOXA3*	1.8(I)	3(I)	270.6(P)	1480.2(P)	−0.3(D)
204391_x_at	Tripartite motif-containing 24	*TRIM24*	1.7(I)	1.8(I)	902.5(P)	1211.9(P)	0.9(I)
229231_at	Leucine rich repeat containing 37B	*LRRC37B*	1.7(I)	1.1(I)	735.1(P)	1030.9(P)	0.9(I)

LS180 as well as SW480 cells were treated with emetine and in addition LS180 cells underwent C(C+A)/CA treatment (See Methods). Changes in mRNA amounts produced in cells by treatments recorded as signal $\log_2$ ratio (SLR) as well as absolute hybridization signal intensities in untreated cells were analysed using the Affymetrix GeneChip Human Genome U133Plus2.0 array. In the table, the numbers for the SLRs are accompanied with comments in parenthesis indicating statistically significant increased (I), decreased (D) or not changed (NC). Absolute signal intensities numbers are accompanied in parenthesis with the comments indicating present (P) or absent (A). Selected genes had to satisfy the following parameters: (1) SLR $\geqslant$ 1.7 and SLR $\geqslant$ 1 in LS180 cells following emetine and C(C+A)/CA treatment, respectively, (2) SLR $<$1 in SW480 cells following emetine treatment, (3) absolute signal intensity corresponding to untreated LS180 cells less than that corresponding to untreated SW480 cells, (4) absolute signal intensity corresponding to SW480 cells indicated as present (P) should be $>$500. [a]**In bold** are the genes with identified biallelic inactivating mutations.

Table 1b Reduction of the candidate gene list from Table 1a after additional filtering using GCRMA normalized data

Probe set ID	*Gene title*	*Gene symbol*	*LS180 emetine SLR (change)*	*LS180 C(A+C)/CA SLR (change)*	*SW480-emetine SLR (change)*
219121_s_at*	**RNA-binding motif protein 35A**	***RBM35A***	**5.51**	**1.10**	**0.72**
225897_at	**Myristoylated alanine-rich protein kinase C substrate**	***MARCKS***	**4.68**	**3.99**	**−0.09**
201669_s_at	**Myristoylated alanine-rich protein kinase C substrate**	***MARCKS***	**3.93**	**5.52**	**0.75**
208944_at	**Transforming growth factor, beta receptor II (70/80 kDa)**	***TGFBR2***	**2.75**	**2.56**	**−1.50**
218522_s_at	**BPY2 interacting protein 1**	***BPY2IP1***	**2.99**	**2.92**	**−0.46**
201637_s_at	**Fragile X mental retardation, autosomal homolog 1**	***FXR1***	**2.60**	**1.36**	**0.05**
206777_s_at	Crystallin, beta B2	*CRYBB2*	3.89	1.31	−0.26
222760_at	Zinc finger protein 703	*ZNF703*	2.85	1.31	0.35
225548_at	Shroom	*SHRM*	1.77	1.42	0.54
235521_at	Homeo box A3	*HOXA3*	2.59	3.42	−0.62

The signal log ratios (SLRs) are generated from the pairwise subtraction between signal log values after GCRMA normalization using 'gcrma' library from Bioconductor (www.bioconductor.org).

why GINI2 alone is not as efficient as in the combination with GINI1is that, despite the caffeine pretreatment, the subsequent incubation with actinomycin D results in lower levels of cellular mRNA. As a result, diminished signal intensities of hybridization are seen, effectively decreasing the 'signal to noise ratio' and consequently identifying a larger number of false positives.

Our results demonstrate that analysing changes in mRNA levels produced using both GINI1 and GINI2 protocols treatment to inhibit NMD is a highly efficient way to detect mutant genes. By analysing changes in mRNA levels following the GINI2 protocol, it was possible to distinguish genes that were stabilized by NMD inhibition from false positives that show mRNA increase due to increased transcription induced by a stress response. Semiquantitative RT–PCR analysis (Figure 4) illustrates how false positives produced by stress response to emetine treatment in LS180 cells can be successfully identified using the GINI2 protocol.

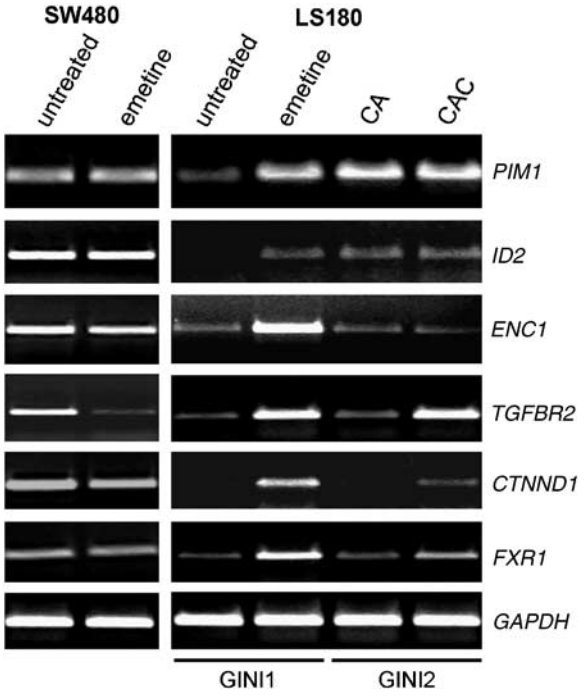

Figure 4 Selection of candidate genes for sequencing using analysis of changes in mRNA levels produced using both GINI1 and GINI2 protocols. These examples illustrate the strategy for prioritizing genes for sequencing following GINI protocols. RT–PCR analysis shows mRNA increases for six genes in LS180 but not in SW480 cells following emetine treatment (GINI1). As the level of mRNA for these genes is lower in LS180 than in SW480 cells, this suggests that basing only on GINI1 analysis all six genes would be selected as candidates for sequencing. GINI2 analysis, however, demonstrates the mRNA increases only for the *TGFBR2*, *CTNND1* and *FXR1* genes containing mutations but not for the *PIM1*, *ID2* and *ENC1* genes not harboring mutations in LS180 cells. CAC is the abbreviation for the pretreatment with caffeine followed by treatment with both actinomycin D and caffeine to block both transcription and NMD. CA is the abbreviation for the pretreatment with caffeine followed by treatment with actinomycin D alone.

RT–PCR analysis is shown for six different genes, none of which are mutated in the SW480 cell line. Emetine treatment in the GINI1 strategy in these cells does not result in an increased level of mRNA from these genes. In the LS180 cells, the mRNA levels for these genes in untreated cells is lower than in SW480 cells, which could be due either to normal variation between the cell lines or as a consequence of NMD. These genes, therefore, pass the first requirement in the prioritization. Next, the fact that emetine treatment of LS180 results in increased mRNA levels for these genes suggests this is due to either a stress response or lack of degradation by NMD, fulfilling the second requirement. Using the GINI2 strategy, no mRNA levels increases are seen for three of these genes (*PIM1*, *ID2*, *ENC1*), suggesting that they are stress response genes rather than carrying mutations. For the other three genes (*TGFBR2*, *CTNND1* and *FXR1*), increases in mRNA levels are seen after both GINI1 and GINI2 protocols, and sequencing analysis shows that they all carry mutations.

To identify mutated genes in the MSI-negative HT29 colon cancer cells, we compared mRNA profile changes following NMD inhibition in HT29 cells with that in LS180 cells. Using the same procedures and selection algorithm, we identified only four candidate genes in this comparison. This group included the *MADH4* gene, which has previously been reported to be homozygously mutated in HT29 (Woodford-Richens *et al*., 2001). No mutations were found in the remaining three candidates. When we lowered the stringency of the analytical filter, by using a lower numerical 'cutoff' for the GINI1 or GINI2 fold change thresholds, the number of candidates increased, but still no additional mutant genes were identified in HT29 cells. Furthermore, we could not identify PTC-generating mutations in MSI-negative SW480 colon cancer cells when LS180 or HT29 cells were used as controls.

The genes containing bi-allelic inactivating mutations identified so far in MSI-positive colon cancer cell lines using inhibition of NMD and Affymetrix genechips are described in Table 2. The prevailing mutation type in this study was the deletion or insertion of a single nucleotide in short-coding mononucleotide repeats. Such sequences represent mutational hotspots in cells with inactivated MMR (Markowitz *et al*., 1995; Rampino *et al*., 1997) and facilitated the analysis of these genes in MSI-positive primary colon tumors. To establish whether the genes identified as mutated in the colon cancer cell lines were also mutated in primary colon tumors showing MSI, the mononucleotide repeat-containing regions in these genes were sequenced in 36 tumors obtained from Oncomatrix Inc. (www.oncomatrix.com). Table 2 shows the mutation frequencies in primary colon tumors with MSI for the genes containing coding mononucleotide repeats longer than seven repetitive units which are mutated in cell lines. All analysed genes, except *SMAP-1*, were found to contain mutations in primary tumors with the frequencies largely proportional to the length of the coding microsatellites. However, some genes had higher mutation frequencies than other genes with the same length of the coding microsatellite. Moreover, the $(A)_8$ repeat of the *RBM35A* gene is mutated with higher (9/36, 25%) frequency than the $(A)_9$ repeat of the *RHAMM* gene, which is mutated in six of 36 (17%) tumors, suggesting the stronger selective pressure for the *RBM35A* gene inactivation in primary MSI-positive tumors. The absence of mutations in primary tumors for the *SMAP-1* gene, encoding for the stromal membrane-associated protein 1, suggests that the mutations in the HCT116, SW48 and RKO cells could occur during *in vitro* propagation.

Discussion

The original GINI strategy (GINI1) to identify genes containing nonsense mutations has been used successfully in proof-of-principle experiments to identify genes with known mutations in cell lines. Although it has also been used successfully to identify novel mutant genes in

Table 2 Genes mutated in colon cancer cell lines identified using GINI

Gene	*Gene symbol*	*Mutation*	*Cell line with gene mutated*	*Function*
RNA-binding motif protein 35A	*RBM35A*	DelA in $(A)_8$ nt 1528–35	LS180	RNA binding
Myristoylated alanine-rich protein kinase C substrate	*MARCKS*	DelA in $(A)_{11}$ nt 454–64	LS180, LoVo	Actin filament binding
BPY2 interacting protein 1	*BPY2IP1*	DelC in $(C)_6$ nt 588–93, Del AG in $(AG)_4$ nt 1447–54	LS180	Interacts with natural paclitaxel-like microtubule stabilizer
Fragile X mental retardation, autosomal homolog 1	*FXR1*	DelA in $(A)_8$ nt 364–71	Ls180	RNA binding
Catenin (cadherin-associated protein), delta 1	*CTNND1*	TGG→TGA nt 1431, DelTG in $(TG)_2$ nt 940–4	LS180	Cell-cell adhesion
Ring finger protein 43	*RNF43*	DelC in $(C)_6$ nt 343–48, DelG in $(G)_7$ nt 1969–75, DelTG in $(TG)_3$ nt 891–5	HCT116, SW48	Protein ubiquitination
SEC31-like 1 (Saccharomyces cerevisiae)	*SEC31L*	DelA in $(A)_9$ nt 1376–84	HCT116	ER to Golgi transport
Nuclear receptor co-repressor 1	*NCOR1*	DelC in $(C)_4$ nt 5494–7, InsA in $(A)_6$ nt 1568–73, DelTG in $(TG)_2$ nt 2463–6	RKO, HCT116	Regulation of transcription
mutS homolog 3 (Escherechia coli)	*MSH3*	DelA in $(A)_8$ nt 1141–8	RKO, HCT116	DNA mismatch repair
Hyaluronan-mediated motility receptor	*RHAMM*	DelA in $(A)_9$ nt 1990–8	RKO	Cell motility
PHD finger protein 14	*PHF14*	DelA in $(A)_7$ nt 530–6	HCT116	Regulation of transcription, DNA-dependent
Zinc finger protein 294	*ZNF294*	DelA in $(A)_{11}$ nt 1597–1607	Lovo, HCT116, RKO	Protein ubiquitination
Stromal membrane-associated protein 1	*SMAP-1*	Del A in $(A)_{10}$ nt 425–434	HCT116, SW48, RKO	Regulation of GTPase activity
HLA-B associated transcript 3	*BAT3*	DelC in $(C)_8$ nt 1180–7	LoVo	Implicated in the control of apoptosis and regulating heat shock protein
Beta-2-microglobulin	*B2M*	Del CT in $(CT)_4$ nt 37–44	LoVo	MHC class I receptor activity

Abbreviations: Del, deleted; ER, estrogen receptor; Ins, inserted; nt, nucleotides.

different systems, the problems associated with distinguishing stress response genes from mutated genes led to extensive unnecessary sequencing of genes that turned out not to carry mutations even despite custom filtering of the data. Even the use of actinomycin D to suppress *de novo* mRNA synthesis provided limited advantages, as the overall reduction in mRNA levels as a result of this treatment prevented detection of events related to moderately expressed genes. In this report, we describe a second-generation version of GINI analysis (GINI2), which uses caffeine as the inhibitor of NMD and incorporates a pretreatment step to boost overall mRNA levels. Using this strategy, in combination with the original GINI protocol, we have identified a series of genes, which are mutated in colon caner cell lines with MSI. In many cases, these genes have also been shown to be mutated in primary colon cancer tumors with MSI. Thus, the GINI2 approach has proved a significant development in the search for colon cancer-related genes. Using a modified GINI strategy, we have analysed five MSI-positive colon cancer cell lines and identified 13 novel genes that are mutated in colon cancers with MSI.

As arbitrarily chosen cutoff thresholds were used to select candidate genes for sequencing analysis, we have likely missed some genes harboring NMD-activating mutations in the cell lines analysed. For example, our series of mutant genes does not include *BAX*, *RIZ*, *PTEN*, *RAD50*, *CASPASE-5*, *IGFIIR*, *MSH6*, *DNA-PKCs* and many others that have been reported to be mutated at high frequencies in colorectal cancers with MSI. It is possible that many of these genes do not contain bi-allelic frameshift mutations in the cell lines analysed. The *BAX* gene, however, is known to harbor mutations in both alleles in LoVo and LS180 cells (Rampino *et al*., 1997), but has not been identified as the target for sequencing using our analysis. This fact is in agreement with recent publication describing differential NMD of mutated mRNAs in MMR-deficient colorectal cancers (El-Bchiri *et al*., 2005). The cause for the difference in the efficiency of NMD for the degradation of mutant mRNAs between individual genes is unknown. It is possible that the extent of the increase in mutant mRNA levels produced by NMD inhibition depends on the half-life of the corresponding wild-type mRNA transcripts. For example, if non-mutant mRNA for a hypothetical gene is degraded regardless of NMD immediately after the first round of translation, then the inhibition of NMD for the same gene harboring PTC will not result in accumulation of the mutant transcript. If this suggestion is true, then identifying mutant genes using GINI analysis for the transcripts with short half-lives will be less efficient than for the transcripts with longer half-lives. Loosening the stringency of the

analytical filter, by using a lower numerical 'cutoff' for the fold change thresholds in GINI1 or GINI2 will increase the number of candidates for sequence analysis, which might increase the number of mutant genes identified. This may, however, also result in a higher number of false positives, in turn reducing the efficiency of identifying mutant genes.

The GINI strategy can efficiently identify only those mutations, which affect both alleles of a gene and result in PTC inactivating gene function. Such mutations suggest possible tumor suppressor function of these genes. Identifying inactivating mutations in both alleles of the *CTNND1* gene in LS180 cells (Figure 5a) illustrates the efficiency of GINI for the identification of candidate

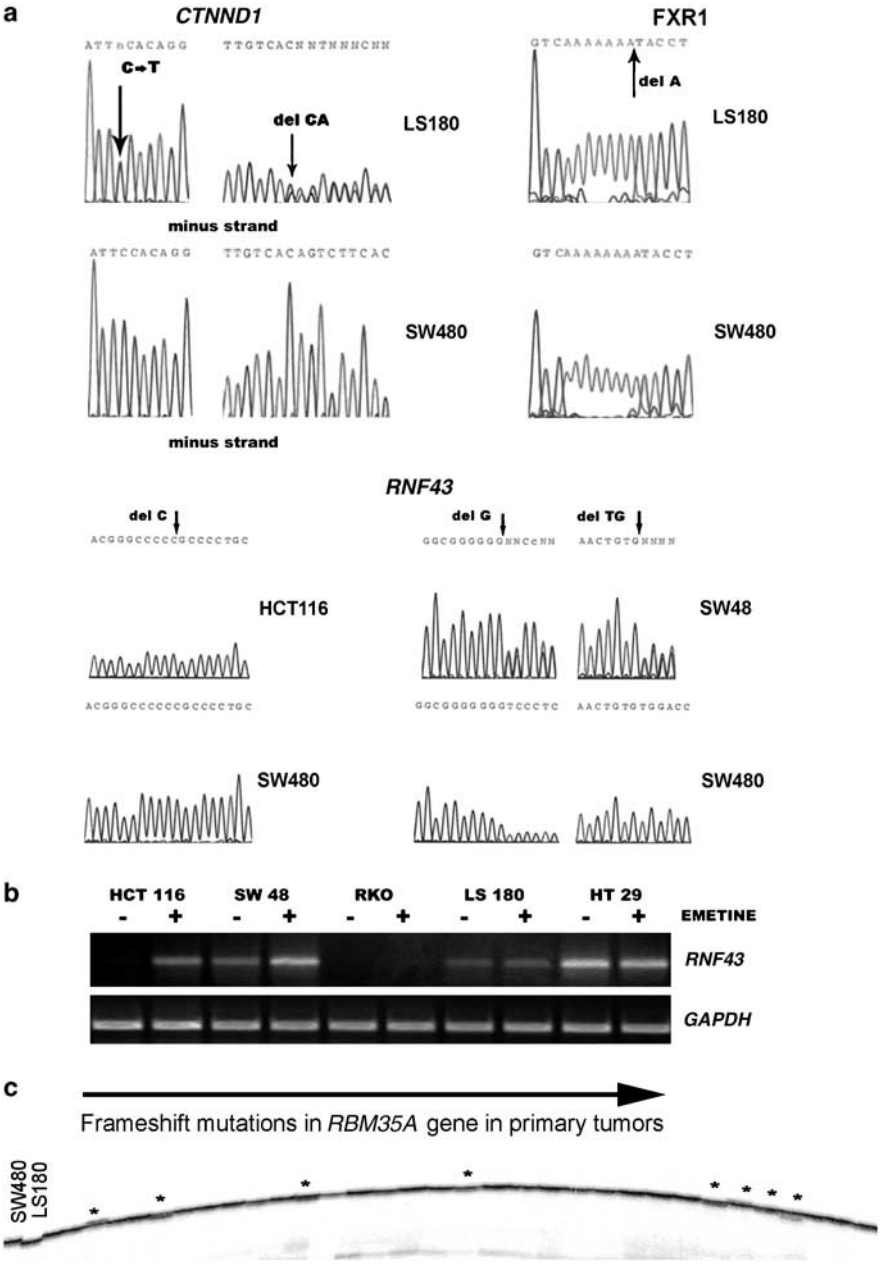

Figure 5 Mutation analysis of candidate genes selected using GINI. (**a**) Sequencing analysis identified two heterozygous mutations in LS180 cells in the *CTNND1* gene (one is a G to A substitution resulting in a TGA stop codon and the other is TG deletion in a $(TG)_2$ repeat (shown on the minus strand sequencing chromatogram). A homozygous deletion of one adenosine in *FXR1* gene is seen, and a homozygous deletion of one cytosine and two heterozygous mutation (a deletions of a G in a $(G)_7$ and TG in a $(TG)_3$ repeats) are seen in *RNF43* in HCT116 and SW48 cells, respectively. Wild-type sequences of the corresponding normal region from SW480 cells are shown as controls. (**b**) RT–PCR analysis of *RNF43* mRNA levels in both untreated and emetine treated colon cancer cells. Cell lines HCT116 and SW48, with identified inactivating mutations in the *RNF43* gene, show mRNA increases following inhibition of NMD with emetine, unlike LS180 and HT29, which do not have *RNF43* mutations. Cell line RKO does not express RNF43 mRNA. (**c**) Denaturing sequencing gel electrophoresis shows shifts in mobility for nine PCR-amplified fragments indicating frameshift mutations. Frameshift mutations in the *RBM35A* gene in colon tumors with MSI (indicated by the *) sign were detected by autography of PCR-amplified genomic DNA region containing the $(A)_8$ tract.

tumor suppressor genes. *CTNND1* has been also found to be mutated by others in SW48 MSI-positive colon cancer cells and the tumor suppressor function of this gene had been supported by functional studies (Ireton *et al*., 2002; Thoreson and Reynolds, 2002).

The majority (12/13) of the mutations in colon cancer cell lines identified using GINI analysis occurred in coding microsatellite sequences in the genes involved. On a background of inactivated MMR, it is possible that these mutations could occur as a consequence of the high overall instability of microsatellite DNA (bystander effect), rather than being tumor cell promoting events. Clearly, the only way to distinguish between these possibilities is to perform functional studies to determine whether they play a role in the development of the malignant phenotype. GINI analysis provides the candidate genes to perform these analyses. There have, however, been some estimates of the mutation frequencies in coding microsatellites of the genes contributing to tumor development (target genes) (Boland *et al*., 1998; Perucho, 1999a, b; Duval *et al*., 2001, 2002; Zhang *et al*., 2001; Suzuki *et al*., 2002; Woerner *et al*., 2003), which appears to depend largely on the length of the repeat (Kunkel *et al*., 1994; Chen *et al*., 1995). According to statistical analyses describing the relationship between the mutation frequency and the relevance to tumor development, the frequency of the mutations in coding microsatellites in bystander genes (not related to tumorigenesis) are no higher than the frequency of mutations in non-coding microsatellites of the same length and type (Duval *et al*., 2002; Woerner *et al*., 2003). The frequency of mutations seen in many non-coding repeats that are less than nine mononucleotides, has been shown to be lower than 6% (Suzuki *et al*., 2002) and so the mutations seen in the *RBM35A* gene, for example, are less likely to be due to a bystander effect, as this gene was mutated in 25% of primary tumors with MSI (Table 2 and Figure 5).

To our knowledge, there are no studies describing the frequency of mutations at non-coding microsatellites that are less than eight nucleotides long. During our screen for mutations using PCR products derived from genomic regions containing mononucleotide repeats in the regions flanking exons in over 100 sequencing analyses, we have never encountered mutations in these sequences. We suggest, therefore, that bi-allelic frameshift mutations occurring in the short-coding microsatellites of *BPY2IP1*, *NCOR1* and *RNF43* genes (Table 2) are unlikely to be due to background instability of coding microsatellite repeats. Moreover, each allele of these genes was inactivated by different heterozygous mutations, indicating the absence of a particularly unstable repetitive sequence in the coding DNA in these genes. Instead, these mutations may be considered to provide a selective advantage as a result of gene inactivation.

The protein encoded by *BPY2IP1*, also known as cell death inducer *C19ORF5*, is a hyperstabilized microtubule-specific binding protein, which causes mitochondrial aggregation and cell death (Liu *et al*., 2005c). Its involvement in tumor suppression is suggested by its interaction with natural paclitaxel-like microtubule stabilizer and candidate tumor suppressor *RASSF1A* (Liu *et al*., 2002, 2005a–c), which is frequently silenced by promoter hypermethylation in colon tumors with MSI (Oliveira *et al*., 2005).

Although the role of the *NCOR1* gene in suppression of colon tumorigenesis has not been reported, there is evidence for its involvement in the development of breast cancer. *NCOR1* associates with estrogen receptor-α (ER-α) preventing tamoxifen from stimulating proliferation in breast cancer cells. This effect is achieved through repression of a subset of target genes involved in ER-α function and cell proliferation (Keeton and Brown, 2005). It has also been shown that low NCOR1 expression in breast tumors was associated with a significantly shorter relapse-free survival (Girault *et al*., 2003) and that patients with high *NCOR1* expression levels have a better prognosis than those with low expression (Zhang *et al*., 2005). Mutational inactivation of this gene in colon cancer cell lines also suggests a role for *NCOR1* in colon cancer.

The RING finger protein 43 (*RNF43*), which shows homozygous deletion of one nucleotide in a $(C)_6$ repeat in HCT116 cells, and two heterozygous mutations in other parts of the gene in SW48 cells (Figure 5a), does not express mRNA in RKO cells as shown by RT–PCR analysis. MSI in RKO cells is known to be the consequence of *hMLH1* gene silencing due to promoter hypermethylation (Veigl *et al*., 1998). It is possible, therefore, that the loss of expression of the *RNF43* gene in RKO cells may also be due to an epigenetic mechanism. Loss of gene expression in one cell line, and the mutational inactivation in the others, argues in favor of the selective advantage provided by inactivation of the *RNF43* gene. The presence of a RING finger domain in the protein sequence can be associated with the ubiquitin-protein ligase activity of the gene product, suggesting a possible involvement in the mechanism of cell cycle progression as well as in the regulation of the expression of other genes by initiating the degradation of their protein products (Fang *et al*., 2003).

In addition to *RNF43* and *NCOR1*, several other genes that contain bi-allelic inactivating mutations in colon cancer cell lines may affect the gene expression profiles of cancer cells. The *RBM35A* gene, which contains an RNA-binding motif may be involved in mRNA splicing or degradation. *ZNF294*, which is mutated in all of the MSI-positive colon cancer cell lines analysed, also contains a RING finger motif which, like *RNF43*, may regulate the expression of other genes through its ubiquitin-protein ligase activity. *FXR1*, which is mutated in LS180 cells, has recently been shown to post-transcriptionally regulate the expression of genes containing AU-rich elements within the 3′-untranslated regions of their mRNA. This regulation modulates translational efficacy and mRNA stability (Garnon *et al*., 2005). The plant homeodomain finger protein (PHF14) gene that is mutated in HCT116 cells contains a plant homeodomain (PHD) finger domain known to be involved in chromatin-mediated transcriptional regulation (Aasland *et al*., 1995). Although some

of these genes may not contribute to colon tumorigenesis directly, they may regulate the expression of known or yet unknown oncogenes or tumor suppressor genes that function downstream of these mutated genes. Functional *in vivo* studies in mice are required to provide a definite proof for the role of mutant genes in colon cancer development.

Only one mutant gene (*SMAD4* in HT29 cells) has been identified in two MSI-negative colon cancer cell lines HT29 and SW480 using the same GINI2 strategy. This is in agreement with the lower mutation rate reported in cells with functional MMR. It is noteworthy that both of these cell lines have *p53* mutations, whereas MSI-positive cell lines usually have wild-type *p53*. It is possible, therefore, that the higher frequency of detectable mutant genes in cells with inactivated MMR can compensate the absence of *p53* mutations. It is also possible that epigenetic mechanisms of gene inactivation may substitute for the mutational inactivation of genes in MSI-negative cells.

Methods

Cell culture and drugs treatments
Colon cancer cell lines were obtained from American Type Culture Collection repository and maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics. All NMD inhibiting drugs were added when cells achieved 50–90% of confluence.

The emetine treatment (*GINI1*) protocol was as described previously (Ionov *et al*., 2004b). Briefly, cells were incubated for 8 h in tissue culture medium containing 100 μg/ml of emetine before total RNA was extracted from emetine treated or untreated cells. The caffeine, cantharidin and okadaic acid treatment protocols were as described for emetine using concentrations of 10 mM for caffeine and 100 nM for okadaic acid or cantharidin. In the latter two cases, the incubation times were reduced to 3 h to reduce cell death.

The GINI2 treatment was as follows. Cells were seeded in two tissue culture plates and caffeine (10 mM) was added to both plates. Following 4-h incubation, the medium from both plates was removed, cells were washed twice with phosphate-buffered saline and actinomycin D (2 μg/ml) together with caffeine (10 mM) was added to one plate and actinomycin D alone was added to the other plate. Following further 4-h incubation, total RNA from both plates was prepared and used for Northern blotting as well as for Affymetrix U133Plus2.0 oligonucleotide array analysis.

RNA isolation and Northern blot
Total RNAs for microarray and Northern blot analysis were prepared using TRIZOL reagents (Invitrogen, Carlsbad, CA, USA). For Northern blots, 20 μg of total RNA was size fractionated on 1% agarose formaldehyde gels and then transferred to nylon membranes (Hybond-N, Amersham, Boston, MA, USA). Hybridizations with radiolabeled probes were carried out at 68°C using MyracleHyb solution (Stratagene, Cedar Creek, TX, USA) according to the manufacturer's protocol. Radioactivity was detected and quantified using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA, USA). Radiolabelled probes generated by PCR from RNA-derived cDNA were generated using the random-primed DNA labeling kit (Roche, Indianapolis, IN, USA).

Affymetrix oligonucleotide array data analysis
Messenger RNA levels in drug treated and untreated cells were measured using the Affymetrix U133Plus2.0 array. GeneChip expression array analysis was performed using Affymetrix Microarray Suite software version 5 (GeneChip Analysis Suite). This software uses two independent sets of algorithms, a quantitative algorithm (robust estimator of the mean difference in probe intensities) that computes the raw signal and the SLR directly from the hybridization intensities of the probes, as well as a confidence algorithm (nonparametric) that provides *P* values to estimate the confidence in detection (absent or present) or change (increase, decrease or no change) of a specific target.

An adaptation of RMA (Irizarry *et al*., 2003), with specific correction for GC biases known as GC-RMA (Wu *et al*., 2004) procedure was applied for the .CEL files of Affymetrix U133Plus2.0 data to further filter candidates for sequencing analysis. The resultant SLRs were generated based on pairwise comparison between the various treatments and control groups. GC-RMA uses only the perfect match (PM) intensity values and ignores the mismatch (MM) intensities which have been shown to introduce variation (Naef *et al*., 2002). The normalized PM values were then log transformed and all the probes in a set representing specific genes were analysed using Tukey's median polishing procedure.

Sequencing analysis
One microgram of total RNA from emetine treated colon cancer cells was reverse transcribed using the SuperScript II protocol (Invitrogen, Carlsbad, CA, USA). Overlapping PCR primer sets were used to generate products spanning the entire open reading frames for candidate genes. Primers for sequencing analysis were designed using Primer3 software available online (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Genomic DNA samples from colon tumors with MSI were purchased from Oncomatrix (Ocean Side, CA, USA). Primers for PCR amplification of genomic DNA fragments containing coding microsatellite sequences were designed using Primer3 software and are available upon request. The PCR products were gel purified and sequenced using the Applied Biosystems' PRISM 3100 Genetic Analyzer.

Analysis of frameshift mutations in DNA from primary colon tumors
Regions of approximately 100 bp encompassing the mononucleotide repeats of specific genes were amplified by PCR using specific primers (sequences are available upon request). PCR was carried out with Vent DNA polymerase (New England Biolabs, Boston, MA, USA) with one cycle at 94°C for 4 min followed by 30 cycles of 94, 54 and 68°C, for 30 s in each cycle in the presence of 0.2 μCi of [^{33}P]dCTP (1 Ci = 37 GBq) (PerkinElmer, Boston, MA, USA). PCR products were electrophoresed in a SequaGel XR (National Diagnostics, Atlanta, Georgia, USA) polyacrylamide gel. The gel was dried on filter paper and subjected to autoradiography.

Acknowledgements

This work was supported by grants Roswell Park Alliance Foundation grant, 1R01 CA109256-01 from the National Cancer Institute and PC030076 from the Prostate Program of the United States Army Department of Defense (Yurij Ionov) and in part by the RPCI Cancer Center Support Grant CA 16056 for all authors and NIH NSO54543 (John Cowell).

References

Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP *et al*. (1993). Clues to the pathogenesis of familial colorectal cancer. *Science* **260**: 812–816.

Aasland R, Gibson TJ, Stewart AF. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* **20**: 56–59.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW *et al*. (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **58**: 5248–5257.

Brumbaugh KM, Otterness DM, Geisen C, Oliveira V, Brognard J, Li X *et al*. (2004). The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. *Mol Cell* **14**: 585–598.

Chen J, Heerdt BG, Augenlicht LH. (1995). Presence and instability of repetitive elements in sequences the altered expression of which characterizes risk for colonic cancer. *Cancer Res* **55**: 174–180.

Duval A, Reperant M, Hamelin R. (2002). Comparative analysis of mutation frequency of coding and non coding short mononucleotide repeats in mismatch repair deficient colorectal cancers. *Oncogene* **21**: 8062–8066.

Duval A, Rolland S, Compoint A, Tubacher E, Iacopetta B, Thomas G *et al*. (2001). Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. *Hum Mol Genet* **10**: 513–518.

El-Bchiri J, Buhard O, Penard-Lacronique V, Thomas G, Hamelin R, Duval A. (2005). Differential nonsense mediated decay of mutated mRNAs in mismatch repair deficient colorectal cancers. *Hum Mol Genet* **14**: 2435–2442.

Fang S, Lorick KL, Jensen JP, Weissman AM. (2003). RING finger ubiquitin protein ligases: implications for tumorigenesis, metastasis and for molecular targets in cancer. *Semin Cancer Biol* **13**: 5–14.

Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J *et al*. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**: 1027–1038.

Garnon J, Lachance C, Di Marco S, Hel Z, Marion D, Ruiz MC *et al*. (2005). Fragile X-related protein FXR1P regulates proinflammatory cytokine tumor necrosis factor expression at the post-transcriptional level. *J Biol Chem* **280**: 5750–5763.

Girault I, Lerebours F, Amarir S, Tozlu S, Tubiana-Hulin M, Lidereau R *et al*. (2003). Expression analysis of estrogen receptor alpha coregulators in breast carcinoma: evidence that NCOR1 expression is predictive of the response to tamoxifen. *Clin Cancer Res* **9**: 1259–1266.

Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. (2004). Nonsense-mediated decay approaches the clinic. *Nat Genet* **36**: 801–808.

Huusko P, Ponciano-Jackson D, Wolf M, Kiefer JA, Azorsa DO, Tuzmen S *et al*. (2004). Nonsense-mediated decay microarray analysis identifies mutations of EPHB2 in human prostate cancer. *Nat Genet* **36**: 979–983.

Ionov Y, Matsui S, Cowell JK. (2004a). A role for p300/CREB binding protein genes in promoting cancer progression in colon cancer cell lines with microsatellite instability. *Proc Natl Acad Sci USA* **101**: 1273–1278.

Ionov Y, Nowak N, Perucho M, Markowitz S, Cowell JK. (2004b). Manipulation of nonsense mediated decay identifies gene mutations in colon cancer Cells with microsatellite instability. *Oncogene* **23**: 639–645.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**: 558–561.

Ireton RC, Davis MA, van Hengel J, Mariner DJ, Barnes K, Thoreson MA *et al*. (2002). A novel role for p120 catenin in E-cadherin function. *J Cell Biol* **159**: 465–476.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al*. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249–264.

Ishigaki Y, Li X, Serin G, Maquat LE. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**: 607–617.

Keeton EK, Brown M. (2005). Cell cycle progression stimulated by tamoxifen-bound estrogen receptor-alpha and promoter-specific effects in breast cancer cells deficient in N-CoR and SMRT. *Mol Endocrinol* **19**: 1543–1554.

Knudson Jr AG. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* **68**: 820–823.

Konishi M, Kikuchi-Yanoshita R, Tanaka K, Muraoka M, Onda A, Okumura Y *et al*. (1996). Molecular nature of colon tumors in hereditary nonpolyposis colon cancer, familial polyposis, and sporadic colon cancer. *Gastroenterology* **111**: 307–317.

Kunkel TA, Patel SS, Johnson KA. (1994). Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit. *Proc Natl Acad Sci USA* **91**: 6830–6834.

Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R *et al*. (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**: 1215–1225.

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

Liu L, Amy V, Liu G, McKeehan WL. (2002). Novel complex integrating mitochondria and the microtubular cytoskeleton with chromosome remodeling and tumor suppressor RASSF1 deduced by in silico homology analysis, interaction cloning in yeast, and colocalization in cultured cells. *In vitro Cell Dev Biol Anim* **38**: 582–594.

Liu L, Vo A, Liu G, McKeehan WL. (2005a). Distinct structural domains within C19ORF5 support association with stabilized microtubules and mitochondrial aggregation and genome destruction. *Cancer Res* **65**: 4191–4201.

Liu L, Vo A, Liu G, McKeehan WL. (2005b). Putative tumor suppressor RASSF1 interactive protein and cell death inducer C19ORF5 is a DNA binding protein. *Biochem Biophys Res Commun* **332**: 670–676.

Liu L, Vo A, McKeehan WL. (2005c). Specificity of the methylation-suppressed A isoform of candidate tumor suppressor RASSF1 for microtubule hyperstabilization is determined by cell death inducer C19ORF5. *Cancer Res* **65**: 1830–1838.

Maquat LE. (2005). Nonsense-mediated mRNA decay in mammals. *J Cell Sci* **118**: 1773–1776.

Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J *et al*. (1995). Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**: 1336–1338.

Naef F, Hacker CR, Patil N, Magnasco M. (2002). Characterization of the expression ratio noise structure in high-density oligonucleotide arrays. *Genome Biol* **3**: research0018.1–0018.11.

Noensie EN, Dietz HC. (2001). A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. *Nat Biotechnol* **19**: 434–439.

Ohnishi T, Yamashita A, Kashima I, Schell T, Anders KR, Grimson A *et al*. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol Cell* **12**: 1187–1200.

Oliveira C, Velho S, Domingo E, Preto A, Hofstra RM, Hamelin R *et al*. (2005). Concomitant RASSF1A hypermethylation and KRAS/BRAF mutations occur preferentially in MSI sporadic colorectal cancer. *Oncogene* **24**: 7630–7634.

Perucho M. (1999a). Correspondence re: C.R. Boland et al., A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **58**: 5248–5257.

Perucho M. (1999b). Correspondence re: C.R. Boland et al., A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **59**: 249–56.

Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, *et al*. (1997). Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**: 967–969.

Rossi MR, Hawthorn L, Platt J, Burkhardt T, Cowell JK, Ionov Y. (2005). Identification of inactivating mutations in the JAK1, SYNJ2, and CLPTM1 genes in prostate cancer cells using inhibition of nonsense-mediated decay and microarray analysis. *Cancer Genet Cytogenet* **161**: 97–103.

Salovaara R, Roth S, Loukola A, Launonen V, Sistonen P, Avizienyte E *et al*. (2002). Frequent loss of SMAD4/DPC4 protein in colorectal cancers. *Gut* **51**: 56–59.

Suzuki K, Dai T, Suzuki I, Dai Y, Yamashita K, Perucho M. (2002). Low mutation incidence in polymorphic noncoding short mononucleotide repeats in gastrointestinal cancer of the microsatellite mutator phenotype pathway. *Cancer Res* **62**: 1961–1965.

Thibodeau SN, Bren G, Schaid D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* **260**: 816–819.

Thoreson MA, Reynolds AB. (2002). Altered expression of the catenin p120 in human cancer: implications for tumor progression. *Differentiation* **70**: 583–589.

Usuki F, Yamashita A, Higuchi I, Ohnishi T, Shiraishi T, Osame M *et al*. (2004). Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's disease. *Ann Neurol* **55**: 740–744.

Veigl ML, Kasturi L, Olechnowicz J, Ma AH, Lutterbaugh JD, Periyasamy S *et al*. (1998). Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci USA* **95**: 8698–8702.

Vogelstein B, Kinzler KW. (2004). Cancer genes and the pathways they control. *Nat Med* **10**: 789–799.

Weischenfeldt J, Lykke-Andersen J, Porse B. (2005). Messenger RNA surveillance: neutralizing natural nonsense. *Curr Biol* **15**: R559–62.

Woerner SM, Benner A, Sutter C, Schiller M, Yuan YP, Keller G *et al*. (2003). Pathogenesis of DNA repair-deficient cancers: a statistical meta-analysis of putative Real Common Target genes. *Oncogene* **22**: 2226–2235.

Woodford-Richens KL, Rowan AJ, Gorman P, Halford S, Bicknell DC, Wasan HS *et al*. (2001). SMAD4 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway. *Proc Natl Acad Sci USA* **98**: 9719–9723.

Wu Z, Irizarry RA, Gentleman R, Murillo FM, Spencer F. (2004). A model based background adjustment for oligonucleotide expression arrays (May 28, 2004). Johns Hopkins University, Department of Biostatistics Working Papers. Working Paper 1.

Zhang L, Yu J, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B. (2001). Short mononucleotide repeat sequence variability in mismatch repair-deficient cancers. *Cancer Res* **61**: 3801–3805.

Zhang Z, Yamashita H, Toyama T, Sugiura H, Ando Y, Mita K *et al*. (2005). NCOR1 mRNA is an independent prognostic factor for breast cancer. *Cancer Lett* **237**: 123–129.