AD_____

Award Number:

W81XWH-08-1-0584

TITLE:

Dissecting the Molecular Mechanism of RhoC GTPase Expression In The Normal and Malignant Breast

PRINCIPAL INVESTIGATOR:

John Brenner

CONTRACTING ORGANIZATION:

The University of Michigan, Ann Arbor, Michigan 48109-1274

REPORT DATE:

September 2010

TYPE OF REPORT:

Annual Summary

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

DISTRIBUTION STATEMENT:

⊗ Approved for public release; distribution unlimited

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From - To)* |
|---|---|---|
| 01-09-2010 | Annual Summary | 1 SEP 2009 - 31 AUG 2010 |

**4. TITLE AND SUBTITLE**
Dissecting the Molecular Mechanism of RhoC GTPase Expression In The Normal and Malignant Breast

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
W81XWH-08-1-0584

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
John Brenner

diannal@Umich.edu

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
The University of
Michigan
Ann Arbor, Michigan 48109-
0127

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research
and Material 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**
Primary inflammatory breast cancer (IBC) accounts for approximately 3% of new breast cancers in the US. This form of locally advanced breast cancer is rapidly metastatic and, because of this disease's rapid progression, the effectiveness of aggressive multimodality treatment is limited; the 5-year disease-free, mean survival rate is less than 45%, making IBC the most lethal form of breast cancer. Here, we report that RhoC GTPase expression is regulated by the NfκB pathway. Specifically, p65 binds to and activates the RhoC promoter leading to increased RhoC mRNA expression and RhoC-mediated motility and invasion in IBC cell lines, but not control metastatic breast cancer cell lines. Additionally, we report that IBC has an additional copy of chromosome 1, possibly leading to an additional mechanism of increased gene expression. Finally, although we did not find any recurrent gene fusions in IBC by high throughput transcriptome sequencing, by microRNA array, we found that miR-31 and miR-31* are specifically downregulated in IBC cell lines. Taken together, we have identified several molecular alterations which drive the aggressive phenotype of IBC cell lines and propose that these may represent important targets for future studies of IBC.

**15. SUBJECT TERMS**
RhoC GTPase, miR-31, inflammatory breast cancer, metastasis

| **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 | 47 | **19b. TELEPHONE NUMBER** *(include area code)* |

## Table of Contents

**Page**

## Introduction:

Primary inflammatory breast cancer (IBC) accounts for approximately 3% of new breast cancers in the US. This form of locally advanced breast cancer is characterized clinically by erythema, warmth, and dimpling of the skin that arise rapidly, typically within six months. IBC is generally not associated with precursor lesions and is rapidly invasive from the outset, especially to the skin and lymphatics, and is highly angiogenic and metastatic. Because of this disease's rapid progression, the effectiveness of aggressive multimodality treatment is limited; the 5-year disease-free, mean survival rate is less than 45%, making IBC the most lethal form of breast cancer (1). This rapid progression is due to the development of distant metastases, indicating that the tumors quickly acquire the ability to invade and metastasize during tumor development. This suggests that the unique aggressive inflammatory phenotype of IBC is the result of a limited number of concordant genetic alterations. As such, IBC constitutes an excellent paradigm to understand aggressive phenotypes in breast cancer. Previously, our laboratory has found concordant and consistent overexpression and of RhoC GTPase in tissue samples from patients with IBC as compared to stage-matched non-IBC (2, 3). We have also demonstrated that RhoC GTPase occupies an integral role in the aggressive phenotype of IBC (4, 5). With the increasing evidence that RhoC and other ras-homology family proteins play a significant role in other cancers (6, 7), the therapeutic importance of inhibiting RhoC activity is clear, highlighting the crucial need to uncover the the molecular mechanisms leading to RhoC-driven metastatic phenotype of IBC. In spite of this need, however, a model explaining the mechanisms of RhoC overexpression in breast cancer does not exist. The goal of this award is to establish such a model. ***Our central hypothesis was that overexpression of RhoC GTPase in metastatic breast cancer is due to gene amplification, epigenetic deregulation, transcription factor deregulation, and/or enhanced or differential mRNA stability. Because of these cellular and molecular alterations, early stage IBC is subject to rapid metastatic spread through downstream effectors signaling for invasion and angiogenesis.***

**Body:**

As I matriculated through graduate school, I originally thought that I was going to work full time in Dr. Merajver's lab (first graduate school rotation) studying IBC. However, as I was granted this award I was also choosing to transfer into Dr. Arul Chinnaiyan's lab. Fortunately, I was granted permission by the DOD to transfer the award to follow me to continue working on this project from Dr. Chinnaiyan's lab. Because of this, we have been able to establish an effective and highly collaborative meeting with Dr. Merajver where I attend her bi-weekly lab meetings and work with a technician in her lab to help complete this project. This has given me a lot of unique experiences. For example, learning how to create well defined experimental protocols and making sure that the technician has the appropriate materials and controls to execute each experiment. As such, continuing this DOD pre-doctoral grant has given me the opportunity to continue existing collaborations and to continue improving my leadership skills through working with a technician on a daily basis.

In addition to working with a technician, I have also had the opportunity to train five undergraduate students through the University of Michigan Undergraduate Research Opportunities Program and one Master's degree student. The students have learned several different protocols including PCR, restriction digests, Gateway cloning, DNA miniprep, DNA maxiprep, RNA isolation, cDNA synthesis, qRT-PCR, PCR, Western blotting, transfections of

both large DNA vectors and siRNA into mammalian cells, cell culture, production of lentivirus and lentiviral transduction, cell invasion assays, cell growth assays and propidium iodide staining. Additionally, I have led a bi-weekly cancer biology journal club meeting with all of the students in our lab (26 undergraduates). At the end of each semester, I help the students compile their results to present at a lab meeting and at an Undergraduate research forum by both poster presentation and lecture. Importantly, three of the students that I have trained have been awarded NIH summer fellowships that funded their work in the lab for the entire summer.

In addition to directly working with undergraduate students in the lab, as part of the statement of work the training plan I am currently a graduate assistant teaching Cancer Biology for incoming graduate students. For this course, I co-ordinate different lectures and also prepare and give several lectures throughout the semester. For example, this year I will be teaching lectures on GTPase oncogenes like RhoC, DNA damage and translocations as well as the use of high throughput sequencing for modern cancer biology. Additionally, I will give review courses for the other professors' lectures and have weekly open office hours for the students. Finally, as part of the training program I have been able to host several speakers for a University of Michigan speaker series and co-ordinate student discussions with the lectures by providing background reading and a background lecture for incoming speakers.

While I have found reward in the successes that students have experienced after working with them in the various teaching formats, I have also been able to learn several new experimental techniques that I would not otherwise have had the opportunity to learn without this training grant including Solexa high throughput Transcriptome sequencing, Fluorescence in situ hybridization as well as running aCGH and microRNA arrays. Perhaps more interesting is the analysis algorithms that I am helping to develop, including those used to identify novel gene fusions from paired end sequencing data (8), for the analysis of my global profiling data from these IBC cell line samples. While little is known about the molecular origins of inflammatory breast cancer, we have made significant advances not only in the acquisition of large profiling data sets of DNA copy number, microRNA expression and transcriptome sequencing, but also in software development to analyze this data. Currently, we are in the process of completing an integrated analysis from all three profiling platforms. Additionally, we have unexpectedly found that the two IBC cell lines SUM149 and SUM190 have an extra copy of chromosome 1. Because several other stage matched breast cancer cell lines do not have this extra copy of chromosome 1, we are exploring the occurrence of chromosome 1 amplification in IBC clinical samples. The significance of this finding is still unclear, but will be explored in more detail if a clinical correlation is observed.

The opportunity to work on developing novel techniques and protocols for this project has led directly to opportunities to improve my communication and professional skills. Within the last two years, I have presented some of the work at the American Association for Cancer Research Meeting in Denver, Colorado (April 2009 and April 2010). At those meetings, I was a co-author or first author on six posters on both the role of RhoC GTPases in IBC and other breast cancers as well as co-author on an abstract that I presented by podium presentation. Additionally, this research led to a scholarship to attend a keystone conference in Victoria, British Columbia. For this meeting, I wrote a meeting summary that was published as part of the conference proceedings. Following the research for this project, I received an independent nomination to

become an American Association of Cancer Research Associate council member. Finally, this research has led directly to the generation of preliminary data that was used to produce a grant, which I co-authored and is funded through the Susan G. Komen foundation N012788-00 (11-PAF00190).

## Key Research Accomplishments:

**Specific Aim 1:** ***To delineate if and how gene amplification in RhoC GTPase occurs in breast cancer and to identify novel gene fusions in inflammatory breast cancer.***

- Completed RhoC FISH and discovered that IBC cell lines do not have amplification of the RhoC locus, but carry an extra copy of chromosome 1. (**Figure 1**)
- Acquired 244k Agilent aCGH data for several cell lines including the two IBC cell lines, SUM149 and SUM190. Recurrent aberrations between the two IBC cell lines were not observed.
- Completed the Illumina bead station microRNA profiling chip V2 of cell line panel including HME, MCF10A, SUM149, SUM190, HCC1937 and BT20. This led to the identification of has-miR-31 and the anti-sense hsa-miR-31* as downregulated in the IBC cell lines, but not the control cell lines (**Table 1**). This observation was confirmed using Taqman qPCR probes to analyze mature miR-31 and miR-31* expression across the panel of cell lines. Because miR-31 has recently been shown to suppress metastatic breast cancer (Valastyan et. al., 2009), we are currently exploring the specific role of miR-31 in IBC.
- Sequenced the RNA transcriptome of both SUM149 and SUM190 using massively parallel, high throughput paired-end sequencing on a SOLEXA GA2 from Illumina. While we found and pursued several gene fusions, we were unable to identify any recurrent gene fusions using our integrated techniques. As such we have decided to pursue the SOLEXA data in more detail by analyzing the role of non-coding RNAs (ncRNAs) in both IBC and highly metastatic breast cancer. To do this, we are attempting to identify and validate ncRNAs that are specifically expressed in either IBC, triple negative or metastatic breast cancer. Additionally, we have generated ChIP-Sequencing (ChIP-SEQ) libraries of 17-β-estradiol treated MCF7 and BT474 cells in order to assess which of these ncRNAs may be estrogen responsive. To demonstrate the success of these experiments, **Figure 2** shows ChIP-Seq coverage maps of the estrogen-regulated gene *GREB1* from MCF7 and BT474 cell lines starved for 48 hours and then treated with either 1nM 17-β-estradiol or vehicle for 48 hours. ChIP-assays were performed with antibodies against ERα or a histone mark of activated transcription, H3K4-Me3.

**Specific Aim2:** ***To determine how DNA methylation status and histone modifications regulate the RhoC GTPase promoter, and to assess the ability of the small molecule drugs 5-azacytidine and Trichostatin A to alter the metastatic phenotype depicted by an IBC cell line model.***

- Completed Illumina bead station microRNA profiling chip V2 of cell line panel including HME, MCF10A, SUM149, SUM190, MDA-MB-231, HCC1937 and BT20 treated with 5-azacytidine or Trichostatin A.
- Prepared RNA transcriptome libraries of both SUM149 and SUM190 treated 5-azacytidine or Trichostatin A for sequencing on an Illumina SOLEXA GA2.

- Treatment of MCF10A and HME cells with either 5-azacytidine or Trichostatin A revealed no significant increase in RhoC mRNA expression suggesting that the molecular mechanism leading to RhoC overexpression does not involve the activation of genes repressed by either methylation or deacetylation.

**Specific Aim3:** ***To characterize the consequences of down regulating the expression of the transcription factors FoxP3, HoxA3, HoxB7, HoxB8, HoxD9, HoxD10, CREB and NFκB1, all of which contain highly conserved binding sites in the putative RhoC GTPase promoter, on molecular pathways regulating cell proliferation, survival and the metastatic phenotype, using an RNAi model system of human IBC cell lines.***

- Established stable shRNA knockdown cell lines for FoxP3, HoxA3, HoxB7, HoxB8, HoxD9, HoxD10, CREB and NFκB1 in SUM149 cells.
- Identified NFκB1 as a key regulator of RhoC mRNA and protein expression in SUM149 and SUM190 cells. (**Figure 3**)
- Completed chromatin immunoprecipitation assays that demonstrated enhanced p65 binding at 2/3 putative NFκB1 binding sites in the RhoC promoter. This binding pattern was unique to SUM149 cells. (**Figure 4A and B**)
- Established a 4.0kbp RhoC promoter reporter system. Importantly, transient transfections assays with this promoter reporter system demonstrated increased activity in the SUM149 cells, but not in MCF10A cells. This suggests that the RhoC promoter activity is deregulated in IBC leading to RhoC overexpression. (**Figure 4C**)
- Developed site mutants of RhoC promoter reporter system.
- Demonstrated that downregulation of p65 in IBC cells leads to loss of cell motility and invasion. (**Figure 5**)

**Specific Aim4:** ***To determine the distribution and stability of RhoC GTPase transcription variants in altering the half-life of the different mRNAs, thereby, regulating the total RhoC GTPase protein expression.***

- Established RhoC and GAPDH probes for northern blot analysis. This experiment demonstrated that RhoC mRNA decay is not differential between IBC and non-IBC control cell lines.

## Reportable outcomes:

- Published a manuscript detailing the methodology for identification of gene fusions in epithelial cancers, "Chimeric transcript discovery by paired-end transcriptome sequencing." (8)
- Published a review titled, "Translocations in epithelial cancers." (9)
- A manuscript was accepted for publication at *Mol. Cancer Res*, "RhoC Expression and Head and Neck Cancer Metastasis" (*In Press*)
- Completed a book chapter that was accepted for publication, "The Rho GTPases in Cancer" (*In Pres*s)
- Manuscript in preparation, "p65 drives RhoC GTPase expression and the metastatic phenotype in Inflammatory Breast Cancer"
- Research has led to an additional breast cancer grant that I co-authored through the Susan G. Komen foundation.

## Conclusions:

Since the submission of the original application and initiation of the DOD breast cancer training program, I have completed the core courses in Genetics, Biochemistry, Cell Biology and Ethics required by the University's CMB program as well as comprehensive courses in Cancer Biology, Pharmacology, Proteomics, Bioinformatics of Sequence Alignment and Mathematical Models in Biology. I have completed a comprehensive preliminary exam on a subject unrelated to this DOD award (my thesis project) as required by the CMB program. I have been first author or co-author on three manuscripts and one book chapter accepted for publication on work directly disseminating from this DOD Breast cancer award. Finally, I have co-authored a grant emanating from this research which was funded by the Susan G. Komen foundation.

## Figures

**Figure 1. Fluorescence in situ hybridization (FISH) using a *RHOC* locus probe.** Normal breast tissue is shown on the left as well as the IBC cell lines SUM149 and SUM190. An interphase spread of SUM190 is shown. In both SUM149 and SUM190 cells, three copies of chromosome 1 are present as confirmed by additional cytogenetic analysis using a centromeric probe for chromosome 1 leading to the additional copy of *RHOC*. Representative images are shown.

**Figure 2. ChIP-SEQ positive control analysis.** Chromatin immunoprecipitation-sequencing (ChIP-SEQ) using anti-ERα or anti-H3K4-tri-methylation antibodies on MCF7 or BT474 cells treated with or without 1nM 17-β-Estradiol as indicated. Plots show read accumulations in reads per kilobase million were aligned to the genome using HPEAK software as previously described (Yu et. al., 2010). Analysis of the *GREB1* locus reveals increased binding of ERα and H3K4-tri-methylation in both cell lines following stimulation with 17-β-Estradiol.

**Figure 3. p65 regulates RhoC mRNA expression in SUM149 cells.** Following the targeted shRNA screen, p65 was identified as a potential regulator of RhoC mRNA expression. QPCR analysis of SUM149 cells treated with p65 siRNA demonstrates that RhoC mRNA expression decreases with p65 knockdown. p65 knockdown was confirmed and IL6, a known target of p65, was used to demonstrate functional p65 knockdown. Importantly, the p65 siRNA did not alter p105 mRNA expression. Reactions were run in quadruplicate three times. Standard deviation is shown in the error bars.

**Figure 4. p65 binds to the *RHOC* promoter. A)** schematic shows putative p65 binding sites in the *RHOC* proximal promoter. **B)** ChIP analysis of p65 binding in HME, MCF10A, SUM149 and MDA-MB-231 cells demonstrates that p65 is enriched in the SUM149 cell line at p65 binding sites 1 and 3, but not in the control cell lines. **C)** *RHOC* promoter reporter activity demonstrates that the *RHOC* promoter, but not an empty vector control is highly active in the SUM149 cell line, but not in MCF10A cells. Data is shown relative to a renilla control used to normalize for transfections efficiency. All experiments were run in triplicate and standard deviation is shown on the bar plots.

**Figure 5. p65 expression is required for SUM149 cell motility and invasion. A)** Representative photomicrographs of cell motility assays in SUM149 cells treated with shRNA as

indicated. **B)** Quantification of cell motility assays. **C)** As in **A**, except boyden chamber transwell migration assays. Chambers were coated with 100μL matrigel 4 hours prior to seeding cells in serum free media. Forty eight hours later, representative images were taken to assess invasion through 8.0μM pores. Cells were stained with crystal violet. **D)** Quantification of cell invasion. Cells were released from the membrane with acetic acid and quantified by colorimetric analysis at 560nM. Percent maximum invasion is shown. All experiments were run in triplicate.

**Table 1. Analysis of microRNA array data.** MicroRNAs that were greater than two-fold down- or up-regulated were compared across cell lines to identify microRNAs that were recurrently differential among IBC cell lines, but not several other control cell lines.

## References


[1] O. H. Beahrs, J. R. McDonald, and W. G. Smith, *Annals of surgery* **145** (1), 115 (1957).

[2] C. G. Kleer, Y. Zhang, Q. Pan et al., *Breast Cancer Res* **6** (2), R110 (2004).

[3] K. L. van Golen, S. Davies, Z. F. Wu et al., *Clin Cancer Res* **5** (9), 2511 (1999).

[4] A. Hakem, O. Sanchez-Sweatman, A. You-Ten et al., *Genes & development* **19** (17), 1974 (2005).

[5] M. Wu, Z. F. Wu, C. Kumar-Sinha et al., *Breast cancer research and treatment* **84** (1), 3 (2004).

[6] E. Sahai and C. J. Marshall, *Nature reviews* **2** (2), 133 (2002).

[7] K. L. van Golen, Z. F. Wu, X. T. Qiao et al., *Cancer research* **60** (20), 5832 (2000).

[8] S. Valastyan, F. Reinhardt, N. Benaich et al., *Cell* **137** (6), 1032 (2009).

[9] C. A. Maher, N. Palanisamy, J. C. Brenner et al., *Proceedings of the National Academy of Sciences of the United States of America* **106** (30), 12353 (2009).

[10] J. C. Brenner and A. M. Chinnaiyan, *Biochimica et biophysica acta* **1796** (2), 201 (2009).

[11] M. Islam, G. Lin, J. C. Brenner et al., *Mol Cancer Res* **7** (11), 1771 (2009).

[12] J. Yu, J. Yu, R. S. Mani et al., *Cancer cell* **17** (5), 443.

# Chimeric transcript discovery by paired-end transcriptome sequencing

Christopher A. Maher[a,b], Nallasivam Palanisamy[a,b], John C. Brenner[a,b], Xuhong Cao[a,c], Shanker Kalyana-Sundaram[a,b], Shujun Luo[d], Irina Khrebtukova[d], Terrence R. Barrette[a,b], Catherine Grasso[a,b], Jindan Yu[a,b], Robert J. Lonigro[a,b], Gary Schroth[d], Chandan Kumar-Sinha[a,b], and Arul M. Chinnaiyan[a,b,c,e,f,1]

[a]Michigan Center for Translational Pathology, Ann Arbor, MI 48109; Departments of [b]Pathology and [e]Urology, University of Michigan, Ann Arbor, MI 48109; [c]Howard Hughes Medical Institute and [f]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109; and [d]Illumina Inc., 25861 Industrial Boulevard, Hayward, CA 94545

Communicated by David Ginsburg, University of Michigan Medical School, Ann Arbor, MI, May 4, 2009 (received for review March 16, 2009)

**Recurrent gene fusions are a prevalent class of mutations arising from the juxtaposition of 2 distinct regions, which can generate novel functional transcripts that could serve as valuable therapeutic targets in cancer. Therefore, we aim to establish a sensitive, high-throughput methodology to comprehensively catalog functional gene fusions in cancer by evaluating a paired-end transcriptome sequencing strategy. Not only did a paired-end approach provide a greater dynamic range in comparison with single read based approaches, but it clearly distinguished the high-level "driving" gene fusions, such as *BCR-ABL1* and *TMPRSS2-ERG*, from potential lower level "passenger" gene fusions. Also, the comprehensiveness of a paired-end approach enabled the discovery of 12 previously undescribed gene fusions in 4 commonly used cell lines that eluded previous approaches. Using the paired-end transcriptome sequencing approach, we observed read-through mRNA chimeras, tissue-type restricted chimeras, converging transcripts, diverging transcripts, and overlapping mRNA transcripts. Last, we successfully used paired-end transcriptome sequencing to detect previously undescribed ETS gene fusions in prostate tumors. Together, this study establishes a highly specific and sensitive approach for accurately and comprehensively cataloguing chimeras within a sample using paired-end transcriptome sequencing.**

bioinformatics | gene fusions | prostate cancer | breast cancer | RNA-Seq

One of the most common classes of genetic alterations is gene fusions, resulting from chromosomal rearrangements (1). Intriguingly, >80% of all known gene fusions are attributed to leukemias, lymphomas, and bone and soft tissue sarcomas that account for only 10% of all human cancers. In contrast, common epithelial cancers, which account for 80% of cancer-related deaths, can only be attributed to 10% of known recurrent gene fusions (2–4). However, the recent discovery of a recurrent gene fusion, *TMPRSS2-ERG*, in a majority of prostate cancers (5, 6), and *EML4-ALK* in non-small-cell lung cancer (NSCLC) (7), has expanded the realm of gene fusions as an oncogenic mechanism in common solid cancers. Also, the restricted expression of gene fusions to cancer cells makes them desirable therapeutic targets. One successful example is imatinib mesylate, or Gleevec, that targets *BCR-ABL1* in chronic myeloid leukemia (CML) (8–10). Therefore, the identification of novel gene fusions in a broad range of cancers is of enormous therapeutic significance.

The lack of known gene fusions in epithelial cancers has been attributed to their clonal heterogeneity and to the technical limitations of cytogenetic analysis, spectral karyotyping, FISH, and microarray-based comparative genomic hybridization (aCGH). Not surprisingly, *TMPRSS2-ERG* was discovered by circumventing these limitations through bioinformatics analysis of gene expression data to nominate genes with marked overexpression, or outliers, a signature of a fusion event (6). Building on this success, more recent strategies have adopted unbiased high-throughput approaches, with increased resolution, for genome-wide detection of chromosomal rearrangements in cancer involving BAC end sequencing (11), fosmid paired-end sequences (12), serial analysis of gene expression (SAGE)-like sequencing (13), and next-generation DNA sequencing (14). Despite unveiling many novel genomic rearrangements, solid tumors accumulate multiple nonspecific aberrations throughout tumor progression; thus, making causal and driver aberrations indistinguishable from secondary and insignificant mutations, respectively.

The deep unbiased view of a cancer cell enabled by massively parallel transcriptome sequencing has greatly facilitated gene fusion discovery. As shown in our previous work, integrating long and short read transcriptome sequencing technologies was an effective approach for enriching "expressed" fusion transcripts (15). However, despite the success of this methodology, it required substantial overhead to leverage 2 sequencing platforms. Therefore, in this study, we adopted a single platform paired-end strategy to comprehensively elucidate novel chimeric events in cancer transcriptomes. Not only was using this single platform more economical, but it allowed us to more comprehensively map chimeric mRNA, hone in on driver gene fusion products due to its quantitative nature, and observe rare classes of transcripts that were overlapping, diverging, or converging.

## Results

**Chimera Discovery via Paired-End Transcriptome Sequencing.** Here, we employ transcriptome sequencing to restrict chimera nominations to "expressed sequences," thus, enriching for potentially functional mutations. To evaluate massively parallel paired-end transcriptome sequencing to identify novel gene fusions, we generated cDNA libraries from the prostate cancer cell line VCaP, CML cell line K562, universal human reference total RNA (UHR; Stratagene), and human brain reference (HBR) total RNA (Ambion). Using the Illumina Genome Analyzer II, we generated 16.9 million VCaP, 20.7 million K562, 25.5 million UHR, and 23.6 million HBR transcriptome mate pairs ($2 \times 50$ nt). The mate pairs were mapped against the transcriptome and categorized as (*i*) mapping to same gene, (*ii*) mapping to different genes (chimera candidates), (*iii*) nonmapping, (*iv*) mitochondrial, (*v*) quality control, or (*vi*) ribosomal (Table S1). Overall, the chimera candidates represent a minor fraction of the mate pairs, comprising $\approx$<1% of the reads for each sample.

We believe that a paired-end strategy offers multiple advantages over single read based approaches such as alleviating the reliance on sequencing the reads traversing the fusion junction, increased coverage provided by sequencing reads from the ends of a tran-

Author contributions: C.A.M. and A.M.C. designed research; C.A.M., N.P., J.C.B., X.C., S.L., I.K., T.R.B., R.J.L., G.S., C.K.-S., and A.M.C. performed research; C.A.M., S.L., I.K., R.J.L., and G.S. contributed new reagents/analytic tools; C.A.M., N.P., J.C.B., S.K.-S., C.G., J.Y., R.J.L., G.S., C.K.-S., and A.M.C. analyzed data; and C.A.M., N.P., X.C., C.K.-S., and A.M.C. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: arul@umich.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0904720106/DCSupplemental.

CELL BIOLOGY

scribed fragment, and the ability to resolve ambiguous mappings (Fig. S1). Therefore, to nominate chimeras, we leveraged each of these aspects in our bioinformatics analysis. We focused on both mate pairs encompassing and/or spanning the fusion junction by analyzing 2 main categories of sequence reads: chimera candidates and nonmapping (Fig. S2*A*). The resulting chimera candidates from the nonmapping category that span the fusion boundary were merged with the chimeras found to encompass the fusion boundary revealing 119, 144, 205, and 294 chimeras in VCaP, K562, HBR, and UHR, respectively.

**Comparison of a Paired-End Strategy Against Existing Single Read Approaches.** To assess the merit of adopting a paired-end transcriptome approach, we compared the results against existing single read approaches. Although current RNA sequencing (RNA-Seq) studies have been using 36-nt single reads (16, 17), we increased the likelihood of spanning a fusion junction by generating 100-nt long single reads using the Illumina Genome Analyzer II. Also, we chose this length because it would facilitate a more comparable amount of sequencing time as required for sequencing both 50-nt mate pairs. In total, we generated 7.0, 59.4, and 53.0 million 100-nt transcriptome reads for VCaP, UHR, and HBR, respectively, for comparison against paired-end transcriptome reads from matched samples.

Because the UHR is a mixture of cancer cell lines, we expected to find numerous previously identified gene fusions. Therefore, we first assessed the depth of coverage of a paired-end approach against long single reads by directly comparing the normalized frequency of sequence reads supporting 4 previously identified gene fusions [*TMPRSS2-ERG* (5, 6), *BCR-ABL1* (18), *BCAS4-BCAS3* (19), and *ARFGEF2-SULF2* (20)]. As shown in Fig. 1*A*, we observed a marked enrichment of paired-end reads compared with long single reads for each of these well characterized gene fusions.

We observed that *TMPRSS2-ERG* had a >10-fold enrichment between paired-end and single read approaches. The schematic representation in Fig. 1*B* indicates the distribution of reads confirming the *TMPRSS2-ERG* gene fusion from both paired-end and single read sequencing. As expected, the longer reads improve the number of reads spanning known gene fusions. For example, had we sequenced a single 36-mer (shown in red text), 11 of the 17 chimeras, shown in the bottom portion of the long single reads, would not have spanned the gene fusion boundary, but instead, would have terminated before the junction and, therefore, only aligned to *TMPRSS2*. However, despite the improved results only 17 chimeric reads were generated from 7.0 million long single read sequences. In contrast, paired-end sequencing resulted in 552 reads supporting the *TMPRSS2-ERG* gene fusion from ≈17 million sequences.

Because we are using sequence based evidence to nominate a chimera, we hypothesized that the approach providing the maximum nucleotide coverage is more likely to capture a fusion junction. We calculated an *in silico* insert size for each sample using mate pairs aligning to the same gene, and found the mean insert size of ≈200 nt. Then, we compared the total coverage from single reads (coverage is equivalent to the total number of pass filter reads against the read length) with the paired-end approach (coverage is equivalent to the sum of the insert size with the length of each read) (Fig. S2*B*). Overall, we observed an average coverage of 848.7 and 757.3 MB using single read technology, compared with 2,553.3 and 2,363 MB from paired-end in UHR and HBR, respectively. This increase in ≈3-fold coverage in the paired-end samples compared with the long read approach, per lane, could explain the increased dynamic range we observed using a paired-end strategy.

Next we wanted to identify chimeras common to both strategies. The long read approach nominated 1,375 and 1,228 chimeras, whereas with a paired-end strategy, we only nominated 225 and 144 chimeras in UHR and HBR, respectively. As shown in the Venn diagram (Fig. 1*C*), there were 32 and 31 candidates common to both technologies for UHR and HBR, respectively. Within the common UHR chimeric candidates, we observed previously identified gene fusions *BCAS4-BCAS3*, *BCR-ABL1*, *ARFGEF2-SULF2*, and *RPS6KB1-TMEM49* (13). The remaining chimeras, nominated by both approaches, represent a high fidelity set. Therefore, to further assess whether a paired-end strategy has an increased dynamic range, we compared the ratio of normalized mate pair reads against single reads for the remaining chimeras common to both technologies. We observed that 93.5 and 93.9% of UHR and HBR candidates, respectively, had a higher ratio of normalized mate pair reads to single reads (Table S2), confirming the increased dynamic range offered by a paired-end strategy. We hypothesize that the greater number of nominated candidates specific to the long read approach represents an enrichment of false positives, as observed when using the 454 long read technology (15, 21).

**Paired-End Approach Reveals Novel Gene Fusions.** We were interested in determining whether the paired-end libraries could detect novel gene fusions. Among the top chimeras nominated from VCaP, HBR, UHR, and K562, many were already known, including *TMPRSS2-ERG*, *BCAS4-BCAS3*, *BCR-ABL1*, *USP10-ZDHHC7*, and *ARFGEF2-SULF2*. Also ranking among these well known gene fusions in UHR was a fusion on chromosome 13 between *GAS6* and *RASA3* (Fig. S3*A* and Table S2). The fact that *GAS6-RASA3* ranked higher than *BCR-ABL1* suggests that it may be a driving fusion in one of the cancer cell lines in the RNA pool.

Another observation was that there were 2 candidates among the top 10 found in both UHR and K562. This observation was intriguing, because hematological malignancies are not considered to have multiple gene fusion events. In addition to *BCR-ABL1*, we were able to detect a previously undescribed interchromosomal gene fusion between exon 23 of *NUP214* located at chromosome 9q34.13 with exon 2 of *XKR3* located at chromosome 22q11.1. Both of these genes reside on chromosome 22 and 9 in close proximity to *BCR* and *ABL1*, respectively (Fig. S3*B*). We confirmed the presence of *NUP214-XKR3* in K562 cells using qRT-PCR, but were unable to detect it across an additional 5 CML cell lines tested (SUP-B15, MEG-01, KU812, GDM-1, and Kasumi-4) (Fig. S3*C*). These results suggest that *NUP214-XKR3* is a "private" fusion that originated from additional complex rearrangements after the translocation that generated *BCR-ABL1* and a focal amplification of both gene regions.

Although we were able to detect *BCR-ABL1* and *NUP214-XKR3* in both UHR and K562, there was a marked reduction in the mate pairs supporting these fusions in UHR. Although a diluted signal is expected, because UHR is pooled samples, it provides evidence that pooling samples can serve as a useful approach for nominating top expressing chimeras, and potentially enrich for "driver" chimeras.

**Previously Undescribed Prostate Gene Fusions.** Our previous work using integrative transcriptome sequencing to detect gene fusions in cancer revealed multiple gene fusions, demonstrating the complexity of the prostate transcriptomes of VCaP and LNCaP (15). Here, we exploit the comprehensiveness of a paired-end strategy on the same cell lines to reveal novel chimeras. In the circular plot shown in Fig. S4*A*, we displayed all experimentally validated paired-end chimeras in the larger red circle. We found that all of the previously discovered chimeras in VCaP and LNCaP comprised a subset of the paired-end candidates, as displayed in the inner black circle.

As expected, *TMPRSS2-ERG* was the top VCaP candidate. In addition to "rediscovering" the *USP10-ZDHHC7*, *HJURP-INPP4A*, and *EIF4E2-HJURP* gene fusions, a paired-end approach revealed several previously undescribed gene fusions in VCaP. One such example was an interchromosomal gene fusion between *ZDHHC7*, on chromosome 16, with *ABCB9*, residing on chromosome 12, that was validated by qRT-PCR (Fig. S3*D*). Interestingly, the 5′ partner, *ZDHHC7*, had previously been validated as a complex intrachro-

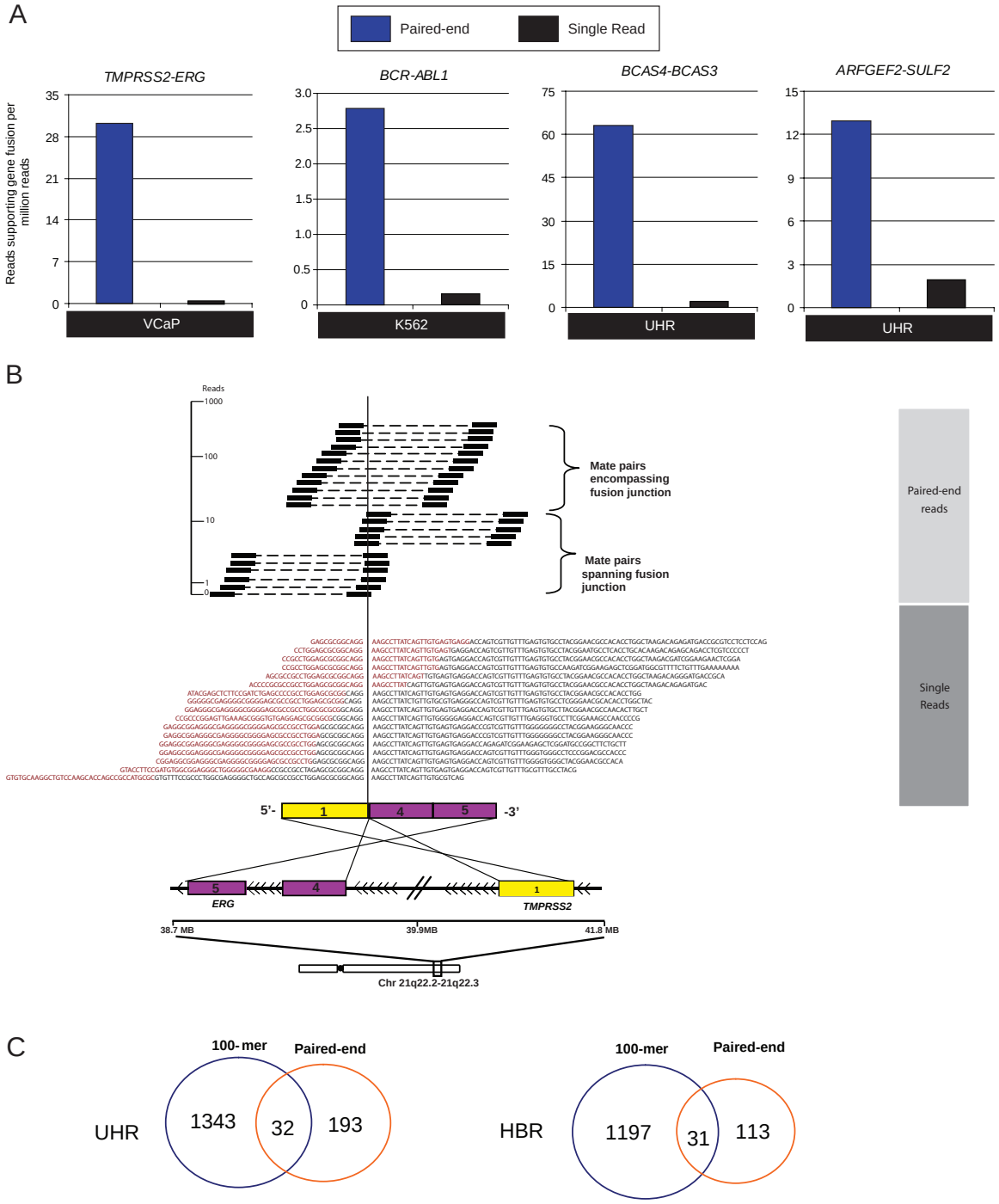



**Fig. 1.** Dynamic range and sensitivity of the paired-end transcriptome analysis relative to single read approaches. (*A*) Comparison of paired-end (blue) and long single transcriptome reads (black) supporting known gene fusions *TMPRSS2-ERG*, *BCR-ABL1*, *BCAS4-BCAS3*, and *ARFGEF2-SULF2*. (*B*) Schematic representation of *TMPRSS2-ERG* in VCaP, comparing mate pairs with long single transcriptome reads. (*Upper*) Frequency of mate pairs, shown in log scale, are divided based on whether they encompass or span the fusion boundary; (*Lower*) 100-mer single transcriptome reads spanning *TMPRSS2-ERG* fusion boundary. First 36 nt are highlighted in red. (*C*) Venn diagram of chimera nominations from both a paired-end (orange) and long single read (blue) strategy for UHR and HBR.

mosomal gene fusion with *USP10* (15). Both fusions have mate pairs aligning to the same exon of *ZDHHC7* (15), suggesting that their breakpoints are in adjacent introns (Fig. S3*D*).

Another previously undescribed VCaP interchromosomal gene fusion that we discovered was between exon 2 of *TIA1*, residing on chromosome 2, with exon 3 of *DIRC2*, or disrupted in renal carcinoma 2, located on chromosome 3. *TIA1-DIRC2* was validated by qRT-PCR and FISH (Fig. S5). In total, we confirmed an additional 4 VCaP and 2 LNCaP chimeras (Fig. S6). Overall, these fusions demonstrate that paired-end transcriptome sequencing can nominate candidates that have eluded previous techniques, including other massively parallel transcriptome sequencing approaches.

**Distinguishing Causal Gene Fusions from Secondary Mutations.** We were next interested in determining whether the dynamic range provided by paired-end sequencing can distinguish known high-

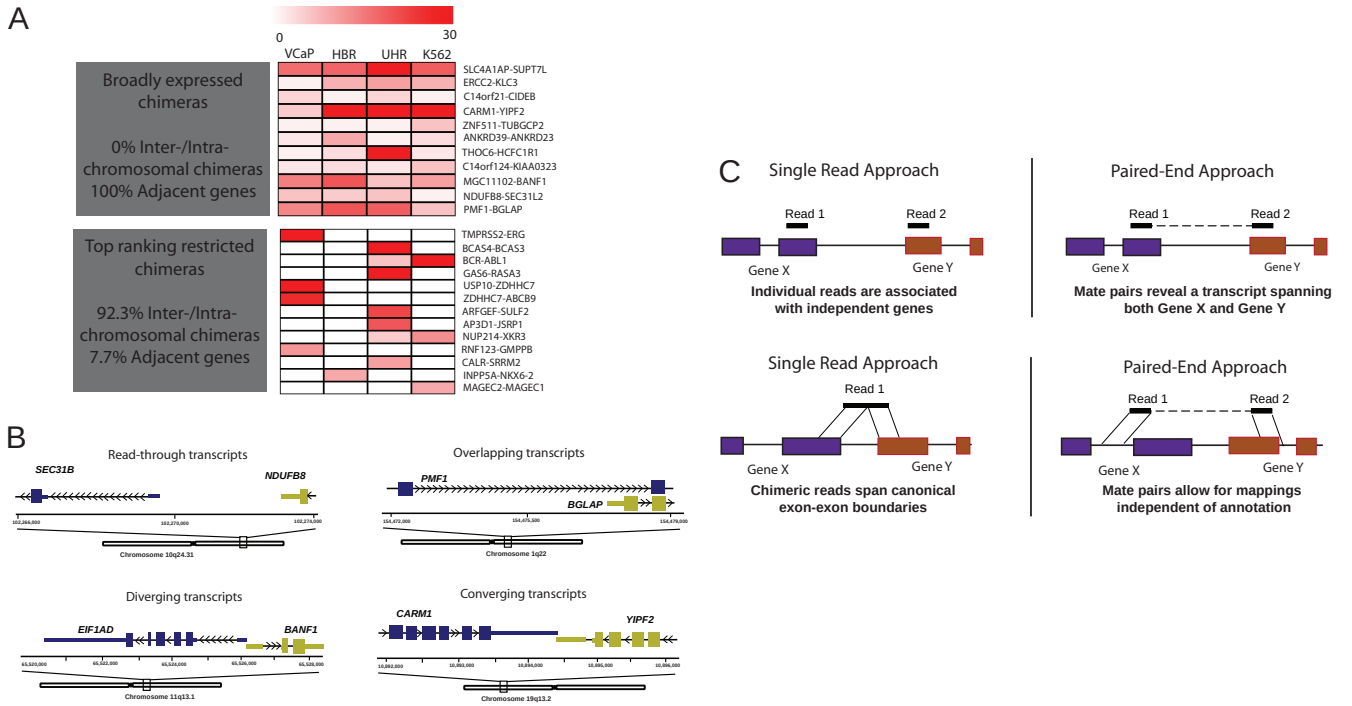



**Fig. 2.** RNA based chimeras. (*A*) Heatmaps showing the normalized number of reads supporting each read-through chimera across samples ranging from 0 (white) to 30 (red). (*Upper*) The heatmap highlights broadly expressed chimeras in UHR, HBR, VCaP, and K562. (*Lower*) The heatmap highlights the expression of the top ranking restricted gene fusions that are enriched with interchromosomal and intrachromosomal rearrangements. (*B*) Illustrative examples classifying RNA-based chimeras into (*i*) read-throughs, (*ii*) converging transcripts, (*iii*) diverging transcripts, and (*iv*) overlapping transcripts. (*C Upper*) Paired-end approach links reads from independent genes as belonging to the same transcriptional unit (*Right*), whereas a single read approach would assign these reads to independent genes (*Left*). (*Lower*) The single read approach requires that a chimera span the fusion junction (*Left*), whereas a paired-end approach can link mate pairs independent of gene annotation (*Right*).

level "driving" gene fusions, such as known recurrent gene fusions *BCR-ABL1* and *TMPRSS2-ERG*, from lower level "passenger" fusions. Therefore, we plotted the normalized mate pair coverage at the fusion boundary for all experimentally validated gene fusions for the 2 cell lines that we sequenced harboring recurrent gene fusions, VCaP and K562. As shown in Fig. S4*B*, we observed that both driver fusions, *TMPRSS2-ERG* and *BCR-ABL1*, show the highest expression among the validated chimeras in VCaP and K562, respectively. This observation suggests a paired-end nomination strategy for selecting putative driver gene fusions among private nonspecific gene fusions that lack detectable levels of expression across a panel of samples (15).

**Previously Undescribed Breast Cancer Gene Fusions.** Our ability to detect previously undescribed prostate gene fusions in VCaP and LNCaP demonstrated the comprehensiveness of paired-end transcriptome sequencing compared with an integrated approach, using short and long transcriptome reads. Therefore, we extended our paired-end analysis by using breast cancer cell line MCF-7, which has been mined for fusions using numerous approaches such as expressed sequence tags (ESTs) (22), array CGH (23), single nucleotide polymorphism arrays (24), gene expression arrays (25), end sequence profiling (20, 26), and paired-end diTag (PET) (13).

A histogram (Fig. S4*C*) of the top ranking MCF-7 candidates highlights *BCAS4-BCAS3* and *ARFGEF-SULF2* as the top 2 ranking candidates, whereas other previously reported candidates, such as *SULF2-PRICKLE*, *DEPDC1B-ELOVL7*, *RPS6KB1-TMEM49*, and *CXorf15-SYAP1*, were interspersed among a comprehensive list of previously undescribed putative chimeras. To confirm that these previously undescribed nominations were not false positives, we experimentally validated 2 interchromosomal and 3 intrachromosomal candidates using qRT-PCR (Fig. S6). Overall, not only was a paired-end approach able to detect gene fusions that have eluded numerous existing technologies, it has revealed 5 previously undescribed mutations in breast cancer.

**RNA-Based Chimeras.** Although many of the inter and intrachromosomal rearrangements that we nominated were found within a single sample, we observed many chimeric events shared across samples. We identified 11 chimeric events common to UHR, VCaP, K562, and HBR (Table S3). Via heatmap representation (Fig. 2*A*) of the normalized frequency of mate pairs supporting each chimeric event, we can observe these events are broadly transcribed in contrast to the top restricted chimeric events. Also, we found that 100% of the broadly expressed chimeras resided adjacent to one another on the genome, whereas only 7.7% of the restricted candidates were neighboring genes. This discrepancy can be explained by the enrichment of inter and intrachromosomal rearrangements in the restricted set.

Unlike, previously characterized restricted read-throughs, such as *SLC45A3-ELK4* (15), which are found adjacent to one another, but in the same orientation, we found that the majority of the broadly expressed chimera candidates resided adjacent to one another in different orientations. Therefore, we have categorized these events as (*i*) read-throughs, adjacent genes in the same orientation, (*ii*) diverging genes, adjacent genes in opposite orientation whose 5′ ends are in close proximity, (*iii*) convergent genes, adjacent genes in opposite orientation whose 3′ ends are in close proximity, and (*iv*) overlapping genes, adjacent genes who share common exons (Fig. 2*B*). Based on this classification, we found 1 read-through, 2 convergent genes, 6 divergent genes, and 2 overlapping genes. Also, we found that ≈81.8% of these chimeras had at least 1 supporting EST, providing independent confirmation of the event (Table S3). In contrast to paired-end, single read ap-

proaches would likely miss these instances as each mate would have aligned to their respective genes based on the current annotations (Fig. 2*C*). Also, these instances may represent extensions of a transcriptional unit, which would not be detectable by a single read approach that identifies chimeric reads that span exon boundaries of independent genes. Overall, we believe that many of these broadly expressed RNA chimeras represent instances where mate pairs are revealing previously undescribed annotation for a transcriptional unit.

**Previously Undescribed ETS Gene Fusions in Clinically Localized Prostate Cancer.** Given the high prevalence of gene fusions involving ETS oncogenic transcription factor family members in prostate tumors, we applied paired-end transcriptome sequencing for gene fusion discovery in prostate tumors lacking previously reported ETS fusions. For 2 prostate tumors, aT52 and aT64, we generated 6.2 and 7.4 million transcriptome mate pairs, respectively. In aT64, we found that *HERPUD1*, residing on chromosome 16, juxtaposed in front of exon 4 of *ERG* (Fig. 3*A*), which was validated by qRT-PCR (Fig. S6) and FISH (Fig. 3*B*), thus identifying a third 5′ fusion partner for *ERG*, after *TMPRSS2* (6) and *SLC45A3* (27), and presumably, *HERPUD1* also mediates the overexpression of ERG in a subset of prostate cancer patients. Also, just as *TMPRSS2* and *SLC45A3* have been shown to be androgen regulated by qRT-PCR (5), we found *HERPUD1* expression, via RNA-Seq, to be responsive to androgen treatment (Fig. S7). Also, ChIP-Seq analysis revealed androgen binding at the 5′ end of *HERPUD1* (Fig. S7).

Also, in the second prostate tumor sample (aT52), we discovered an interchromosomal gene fusion between the 5′ end of a prostate cDNA clone, *AX747630* (*FLJ35294*), residing on chromosome 17, with exon 4 of *ETV1*, located on chromosome 7 (Fig. 3*C*), which was validated via qRT-PCR (Fig. S6) and FISH (Fig. 3*D*). Interestingly, this fusion has previously been reported in an independent sample found by a fluorescence in situ hybridization screen (27); thus, demonstrating that it is recurrent in a subset of prostate cancer patients. As previously reported, gene expression via RNA-Seq confirmed that *AX747630* is an androgen-inducible gene (Fig. S7). Also, ChIP-Seq revealed androgen occupancy at the 5′ end of *AX747630* (Fig. S7).

## Discussion

This study demonstrates the effectiveness of paired-end massively parallel transcriptome sequencing for fusion gene discovery. By using a paired-end approach, we were able to rediscover known gene fusions, comprehensively discover previously undescribed gene fusions, and hone in on causal gene fusions. The ability to detect 12 previously undescribed gene fusions in 4 commonly used cell lines that eluded any previous efforts conveys the superior sensitivity of a paired-end RNA-Seq strategy compared with existing approaches. Also, it suggests that we may be able to unveil previously undescribed chimeric events in previously characterized samples believed to be devoid of any known driver gene fusions as exemplified by the discovery of previously undescribed ETS gene fusions in 2 clinically localized prostate tumor samples that lacked known driver gene fusions.

By analyzing the transcriptome at unprecedented depth, we have revealed numerous gene fusions, demonstrating the prevalence of a relatively under-represented class of mutations. However, one of the major goals remains to discover recurrent gene fusions and to distinguish them from secondary, nonspecific chimeras. Although quantifying expression levels is not proof of whether a gene fusion is a driver or passenger, because a low-level gene fusion could still be causative, it still of major significance that a paired-end strategy clearly distinguished known high-level driving gene fusions, such as *BCR-ABL1* and *TMPRSS2-ERG*, from potential lower level passenger chimeras. Overall, these fusions serve as a model for employing a paired-end nomination strategy for prioritizing leads likely to be high-level driving gene fusions, which would subsequently undergo further functional and experimental evaluation.

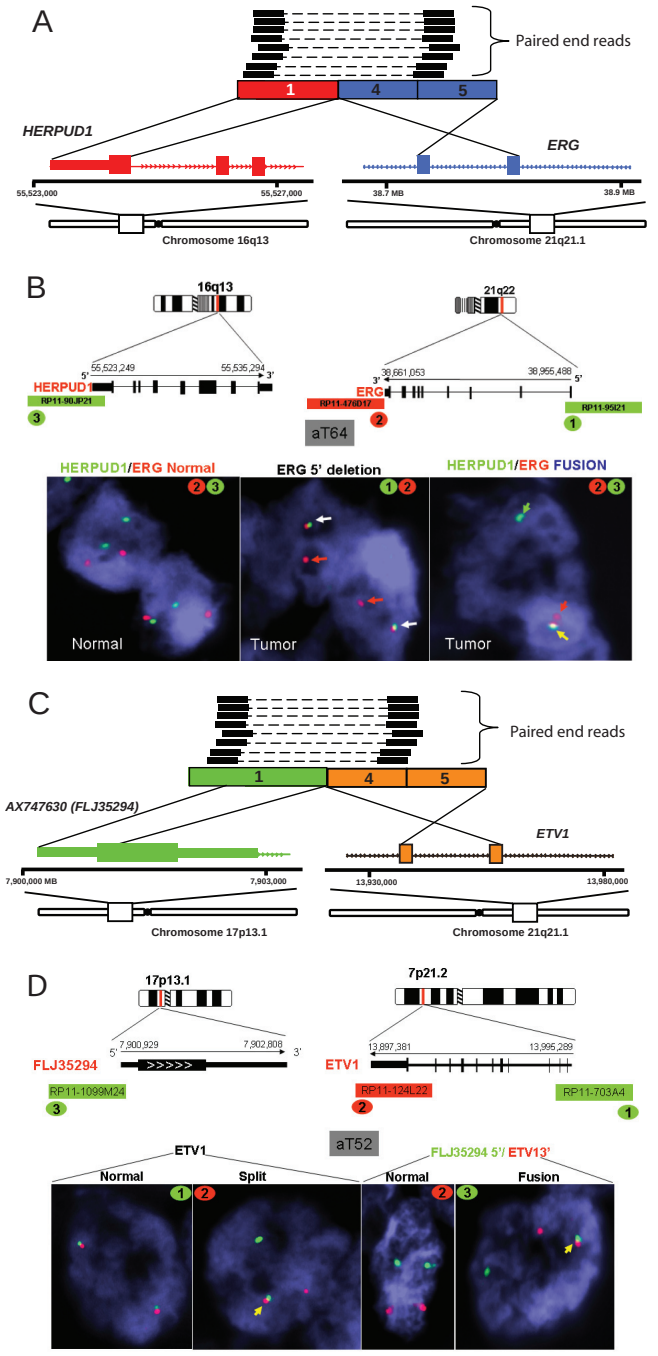



**Fig. 3.** Discovery of previously undescribed ETS gene fusions in localized prostate cancer. (*A*) Schematic representation of the interchromosomal gene fusion between exon 1 of *HERPUD1* (red), residing on chromosome 16, with exon 4 of *ERG* (blue), located on chromosome 21. (*B*) Schematic representation showing genomic organization of *HERPUD1* and *ERG* genes. Horizontal red and green bars indicate the location of BAC clones. (*Lower*) FISH analysis using BAC clones showing *HERPUD1* and *ERG* in a normal tissue (*Left*), deletion of the ERG 5′ region in tumor (*Center*), and *HERPUD1-ERG* fusion in a tumor sample (*Right*). (*C*) Schematic representation of the interchromosomal gene fusion between *FLJ35294* (green), residing on chromosome 17, with exon 4 of *ETV1* (orange) located on chromosome 21. (*D Upper*) Schematic representation of the genomic organization of *FLJ35294* and *ETV1* genes. (*Lower*) FISH analysis using BAC clones showing split of *ETV1* in tumor sample (*Left*) and the colocalization of *FLJ35294* and *ETV1* in a tumor sample (*Right*).

CELL BIOLOGY

One of the major advantages of using a transcriptome approach is that it enables us to identify rearrangements that are not detectable at the DNA level. For example, conventional cytogenetic methods would miss gene fusions produced by paracentric inversions, or sub microscopic events, such as *GAS6-RASA3*. Also, transcriptome sequencing can unveil RNA chimeras, lacking DNA aberrations, as demonstrated by the discovery of a recurrent, prostate specific, read-through of *SLC45A3* with *ELK4* in prostate cancers. Further classification of RNA based events using paired-end sequencing revealed numerous broadly expressed chimeras between adjacent genes. Although these events were not necessarily read-throughs events, because they typically had different orientations, we believe they represent extensions of transcriptional units beyond their annotated boundaries. Unlike single read based approaches, which require chimeras to span exon boundaries of independent genes, we were able to detect these events using paired-end sequencing, which could have significant impact for improving how we annotate transcriptional units.

Overall, we have demonstrated the advantages of employing a paired-end transcriptome strategy for chimera discovery, established a methodology for mining chimeras, and extensively catalogued chimeras in a prostate and hematological cancer models. We believe that the sensitivity of this approach will be of broad impact and significance for revealing novel causative gene fusions in various cancers while revealing additional private gene fusions that may contribute to tumorigenesis or cooperate with driver gene fusions.

## Methods

**Paired-End Gene Fusion Discovery Pipeline.** Mate pair transcriptome reads were mapped to the human genome (hg18) and Refseq transcripts, allowing up to 2 mismatches, using Efficient Alignment of Nucleotide Databases (ELAND) pair within the Illumina Genome Analyzer Pipeline software. Illumina export output files were parsed to categorize passing filter mate pairs as (*i*) mapping to the same transcript, (*ii*) ribosomal, (*iii*) mitochondrial, (*iv*) quality control, (*v*) chimera candidates, and (*vi*) nonmapping. Chimera candidates and nonmapping categories were used for gene fusion discovery. For the chimera candidates category, the following criteria were used: (*i*) mate pairs must be of high mapping quality (best unique match across genome), (*ii*) best unique mate pairs do not have a more logical alternative combination (i.e., best mate pairs suggest an interchromosomal rearrangement, whereas the second best mapping for a mate reveals the pair have a alignment within the expected insert size), (*iii*) the sum of the distances between the most 5′ and 3′ mate on both partners of the gene fusion must be $<$500 nt, and (*iv*) mate pairs supporting a chimera must be nonredundant.

In addition to mining mate pairs encompassing a fusion boundary, the nonmapping category was mined for mate pairs that had 1 read mapping to a gene, whereas its corresponding read fails to align, because it spans the fusion boundary. First, the annotated transcript that the ''mapping'' mate pair aligned against was extracted, because this transcript represents one of the potential partners involved in the gene fusion. The ''nonmapping'' mate pair was then aligned against all of the exon boundaries of the known gene partner to identify a perfect partial alignment. A partial alignment confirms that the nonmapping mate pair maps to our expected gene partner while revealing the portion of the nonmapping mate pair, or overhang, aligning to the unknown partner. The overhang is then aligned against the exon boundaries of all known transcripts to identify the fusion partner. This process is done using a Perl script that extracts all possible University of California Santa Cruz (UCSC) and Refseq exon boundaries looking for a single perfect best hit.

Mate pairs spanning the fusion boundary are merged with mate pairs encompassing the fusion boundary. At least 2 independent mate pairs are required to support a chimera nomination, which can be achieved by (*i*) 2 or more nonredundant mate pairs spanning the fusion boundary, (*ii*) 2 or more nonredundant mate pairs encompassing a fusion boundary, or (*iii*) 1 or more mate pairs encompassing a fusion boundary and 1 or more mate pairs spanning the fusion boundary. All chimera nominations were normalized based on the cumulative number of mate pairs encompassing or spanning the fusion junction per million mate pairs passing filter.

**RNA Chimera Analysis.** Chimeras found from UHR, HBR, VCaP, and K562 were grouped based on whether they showed expression in all samples, ''broadly expressed,'' or a single sample, ''restricted expression.'' Because UHR is comprised of K562, chimeras found in only these 2 samples were also considered as restricted. Heatmap visualization was conducted by using TIGR's MultiExperiment Viewer (TMeV) version 4.0 (www.tm4.org).

**Additional Details.** Additional details can be found in *SI Text*.

**ACKNOWLEDGMENTS.** We thank Lu Zhang, Eric Vermaas, Victor Quijano, and Juying Yan for assistance with sequencing, Shawn Baker and Steffen Durinck for helpful discussions, Rohit Mehra and Javed Siddiqui for collecting tissue samples, and Bo Han and Kalpana Ramnarayanan for technical assistance. C.A.M. was supported by a National Institutes of Health (NIH) Ruth L. Kirschstein postdoctoral training grant, and currently derives support from the American Association of Cancer Research Amgen Fellowship in Clinical/Translational Research and the Canary Foundation and American Cancer Society Early Detection Postdoctoral Fellowship. J.Y. was supported by NIH Grant 1K99CA129565-01A1 and Department of Defense (DOD) Grant PC080665. A.M.C. was supported in part by the NIH (Prostate SPORE P50CA69568, R01 R01CA132874), the DOD (BC075023, W81XWH-08-0110), the Early Detection Research Network (U01 CA111275), a Burroughs Welcome Foundation Award in Clinical Translational Research, a Doris Duke Charitable Foundation Distinguished Clinical Investigator Award, and the Howard Hughes Medical Institute. This work was also supported by National Center for Integrative Biomedical Informatics Grant U54 DA021519.

1. Futreal PA, et al. (2004) A census of human cancer genes. *Nat Rev* 4:177–183.
2. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM (2008) Recurrent gene fusions in prostate cancer. *Nat Rev* 8:497–511.
3. Mitelman F, Johansson B, Mertens F (2004) Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nat Genet* 36:331–334.
4. Mitelman F, Mertens F, Johansson B (2005) Prevalence estimates of recurrent balanced cytogenetic aberrations and gene fusions in unselected patients with neoplastic disorders. *Gene Chromosome Canc* 43:350–366.
5. Tomlins SA, et al. (2007) Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* 448:595–599.
6. Tomlins SA, et al. (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310:644–648.
7. Soda M, et al. (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448:561–566.
8. Druker BJ, et al. (2006) Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *New Engl J Med* 355:2408–2417.
9. Druker BJ, et al. (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med* 2:561–566.
10. Kantarjian H, et al. (2002) Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *New Engl J Med* 346:645–652.
11. Volik S, et al. (2003) End-sequence profiling: Sequence-based analysis of aberrant genomes. *Proc Natl Acad Sci USA* 100:7696–7701.
12. Tuzun E, et al. (2005) Fine-scale structural variation of the human genome. *Nat Genet* 37:727–732.
13. Ruan Y, et al. (2007) Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs). *Genome Res* 17:828–838.
14. Campbell PJ, et al. (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 40:722–729.
15. Maher CA, et al. (2009) Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458:97–101.
16. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18:1509–1517.
17. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628.
18. Shtivelman E, Lifshitz B, Gale RP, Canaani E (1985) Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* 315:550–554.
19. Barlund M, et al. (2002) Cloning of BCAS3 (17q23) and BCAS4 (20q13) genes that undergo amplification, overexpression, and fusion in breast cancer. *Gene Chromosome Canc* 35:311–317.
20. Hampton OA, et al. (2009) A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. *Genome Res* 19:167–177.
21. Zhao Q, et al. (2009) Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line. *Proc Natl Acad Sci USA* 106:1886–1891.
22. Hahn Y, et al. (2004) Finding fusion genes resulting from chromosome rearrangement by analyzing the expressed sequence databases. *Proc Natl Acad Sci USA* 101:13257–13261.
23. Shadeo A, Lam WL (2006) Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res* 8:R9.
24. Huang J, et al. (2004) Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum Genom* 1:287–299.
25. Neve RM, et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 10:515–527.
26. Volik S, et al. (2006) Decoding the fine-scale structure of a breast cancer genome and transcriptome. *Genome Res* 16:394–404.
27. Han B, et al. (2008) A fluorescence in situ hybridization screen for E26 transformation-specific aberrations: Identification of DDX5-ETV4 fusion protein in prostate cancer. *Cancer Res* 68:7629–7637.

# RhoC Expression and Head and Neck Cancer Metastasis

**Mozaffarul Islam,[1] Giant Lin,[2] John C. Brenner,[3] Quintin Pan,[1] Sofia D. Merajver,[4] Yanjun Hou,[1] Pawan Kumar,[1] and Theodoros N. Teknos[1]**

*[1]Department of Otolaryngology-Head and Neck Surgery and Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio; [2]Department of Otolaryngology-Head and Neck Surgery, [3]Department of Pathology, Program in Cellular and Molecular Biology, and [4]Department of Internal Medicine and Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan*

## Abstract

**RhoC protein, a known marker of metastases in aggressive breast cancers and melanoma, has also been found to be overexpressed in certain head and neck cancers, thus we investigated the correlation between RhoC expression and the metastatic behavior of head and neck squamous cell carcinoma. Selective inhibition of RhoC expression was achieved using lentiviral small hairpin RNA (shRNA) transduced and tracked with green fluorescent protein to achieve 70% to 80% RhoC inhibition. Fluorescence microscopy of the RhoC knockdown stable clones showed strong green fluorescence in the majority of cells, signifying a high efficiency of transduction. Importantly, quantitative real-time PCR showed no significant decrease in the mRNA expression levels of other members of the Ras superfamily. Cell motility and invasion were markedly diminished in RhoC-depleted cell lines as compared with control transduced lines. H&E staining of lung tissue obtained from severe combined immunodeficiency mice, which had been implanted with RhoC knockdown cells, showed a marked decrease in lung metastasis and inflammation of the blood vessels. The cultured lung tissue showed a significant decrease in cell growth in mice implanted with RhoC-depleted cell lines as compared with shRNA-scrambled sequence control lines. Microscopic studies of CD31 expression revealed substantial quantitative and qualitative differences in the primary tumor microvessel density as compared with parental and shRNA-scrambled controls. This study is the first of its kind to establish the involvement of RhoC specifically in head and neck metastasis. These findings suggest that RhoC warrants further investigation to delineate its robustness as a novel potentially therapeutic target. (Mol Cancer Res 2009;7(11):1771–80)**

## Introduction

Head and neck cancer is the sixth most common cancer worldwide (1). According to the statistical report of the American Cancer Society, ~70,000 new head and neck squamous cell carcinomas (HNSCC) will be diagnosed this year in the United States (2). In contrast to other epithelial cancers for which effective screening exists, most of the patients with head and neck cancer are diagnosed at a very late stage (stage III and IV). Despite advancements in surgical procedures, chemotherapy, and radiation therapy, survival rates have not improved in the last several decades (3). Furthermore, it has been shown that the high rate of morbidity is due to both locoregional recurrence and distant metastases.

In the past decade, numerous studies have shown that the Rho family of GTPases (RhoA, RhoB, RhoC, Rac1, Rac2, Rac3, and CDC42) is involved in instilling a metastatic phenotype into localized cancerous cells that are localized to the organ of origin. RhoA and RhoC are overexpressed in a number of tumor types (4, 5) suggesting an oncogenic role. Among the members of the Ras homology protein family, RhoC (molecular mass of 21 kDa) has been implicated in a wide range of cellular activities, including downstream expression of inflammatory genes and chemokines, cell proliferation, intracellular signaling, and cytoskeletal organization (6). More significantly, RhoC plays a central role in assembling focal adhesion by modulating the orientation of cytoskeletal fibers, resulting in cell polarity, increased cell motility, and consequently, increased invasiveness (7-9). In addition, signaling mediated by Rho proteins through Rho-activating kinase regulate proteins that in turn regulate actin polymerization such as cofilin, profilin, and formin homology proteins (10). Interestingly, high levels of RhoC and Rho-activating kinase are also associated with membrane blebbing, a phenomenon that is observed in motile or invasive cells (10, 11).

RhoC overexpression is now well documented in a wide range of malignant cancers, suggesting an important role in changing noninvasive carcinomas into invasive forms. Interestingly, overexpression of RhoC has been reported in inflammatory breast cancer and exclusively in invasive breast carcinoma (12-15). Other tumor types in which overexpression of RhoC has been reported are ovarian carcinoma (16), esophageal squamous cell carcinoma (17), pancreatic cancer (18), gastric cancer (17, 19), and human melanoma (11, 20). In addition, functional studies have shown that RhoC can act

Received 10/31/08; revised 8/4/09; accepted 9/2/09; published OnlineFirst 10/27/09.

**Grant support:** National Cancer Institute Head and Neck Cancer Specialized Programs of Research Excellence (T.N. Teknos and S.D. Merajver), NIH CA 77612 (S.D. Merajver), Breast Cancer Research Foundation (S.D. Merajver), and Burroughs Wellcome Fund (S.D. Merajver).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Requests for reprints:** Mozaffarul Islam, Department of Otolaryngology-Head and Neck Surgery, Cramblett Medical Clinic, Suit 4A, 456 West Tenth Avenue, Columbus, OH 43210. Phone: 614-292-3306; Fax: 614-688-4761. E-mail: Mozaffarul.Islam@osumc.edu.



doi:10.1158/1541-7786.MCR-08-0512

as a transforming oncogene when it is overexpressed in human mammary epithelia converting these normally immobile cells into highly motile and invasive malignant cells (12, 21). Thus, a wide range of current studies reveal the important role of RhoC in cancer metastasis.

However, very few studies to date have investigated the role of RhoC in head and neck cancer. Studies on gene expression profiling of stage III and IV regionally metastatic HNSCC showed that there are elevated levels of RhoC when compared with stage I and II localized malignancy (22). Furthermore, in our laboratory, we have shown that RhoC expression is elevated in the tumors of patients with HNSCC when compared with normal squamous cell epithelium (21). More importantly, our study showed that increased RhoC expression is strongly associated with lymph node metastasis and could also be used to predict metastasis even in small ($T_1$ and $T_2$) primary tumors (23). In the present study, we investigated the role of RhoC in head and neck metastasis by inhibiting its function using RNA interference. Our *in vitro* findings determined that inhibiting RhoC function strongly reduced cell motility and invasion. Furthermore, we observed a remarkable reduction in tumor metastasis and microvessel density in severe combined immunodeficiency (SCID) mice injected with RhoC knockdown cell lines. These findings suggest that inhibition of RhoC function in HNSCC can diminish a tumor's aggressive behavior, thus opening new possibilities for future drug therapies targeting this pathway.

## Results

### *RhoC mRNA Expression Is Greatly Reduced in Knockdown Clones from HNSCC Cell Lines*

To understand the role of RhoC expression in head and neck metastasis, we constructed cellular reagents in which RhoC expression was downregulated by small hairpin RNA (shRNA) in squamous cell carcinoma cell lines from the University of Michigan (UM-SCC-11A and UM-SCC-1). These cells exhibit a strong invasive phenotype and have shown in our previous studies that RhoC is constitutively active in these lines (23).

The inhibition of RhoC expression was achieved using RNA interference and lentiviral transfection and transduction technology. After lentiviral infection, positive (stable) clones were selected using puromycin antibiotics. Fluorescence microscopy of the stable clones showed a strong green fluorescence in the majority of the cells, signifying a high efficiency of transfection (Fig. 1).

We then tested the effectiveness of shRNA in depleting RhoC mRNA expression in the lentivirally infected cell lines using quantitative real-time PCR (qRT-PCR). Because only a small number of specific gene sequences are capable of activating the RNA degradation pathway, we used two different RhoC knockdown clones (i.e., C1 and C2 along with a parental and shRNA-scrambled sequence infected control) to ensure the effectiveness of depleting levels of RhoC. The results show greatly reduced expression levels of RhoC gene in the C1 and C2 RhoC knockdown clones, whereas normal RhoC expression was observed in clones with a shRNA-scrambled sequence (Fig. 2). The relative RhoC mRNA expression in parental, shRNA-scrambled control and RhoC knockdown clones 1 and 2 were evaluated by qRT-PCR and the $C_T$ values thus obtained were normalized using two housekeeping genes as described in Materials and Methods. As shown in Fig. 2A, RhoC mRNA expression decreased ~75% and 80% in RhoC knockdown clone 1 and clone 2 of UM-SCC-1, respectively. A similar decrease of 40% and 70% in RhoC mRNA levels was observed in RhoC knockdown clone 1 and clone 2 of UM-SCC-11A, respectively. However, the control shRNA-scrambled sequence in either of the cell lines did not show any significant reduction in RhoC mRNA expression level (Fig. 2A). To confirm that only RhoC expression was being inhibited, the mRNA levels of other Rho family members, Cdc42, Rac1, and Rac2 were also analyzed by qRT-PCR. As shown in Fig. 2B, C, and D, the expression levels of Cdc42, Rac1, and Rac2 are very similar to the parental lines, thus confirming that our shRNA process is highly specific to RhoC only. These studies provided a clear insight about the "switching off" of the RhoC machinery by decreasing total levels of RhoC mRNA expression, and therefore, further detailed studies on its functional roles are defensible. One of the most basic clinical questions that arise at this point is how inhibition of the RhoC transcript affects metastasis in head and neck cancer. To address this question, we investigated two characteristic behaviors of metastatic cells, invasion and motility in the transduced cell lines.

### *RhoC Knockdown Clones Show a Decrease in Cell Invasion and Motility*

In the invasion assays, RhoC-depleted clones of UM-SCC-11A and UM-SCC-1 were remarkably less invasive and motile compared with the parental or shRNA-scrambled controls (Fig. 3). Notably, cell invasion was decreased by 50% and 75% in RhoC knockdown clones 1 and 2, respectively, in the transduced UM-SCC-11A cell line (Fig. 3I). A similar decrease of 60% and 80% in clones 1 and 2, respectively, was observed in UM-SCC-1 lines (Fig. 3J) when compared with their parental or shRNA-scrambled controls ($n = 3$; $P < 0.003$).

We hypothesized that RhoC plays an important role in cell motility in HNSCC. We therefore investigated the effect of RhoC on cell motility using the scratch model. A noticeable decrease in cell motility was observed in RhoC knockdown clones as compared with the parental or shRNA-scrambled sequence control lines (Fig. 4A and B; $n = 3$, $P < 0.005$). These *in vitro* assays provide evidence for the first time that RhoC plays an important role in cell invasion and motility and suggest that RhoC is important for metastasis in head and neck cancer.

### *RhoC Plays an Important Role in Lung Metastasis and Microvessel Density Formation*

Besides localized tumor growth, lung metastasis is a common and frequent occurrence in patients with head and neck cancers (24). Keeping this aspect in view, we designed an *in vivo* study in which we could analyze the effect of RhoC inhibition on lung metastasis and primary tumor vascularity. This was achieved by injecting transduced cell lines through the tail veins of SCID mice and analyzing them for lung metastasis. Because both clones gave similar results in our

cell motility and invasion assays, we selected RhoC knockdown clone 2 for our subsequent *in vivo* studies and all results discussed hereafter are based on this clone. The xenograft mice were sacrificed 2 weeks after implantation and their lungs were analyzed for metastasis using H&E stain. As shown in Fig. 5A and B, in the mice injected with UM-SCC-11A parental or shRNA-scrambled control, a large metastatic focus and inflamed blood vessels were observed in the

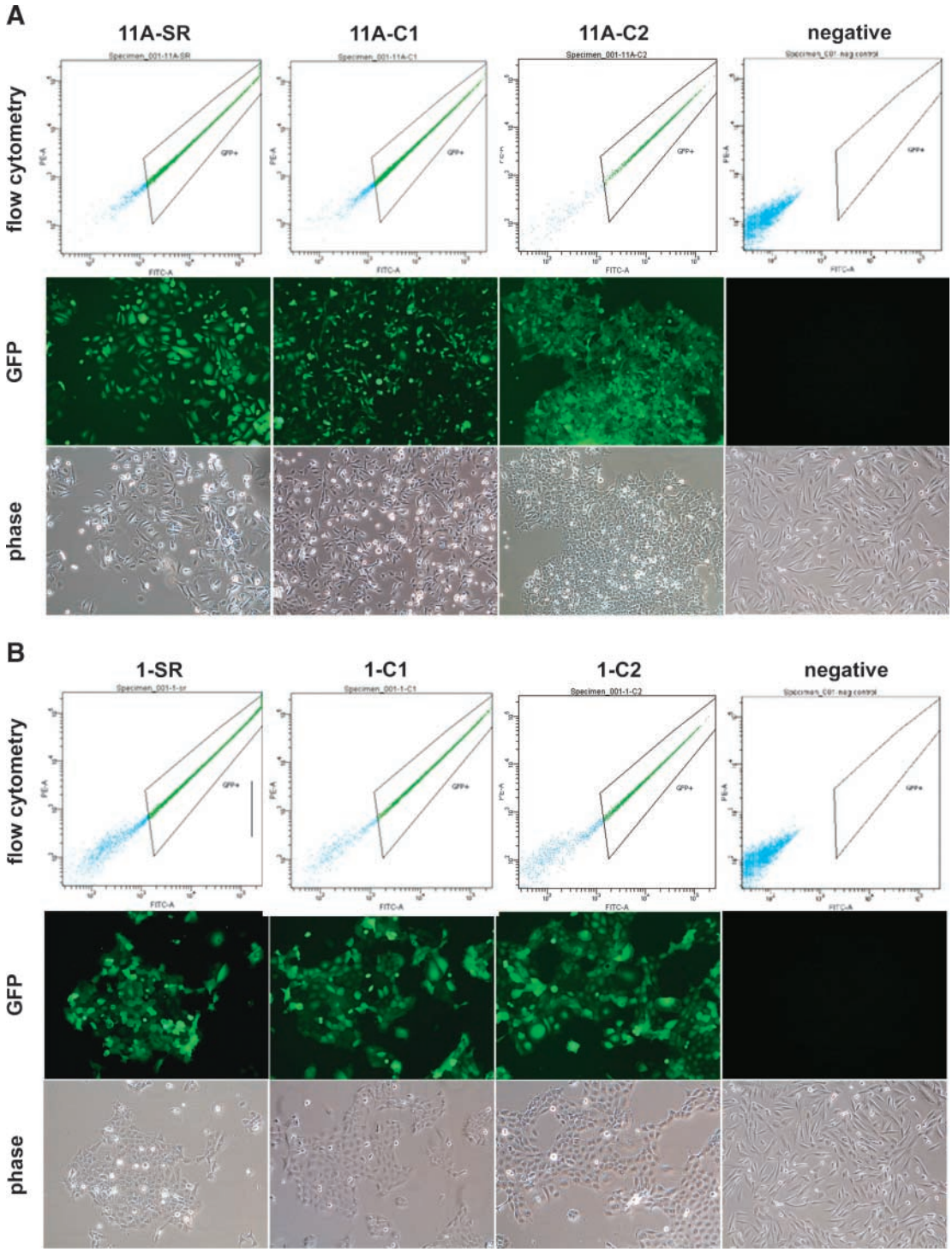



**FIGURE 1.** Lentivirus-infected cells showing GFP expression levels. **A.** UM-SCC-11A cell line transfected with shRNA-scrambled sequence control (*SR*), RhoC knockdown clone 1 (*C1*), RhoC knockdown clone 2 (*C2*), and uninfected cells as controls (*negative*). Histograms obtained by flow cytometry (*top*), and GFP-labeled cells in fluorescent (*middle*) and bright lights (*bottom*). **B.** UM-SCC-1. All other notations are the same as described above. As shown by the GFP expression levels, a high number of cells (80-90%) were successfully infected with recombinant lentivirus.

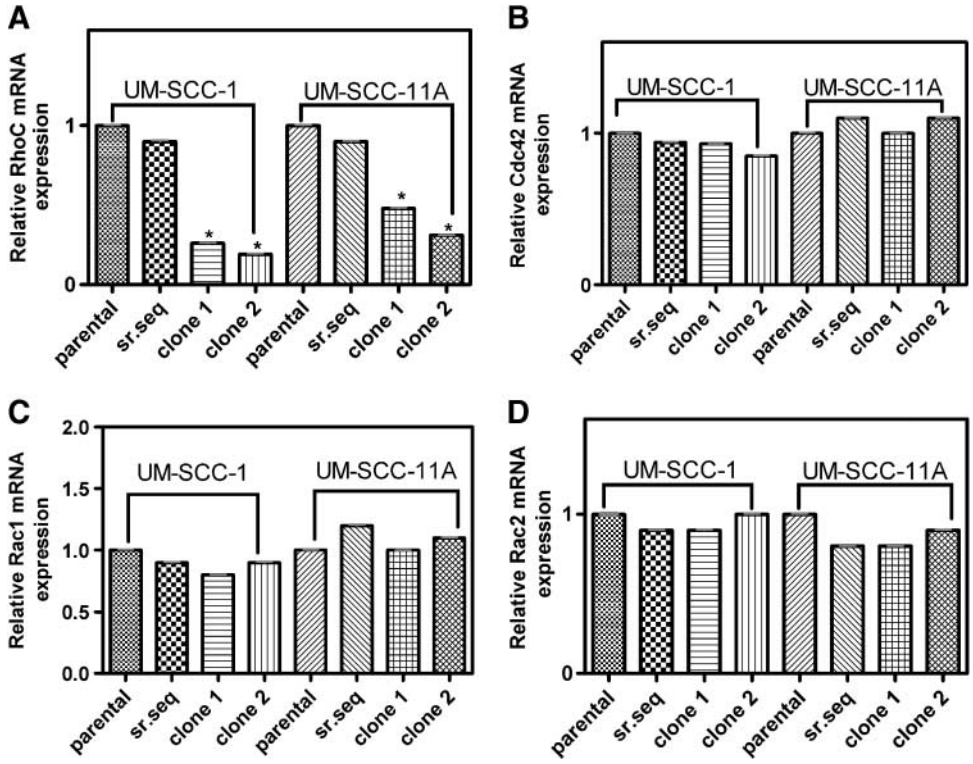



**FIGURE 2.** qRT-PCR of cell lines UM-SCC-1 and UM-SCC-11A showing the relative mRNA expression levels of RhoC (**A**), Cdc42 (**B**), Rac1 (**C**), and Rac2 (**D**) in parental (control), shRNA-scrambled sequence control, and RhoC knockdown clones 1 and 2 after selection and establishment of positive clones. Results were analyzed using $2^{-\Delta\Delta CT}$ methods. A significant decrease in mRNA levels of RhoC knockdown clones were obtained whereas the expression of Cdc42, Rac1, and Rac2 remained unchanged ($P < 0.05$).

lung region (marked by arrows). A similar set of results were obtained for the UM-SCC-1 parental and shRNA-scrambled equence control (Fig. 5D and F). In contrast, mice injected with RhoC knockdown clone have very small metastatic tissue with barely visible patches of inflamed blood vessels in UM-SCC-11A and UM-SCC-1, respectively (Fig. 5C and F).

In addition, the remaining dissected lung tissues were cultured for observation of cell growth by the metastatic tumors. The bar graph shows the number of cancer cells grown in digested lung of mice which includes parental, shRNA-scrambled control and RhoC knockdown clones. Interestingly, there is a 67% and 58% decrease in cell number in RhoC knockdown clones of UM-SCC-11A and UM-SCC-1, respectively, when compared with their parental lines (Fig. 5G and H). These results strongly suggest that inhibition of RhoC expression greatly reduces metastasis *in vivo*.

Furthermore, to test the angiogenic role of RhoC, parental and RhoC knockdown cells were implanted in the flank region of the SCID mice. Microvessel density of the localized solid primary tumor, which grows into a sizable volume after 12 weeks of implantation in the flank region, was analyzed using CD31 antibody. Microscopic analysis of the CD31-stained tumor revealed a remarkable difference in microvessel formation in mice implanted with RhoC knockdown clones when compared with the corresponding either parental or shRNA-scrambled control. In the control groups, well developed microvessels were observed in the tumors, which was in strong contrast with the poorly developed microvessels in mice implanted with RhoC knockdown clone (Fig. 6). Our results are in coherence with the published work about the essential role of RhoC in angiogenesis (25, 26). A similar pattern of microvessel development was observed in UM-SCC-11A parental, shRNA-scrambled, and RhoC knockdown clone (data not shown). These results suggest that RhoC is required for proper formation of the vascular network in a developing tumor.

## Discussion

Tumor metastasis is well correlated with the overexpression of certain oncogenes. The overexpression of the Rho gene family has been reported in many malignant forms of cancer (27), including pancreatic cancer (18), gastric cancer (17, 19), and human melanoma (11, 20). However, there have been very few studies on whether overexpression of RhoC is involved in head and neck tumor metastasis. Previous studies in our laboratories have shown that RhoC is actively expressed in several well established UM-SCC cell lines. Among the cell lines tested, the UM-SCC-11A and UM-SCC-1 lines exhibited considerably high levels of RhoC-protein (23). In particular, the active form of RhoC (RhoC GTPase) was observed to be constitutively expressed in the UM-SCC lines. Therefore, for our current study, we selected

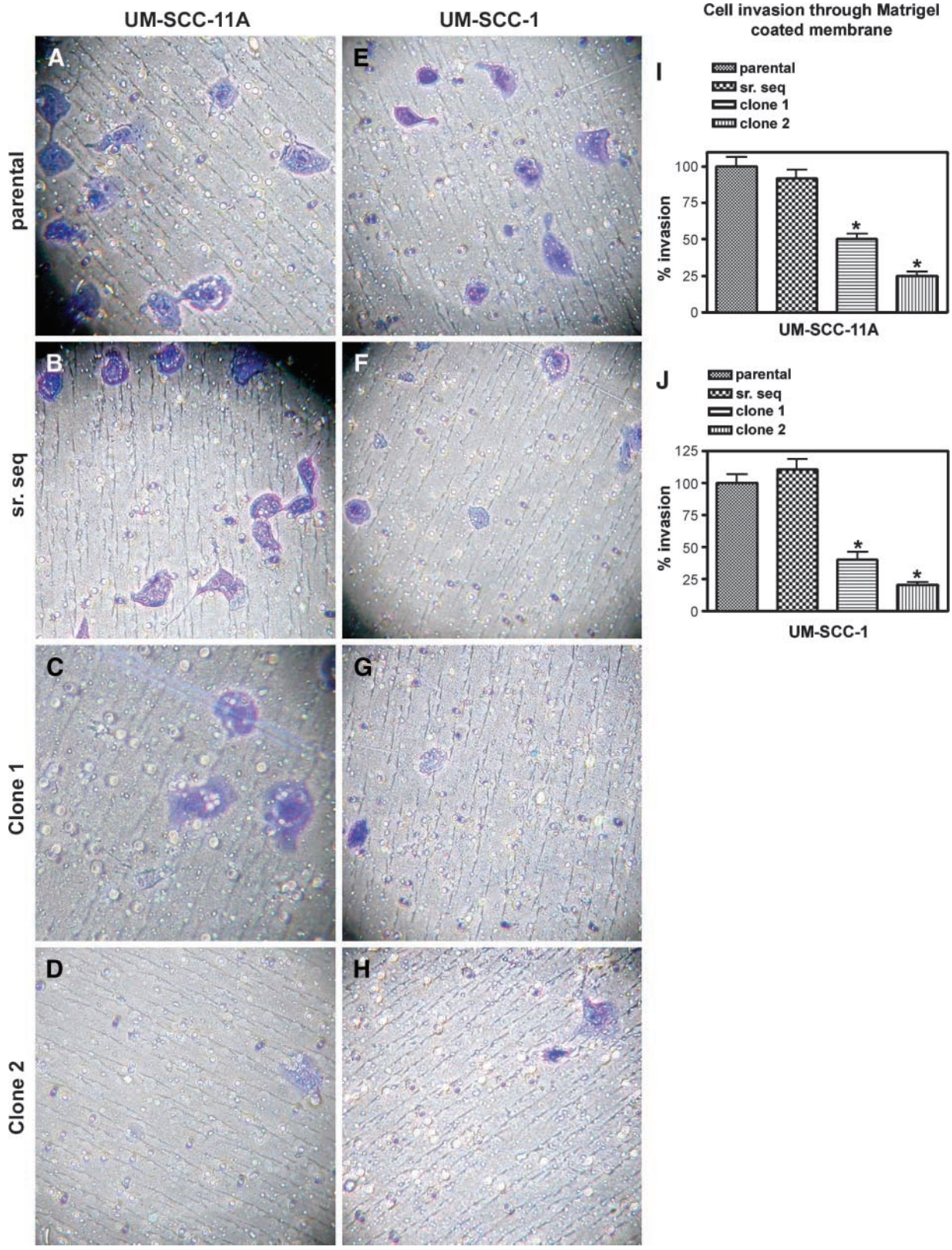



**FIGURE 3.** Cell invasion assay of UM-SCC-11A and UM-SCC-1 lines transfected with RhoC shRNA. **A** and **E.** Parental cell lines; **B** and **F.** shRNA-scrambled controls; **C** and **G.** RhoC knockdown clone 1; **D** and **H.** RhoC knockdown clone 2 of UM-SCC-11A and UM-SCC-1, respectively (magnification, ×40 and ×100). **I** and **J.** Columns, rates of invasion; bars, 95% CI ($P < 0.05$).

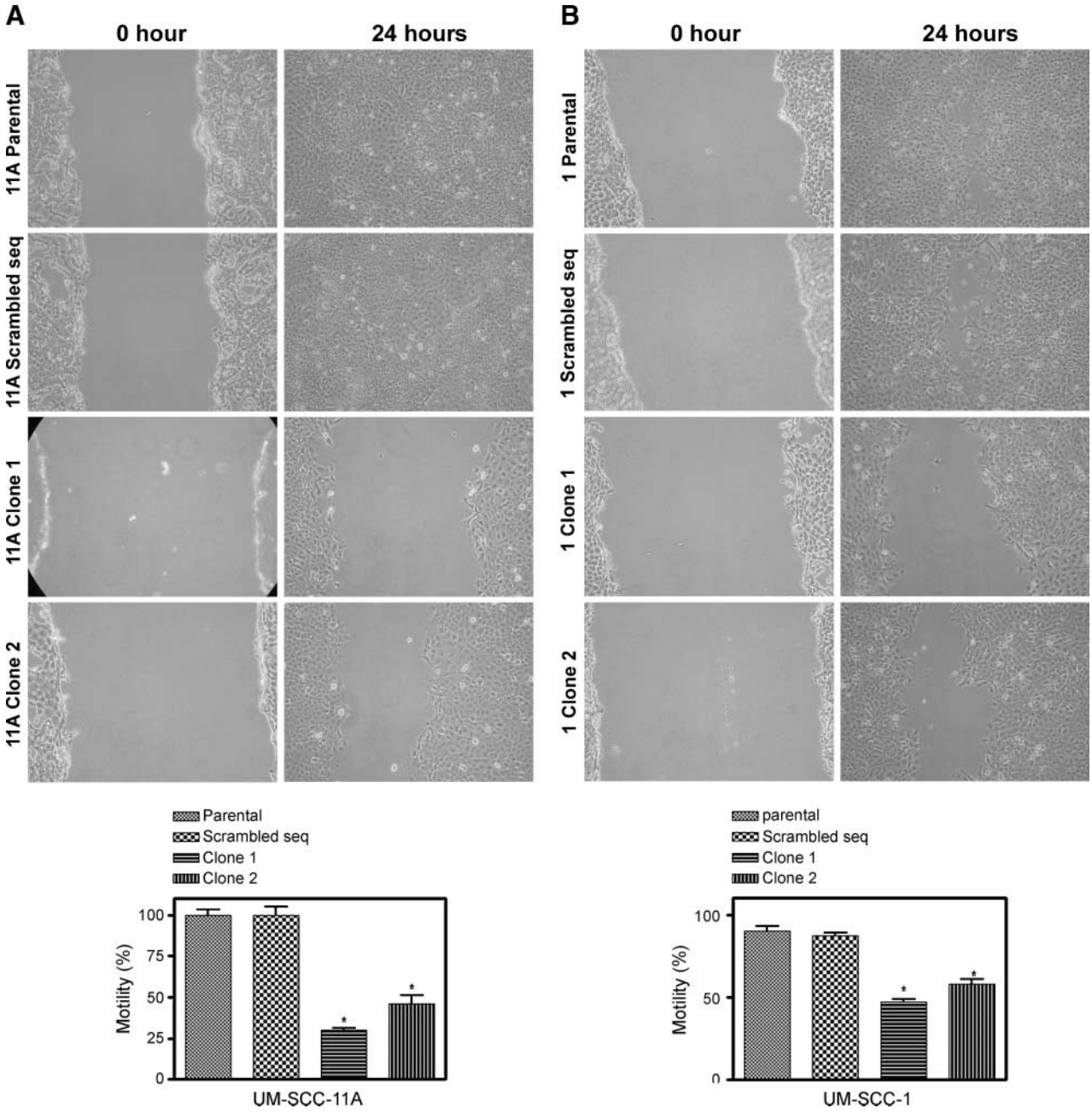



**FIGURE 4.** The effect of RhoC knockdown on cell motility. **A** and **B.** The slow movement of RhoC knockdown cells (after 24 h) as compared with its parental or shRNA-scrambled control in UM-SCC-11A and UM-SCC-1, respectively (magnification, ×40). Columns, percentage of motility with the initial reference point as 0 h ($P < 0.05$).

two UM-SCC lines (UM-SCC-11A and UM-SCC-1) to evaluate the role of RhoC in HNSCC metastasis. Our first and foremost aim was to inhibit RhoC expression in the two selected cell lines and analyze its function *in vitro*. Our expectation was that the motility and invasion would be greatly reduced in RhoC-depleted cell lines as compared with parental lines. In this study, we have shown a successful inhibition of RhoC gene expression and, subsequently, its function using shRNA techniques (Fig. 2). Furthermore, our data show that cell invasiveness and motility which are characteristics of aggressive head and neck cancer cell lines were diminished when RhoC expression was inhibited (Figs. 3 and 4). Therefore, these results suggest that RhoC overexpression drives cell invasion and motility in HNSCC. It is reported that one of the major functions of the Rho family of proteins is to control cytoskeletal organization (28). Cytoskeletal proteins are involved predominantly in cell motility. Therefore, RhoC may control metastasis by modulating cell motility (29). To

facilitate the movement of cells, they need to turn over both cell-extracellular matrices and cell to cell adhesions, which includes both adherence junctions and tight junctions (30, 31). It has also been reported that RhoC plays a predominant role over RhoA in the weakening of adherence junctions, which is an important step towards transforming cells into an invasive phenotype (6). These studies therefore, raise the question as to what effect RhoC inhibition would create *in vivo*. Our *in vivo* results showed that both inflamed blood vessels of lungs and a large volume of lung metastases were present in animals which were administrated by tail vein injection of either parental or shRNA-scrambled sequence (control) cell lines. In contrast, the lungs of mice implanted or injected with RhoC knockdown lines were free from any pathologic findings, specifically very minimal lung metastases and very low level of inflammation in lung tissues and blood vessels (Fig. 5). Furthermore, the level of angiogenesis in the localized tumors was assessed using CD31 antibody and these results showed a remarkable difference both in quality as well quantity of the microvessels in the tumors. The mice implanted with RhoC knockdown lines showed markedly fewer and less poorly developed microvessels as compared with the far greater in number and clearly defined vessels in parental or shRNA-control cell lines (Fig. 6).

The implications of the findings in this study provide a fertile area of research in HNSCC. For instance, recent work has shown that matrix metalloproteinases (MMP), which are well-known mediators of invasive tumor behavior, have been identified as a specific and critical player for the formation of lung metastasis (32, 33). Li et al. reported that the oncogene, *AF1Q*, which is responsible for primary breast tumor growth and pulmonary metastasis are at least, in part, regulates other MMPs and RhoC expression (34). The remodeling of the actin cytoskeleton is a critical and important step in the formation of pulmonary metastasis due to changes in cell shape, polarity, cell interactions, and eventual migration of the cancer cells. Interestingly, studies by Nelson et al. (35) have shown that expression of *MMP3* gene which induces epithelial-mesenchymal transition in mammary epithelial cells is brought about by change in cell shape through Rac1 (also a member of the Rho family) mediated changes in cytoskeletal structure. Clearly, future studies elucidating the specific interactions between MMP2, MMP3, and MMP9 (major MMP proteins in HNSCC) and RhoC are indicated, and may prove to be one of the signaling pathways for RhoC-mediated function.

In conclusion, the findings presented in this study illustrate that in both *in vivo* and *in vitro* conditions, RhoC plays an important role in head and neck cancer progression and metastasis. With additional investigations and ongoing development of RhoC specific inhibitors, this may prove to be an important therapeutic target in this patient population.

## Materials and Methods

### *Cell Culture and Generation of Stable RhoC Knockdown Clones*

UM-SCC-11A and UM-SCC-1 are well established cell lines derived, respectively, from a 65-y-old patient with a $T_2N_{2a}$ of the epiglottis and from a 46-y-old patient with $T_2N_0$ of the false vocal cord (36, 37). These cell lines were grown at 37°C in a humidified atmosphere with 95% air-5% $CO_2$. The cultures were maintained in DMEM (Life Technologies) containing 10% heat-inactivated fetal bovine serum (FBS; Hyclone) and supplemented with 50 μg/mL of penicillin G and 50 μg/mL of streptomycin sulfate.

RhoC knockdown and scrambled sequence constructs with green fluorescence protein (GFP) tag and puromycin resistance sites were synthesized by the vector core facility of

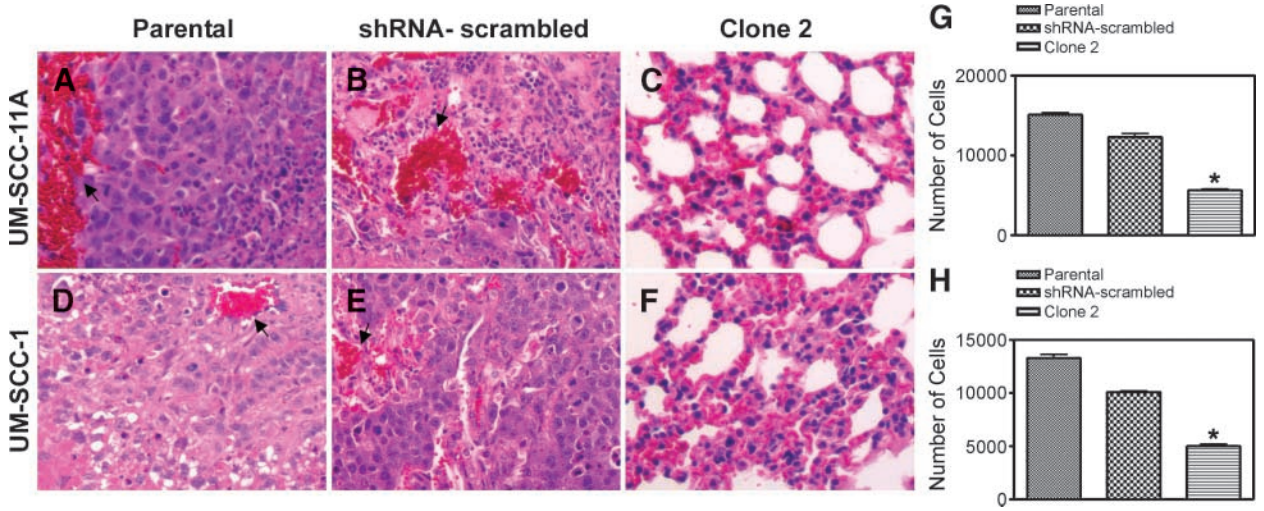



**FIGURE 5.** The effect of RhoC knockdown on lung metastasis in SCID mice injected through the tail vein with UM-SCC-11A and UM-SCC-1 cell lines transfected with RhoC shRNA. The lung sections were stained with H&E dye to show the degree of metastasis. **A** and **D.** Parental; **B** and **E.** shRNA-scrambled controls; **C** and **F.** RhoC knockdown clone 2. Black arrows, inflamed blood vessels present only in parental and scrambled-sequence controls (magnification, ×100). **G** and **H.** Number of cells obtained by culturing the lungs for UMS-CC-11A and UM-SCC-1, respectively, showing a marked reduction in RhoC knockdown clone 2 ($P < 0.05$).

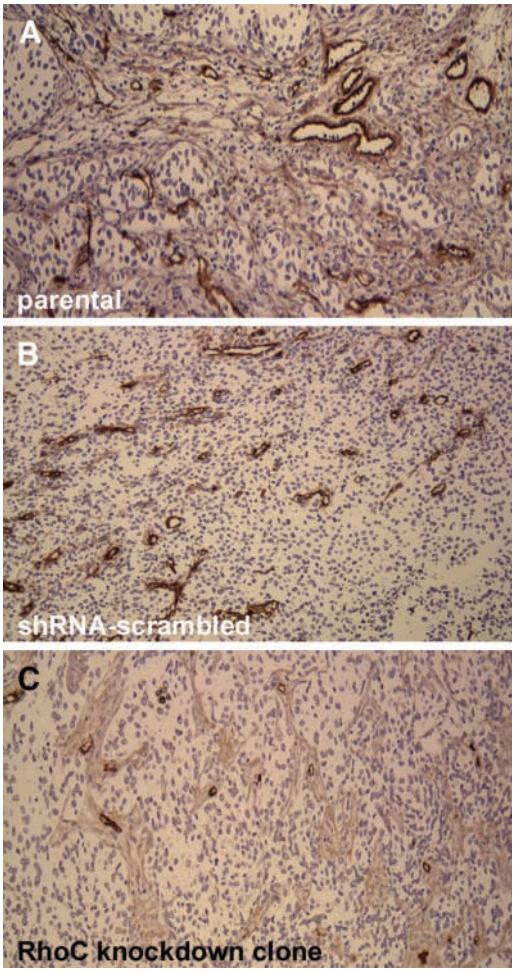



**FIGURE 6.** Assessment of microvessel density in UM-SCC-1 lines using immunostaining with CD31 antibody. **A** and **B.** Parental and shRNA-scrambled sequence with well-developed microvessels. **C.** RhoC knockdown clone 2 with much smaller and poorly developed microvessels (magnification, ×40).

the University of Michigan.[5] The sequences used for RhoC constructs are available in Open Biosystems[6] and include oligo ID V2LHS_69446 and V2LHS_69410, accession number NM_001042678. The sequences of the constructs are 69446 = 5′-ATACTGTCTTTGAGAACTATAT (sense; for RhoC knockdown clone 1) and 69410 = 5′-CACCAG-CACTTTATACACTTC (sense; for RhoC knockdown clone 2). The sequence of shRNA miR nonsilencing (scrambled) control is ATCTCGCTTGGGCGAGAGTAAGTGCTGTT-GACAGTAAGCGATCTCGCTTGGGCGAGAGTAAG-TAGTGAAGCCACAGATGTACTTACTCTCGCCCAGCGA-GAGTGCCTACTGCCTCGGA. This control sequence does not match any known mammalian genes (the sequence had at least three or more mismatches against any gene which was determined via nucleotide alignment/BLAST of target 22mer sequence). This is the nonsilencing shRNAmir hairpin sequence found in the pSM2, pSMP, pGIPZ, pTRIPZ, and pLemiR nonsilencing controls.

293FT cells (Invitrogen) were infected with 250 mmol/L of $CaCl_2$ solution containing RhoC shRNA construct, 25 μmol/L chloroquine and viral particles (i.e., Gag, Pol, and Env) and grown overnight. The medium was changed after 12 h to remove chloroquine and fresh DMEM-10% FBS was added to the growing 293FT cells to produce the virus. The supernatants from the infected cells were collected and 1 mL of this solution was added to growing UM-SCC-11A and UM-SCC-1 lines. Cells were incubated at 37°C and the GFP expression was monitored after 48 h of infection. Positive (stable) clones were selected using puromycin antibiotic (1.6 and 2.0 μg/mL for UM-SCC-11A and UM-SCC-1, respectively). These were then analyzed using fluorescence microscopy which showed a strong green fluorescence in the majority of the cells, signifying a high efficiency of infection (Fig. 1A and B). Furthermore, flow cytometry analyses showed that the number of non-infected cells were significantly low (Fig. 1A and B).

### *Flow Cytometry Analyses*

Approximately 70% to 80% confluent lentivirus-infected cells were harvested using trypsin-EDTA solution and resuspended in phosphate buffer saline containing 3% FBS, 0.5 mm EDTA, and 60 units/mL of DNase. Flow cytometry analysis was done using a BD FACS Aria IIU flow cytometer equipped with a 488 nm, 15 mW, air-cooled argon laser (Analytical Cytometry Laboratory, Ohio State University Comprehensive Cancer Centre). GFP-positive cells were sorted out and grown for subsequent experiments.

### *qRT-PCRs*

Total RNA was isolated according to the standard procedure using TRIzol reagent (Invitrogen). qRT-PCR were conducted with a TaqMan probe system from Applied Biosystems by using the following products: cdc42, Hs03044122_g1; Rac1, Hs01025984_m1; Rac2, Hs01032884_m1; and RhoC, Hs00733980_m1. β-Actin and G3PDH were used as the data normalizers. Relative changes in gene expression were calculated using the $2^{-\Delta\Delta CT}$ method (38).

### *Cell Invasion and Motility Assay*

*Invasion Assay.* Cell invasion assays were done using BD BioCoat Matrigel Invasion Chamber which was obtained from BD Biosciences. The procedure was followed according to the instructions of the manufacturer. Briefly, ~$2.5 \times 10^5$ cells in 2 mL of serum-free DMEM were added at the top of the insert and 1 mL of the medium was added in the bottom well of each insert. FBS albumin was added to the medium in the lower chamber (final concentration of FBS was 10%, v/v), which acted as a chemoattractant. Cells were incubated for 22 h in a humidified cell culture incubator at 37°C, 5% $CO_2$ atmosphere. Next, the noninvading cells at the top of the insert were scraped out with the help of cotton-tipped swab. The invading cells which were attached to the underside of the membrane were fixed in 100% methanol and stained with 1% Toluidine prepared in 100% methanol. After repeated washing of the membrane using distilled water, stained cells were allowed to air-dry at room temperature before it was visualized under a

[5] http://www.med.umich.edu/vcore
[6] http://www.openbiosystems.com/

microscope. A parallel experiment with control inserts (without Matrigel) was also run. Matrigel-invaded cells were counted microscopically at 40× and 100× magnifications.

*Motility Assay.* Cell motility assays were done in 100 mm Petri dishes. At ~80% confluence, cells were washed with PBS and a fine scratch in the form of a groove was made with the help of a sterile pipette tip and immediately photographed. We designated this time as the 0 h. Next, cells were supplemented with DMEM containing 10% FBS and allowed to grow. A migration of cells from the edge of the groove towards the center was monitored microscopically at 40× magnifications after 24 h to assess the extent of scratched area covered. The width of the scratch was measured at 0 h and after 24 h to calculate the percentage of the gape covered by the cells in a 24-h time period.

### *Animal Xenograft*

Athymic SCID mice were obtained from the Jackson Laboratory; 6-wk-old mice were housed in cages of five animals each. Five animals per treatment were selected to receive parental, shRNA-scrambled sequence control and RhoC knockdown clone, resulting in 15 animals per cell line for each set of experiments. Approximately $5 \times 10^6$ UM-SCC-11A and UM-SCC-1 cells were suspended in 100 μL of serum-free DMEM and injected thorough the tail vein and/or in the flank region of mice using a 0.5-inch, 27-gauge needle. Animals were monitored every other day for their general health and activities. At the end of the second week, the animals were euthanized using a $CO_2$ chamber. The lungs were dissected and half of the lungs were fixed in buffered formalin for 6 h, and thereafter transferred to 70% methanol and then processed to form paraffin-embedded tissue blocks (H&E staining was also done). The remaining half of the lungs was digested in collagenase for culturing the cells. At the end of week 12, tumors in the flank region were fully grown. The animals were euthanized and tumors were dissected and fixed in the same way as described above for CD31 staining.

### *Lung Metastases*

Slides of 5-μm-thick sections of lungs were prepared and stained with H&E. Five random fields were microscopically examined in a blind fashion at 100× magnification to detect metastases.

### *Microvessel Density*

Microvessel density in all primary tumors was assessed using antimouse CD31 antibody (PharMingen) at a dilution of 1:250. Five random low-power fields (40× magnification) were selected to visualize the microvessels. The mean was reported in a blind fashion for each tumor.

### *Statistical Analysis*

Statistical analyses were done using Sigma GraphPad prism 4 software. The mean ± SD was reported. Differences were considered to be statistically significant at $P < 0.05$.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Acknowledgments

The authors thank Dr. N.S. Mahfooz for critical reading of the manuscript and Dr. Thomas Carey for providing the cell lines used in this work.

## References

1. Parkin DM. Global cancer statistics in the year 2000. Lancet Oncol 2001;2:533–43.

2. Society AC. Cancer facts and figures 2009.

3. Tran N, Rose BR. Role of human papilomavirus in the etiology of head and neck cancer. Head Neck 2007;29:64–70.

4. Kleer CG, van Golen KL, Zhang Y, Wu ZF, Rubin MA, Merajver SD. Characterization of RhoC expression in benign and malignant breast disease: a potential new marker for small breast carcinomas with metastatic ability. Am J Pathol 2002;160:579–84.

5. Gomez del Pulgar T, Benitah SA, Valeron PF, Espina C, Lacal JC. Rho GTPase expression in tumourigenesis: evidence for a significant link. Bioessays 2005;27:602–13.

6. Sahai E, Marshall CJ. ROCK and Dia have opposing effects on adherens junctions downstream of Rho. Nat Cell Biol 2002;4:408–15.

7. Nobes CD, Hall A. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. Cell 1995;81:53–62.

8. Leung T, Chen XQ, Manser E, Lim L. The p160 RhoA-binding kinase ROK α is a member of a kinase family and is involved in the reorganization of the cytoskeleton. Mol Cell Biol 1996;16:5313–27.

9. Lepage M, Dow WC, Melchior M, et al. Noninvasive detection of matrix metalloproteinase activity *in vivo* using a novel magnetic resonance imaging contrast agent with a solubility switch. Mol Imaging 2007;6:393–403.

10. Olson MF, Sahai E. The actin cytoskeleton in cancer cell motility. Clin Exp Metastasis 2009;26:273–87.

11. Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. Nature 2000;406:532–5.

12. van Golen KL, Wu ZF, Qiao XT, Bao LW, Merajver SD. RhoC GTPase, a novel transforming oncogene for human mammary epithelial cells that partially recapitulates the inflammatory breast cancer phenotype. Cancer Res 2000;60:5832–8.

13. van Golen KL, Davies S, Wu ZF, et al. A novel putative low-affinity insulin-like growth factor-binding protein, LIBC (lost in inflammatory breast cancer), and RhoC GTPase correlate with the inflammatory breast cancer phenotype. Clin Cancer Res 1999;5:2511–9.

14. Kleer CG, Zhang Y, Pan Q, et al. WISP3 and RhoC guanosine triphosphatase cooperate in the development of inflammatory breast cancer. Breast Cancer Res 2004;6:R110–5.

15. Lo AC, Kleer CG, Banerjee M, et al. Molecular epidemiologic features of inflammatory breast cancer: a comparison between Egyptian and US patients. Breast Cancer Res Treat 2008;112:141–7.

16. Horiuchi A, Imai T, Wang C, et al. Up-regulation of small GTPases, RhoA and RhoC, is associated with tumor progression in ovarian carcinoma. Lab Invest 2003;83:861–70.

17. Faried A, Faried LS, Kimura H, et al. RhoA and RhoC proteins promote both cell proliferation and cell invasion of human oesophageal squamous cell carcinoma cell lines *in vitro* and *in vivo*. Eur J Cancer 2006;42:1455–65.

18. Kusama T, Mukai M, Endo H, et al. Inactivation of Rho GTPases by p190 RhoGAP reduces human pancreatic cancer cell invasion and metastasis. Cancer Sci 2006;97:848–53.

19. Pan Q, Bao LW, Teknos TN, Merajver SD. Targeted disruption of protein kinase C epsilon reduces cell invasion and motility through inactivation of RhoA and RhoC GTPases in head and neck squamous cell carcinoma. Cancer Res 2006;66:9379–84.

20. Ruth MC, Xu Y, Maxwell IH, Ahn NG, Norris DA, Shellman YG. RhoC promotes human melanoma invasion in a PI3K/Akt-dependent pathway. J Invest Dermatol 2006;126:862–8.

21. van Golen KL, Wu ZF, Qiao XT, Bao L, Merajver SD. RhoC GTPase overexpression modulates induction of angiogenic factors in breast cells. Neoplasia 2000;2:418–25.

22. Schmalbach CE, Chepeha DB, Giordano TJ, et al. Molecular profiling and the identification of genes associated with metastatic oral cavity/pharynx squamous cell carcinoma. Arch Otolaryngol Head Neck Surg 2004;130:295–302.

23. Kleer CG, Teknos TN, Islam M, et al. RhoC GTPase expression as a potential marker of lymph node metastasis in squamous cell carcinomas of the head and neck. Clin Cancer Res 2006;12:4485–90.

24. Hsu YB, Chu PY, Liu JC, et al. Role of chest computed tomography in head and neck cancer. Arch Otolaryngol Head Neck Surg 2008;134:1050–4.

25. Wang W, Wu F, Fang F, Tao Y, Yang L. RhoC is essential for angiogenesis induced by hepatocellular carcinoma cells via regulation of endothelial cell organization. Cancer Sci 2008;99:2012–8.

26. Merajver SD, Usmani SZ. Multifaceted role of Rho proteins in angiogenesis. J Mammary Gland Biol Neoplasia 2005;10:291–8.

27. Bos JL. The ras gene family and human carcinogenesis. Mutat Res 1988;195: 255–71.

28. Van Aelst L, D'Souza-Schorey C. Rho GTPases and signaling networks. Genes Dev 1997;11:2295–322.

29. Lauffenburger DA, Horwitz AF. Cell migration: a physically integrated molecular process. Cell 1996;84:359–69.

30. Burridge K, Wennerberg K. Rho and Rac take center stage. Cell 2004;116: 167–79.

31. Aspenstrom P, Fransson A, Saras J. Rho GTPases have diverse effects on the organization of the actin filament system. Biochem J 2004;377:327–37.

32. Hiratsuka S, Nakamura K, Iwai S, et al. MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis. Cancer Cell 2002;2:289–300.

33. Ikoma T, Takahashi T, Nagano S, et al. A definitive role of RhoC in metastasis of orthotopic lung cancer in mice. Clin Cancer Res 2004;10:1192–200.

34. Li DQ, Hou YF, Wu J, et al. Gene expression profile analysis of an isogenic tumour metastasis model reveals a functional role for oncogene AF1Q in breast cancer metastasis. Eur J Cancer 2006;42:3274–86.

35. Nelson CM, Khauv D, Bissell MJ, Radisky DC. Change in cell shape is required for matrix metalloproteinase-induced epithelial-mesenchymal transition of mammary epithelial cells. J Cell Biochem 2008;105:25–33.

36. Carey TE, Kimmel KA, Schwartz DR, Richter DE, Baker SR, Krause CJ. Antibodies to human squamous cell carcinoma. Otolaryngol Head Neck Surg 1983;91:482–91.

37. Krause CJ, Carey TE, Ott RW, Hurbis C, McClatchey KD, Regezi JA. Human squamous cell carcinoma. Establishment and characterization of new permanent cell lines. Arch Otolaryngol 1981;107:703–10.

38. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C(T))$ method. Methods 2001;25: 402–8.

Biochimica et Biophysica Acta 1796 (2009) 201–215

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbacan

Review

# Translocations in epithelial cancers

J. Chad Brenner [a,c], Arul M. Chinnaiyan [a,b,c,d,e,*]

[a] Michigan Center for Translational Pathology, University of Michigan, 1400 E. Medical Center Drive, 5316 CCGC, Ann Arbor, MI 48109, USA
[b] Howard Hughes Medical Institute, University of Michigan, 1400 E. Medical Center Drive, 5316 CCGC, Ann Arbor, MI 48109, USA
[c] Department of Pathology, University of Michigan, 1400 E. Medical Center Drive, 5316 CCGC, Ann Arbor, MI 48109, USA
[d] Department of Urology, University of Michigan, 1400 E. Medical Center Drive, 5316 CCGC, Ann Arbor, MI 48109, USA
[e] Comprehensive Cancer Center, University of Michigan, 1400 E. Medical Center Drive, 5316 CCGC, Ann Arbor, MI 48109, USA

ARTICLE INFO

*Article history:*
Received 25 February 2009
Accepted 21 April 2009
Available online 3 May 2009

*Keywords:*
Translocation
Epithelial
Rearrangement
Gene fusion
Chimera
MLL
ERG
ALK
HMGA2
COPA

ABSTRACT

Genomic translocations leading to the expression of chimeric transcripts characterize several hematologic, mesenchymal and epithelial malignancies. While several gene fusions have been linked to essential molecular events in hematologic malignancies, the identification and characterization of recurrent chimeric transcripts in epithelial cancers has been limited. However, the recent discovery of the recurrent gene fusions in prostate cancer has sparked a revitalization of the quest to identify novel rearrangements in epithelial malignancies. Here, the molecular mechanisms of gene fusions that drive several epithelial cancers and the recent technological advances that increase the speed and reliability of recurrent gene fusion discovery are explored.



**Contents**



* Corresponding author. Department of Pathology, University of Michigan, 1400 E. Medical Center Drive, 5316 CCGC, Ann Arbor, MI 48109, USA. Tel.: +1 734 615 4062.
*E-mail address:* arul@umich.edu (A.M. Chinnaiyan).
*URL:* http://www.pathology.med.umich.edu/dynamo/chinnaiyan/index.jsp (A.M. Chinnaiyan).

0304-419X/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbcan.2009.04.005

## 1. Introduction

Throughout history, technological advances are often followed by discoveries that dramatically alter our perceptions of disease etiology. For example, after the term "chromosome" was introduced in the mid-1849's, several German pathologists began using techniques to compare gross mitotic changes in tissue sections from different human malignancies [1]. Almost half of a century later, Theodore Boveri published a critical hypothesis that, "mammalian tumors might be initiated by mitotic abnormalities that resulted in a change in the number of chromosomes in the cell (aneuploidy)", based on the observation that sea urchin embryos would frequently engage in uncommon development following mitotic abnormality [2]. As time passed, breakthroughs arose that dramatically increased the quality and reproducibility of cytogenetic techniques such as the use of colchicine, which arrests cells in mitosis by inhibiting microtubule assembly. As a result of these observations, the general hypotheses regarding the evolution of human disease became increasingly complex; particular pathological conditions were associated with specific chromosomal abnormalities, such as Lejeune's association of Down syndrome with an extra copy of chromosome 21 [3,4].

Advances in technology once again spurred discovery when, in 1958, Rothfels and Siminovitch published a new cytogenetic, air-drying technique for flattening chromosomes [5]. The application of this technology later allowed Hungerford and Nowell to further characterize their initial observation that two patients with chronic myelogenous leukemia (CML) had a characteristic small chromosome [6]. Soon after the initial publication, Hungerford and Nowell were able to report on a series of seven patients, all of which harbored this minute chromosome [7]. This was coined the "Philadelphia chromosome" after the city in which the abnormal chromosome was discovered in accord with the Committee for the Standardization of Chromosomes [8]. The rearrangement leading to the Philadelphia chromosome was eventually characterized as a translocation between chromosomes 9 and 22 [9], resulting in the fusion of the breakpoint cluster region (*BCR*) gene on chromosome 22 with the v-abl Abelson murine leukemia viral oncogene homolog (*ABL1*) gene on chromosome 9 [10]. Later in 1990, Lugo et al. demonstrated that the BCR–ABL1 fusion protein is an active tyrosine kinase, through immunoblotting cell lysates from Rat 1 transfected cells, revealing that cells transfected with either BCR–ABL1 or v-src, but not v-H-ras or v-myc, had a significant increase in total phosphotyrosine content [11]. Understanding the molecular mechanism of BCR–ABL1 led to the development of one of the first molecularly tailored therapies as the small molecule Imatinib was specifically selected for its ability to inhibit BCR–ABL1 kinase activity [12,13]. The success of treating chronic myelogenous leukemia with a specific inhibitor of the BCR–ABL1 chimera led to a strong interest in the discovery of novel gene fusions in other cancer subtypes with the long term goal of designing disease specific therapeutics.

As techniques like the use of chromosome banding for karyotypic analysis were improved, the impact on discovery of novel gene fusions was immediately evident in leukemias and lymphomas. In fact, while BCR–ABL1 is perhaps the most famous gene fusion, the first molecularly characterized chimera was discovered by Zech et al. through the use of karyotypic analysis and is actually involved in the pathogenesis of Burkett's lymphoma and was identified. While this karyotypic analysis demonstrated absence of the distal region on the long arm of chromosome 8 and an extra band in the long arm chromosome 14 distal segment [14], the genes involved in the rearrangement remained elusive until 1982 when it was demonstrated that the translocation altered the *c-MYC* oncogene [15] and that the promoter and 5′ region of the immunoglobulin heavy chain (*IGH*) gene were rearranged such that the *IGH* promoter controls *c-MYC* expression [16]. Although this fusion does not lead to a chimeric protein, it was demonstrated that aberrant *c-MYC* expression through the *IGH* promoter is a necessary component of malignant transformation in Burkett's lymphoma [17].

As with lymphoma research, karyotypic analysis rapidly led to the identification of recurrent breakpoints that seemed to characterize subsets of myeloid leukemia. For example, in 1973, the acute myeloid leukemia 1 (*AML1*) gene was cloned from the breakpoint region of the first recurrent translocation described in leukemia, t(8;21) [18]. In 1991, the *AML1* gene was found to be fused to the eight-twenty one (*ETO*) gene on chromosome 21, which is also known as runt-related transcription factor 1 translocated to 1 (*RUNX1T1*) [19,20].

As the techniques of molecular biology improved, it became easier and easier to obtain the DNA sequence adjacent to chromosomal breakpoints. Since the original identification of *AML1* in myeloid leukemia, over 10 genes have been described to participate in rearrangements with *AML1* [21]. In fact, advances in sequencing technology led to the realization that several genes are recurrently and promiscuously fused to multiple partners; the examples of which are ever increasing. In addition to *AML1*, the other notable example of a promiscuous fusion gene partner is the mixed-lineage leukemia (*MLL*) gene, which is involved in over 40 different rearrangements (reviewed in [22]). In fact, because of the variety and difficulty of discussing all chromosomal aberrations in human malignancies, Mitelman et al. maintain and frequently update an online database of rearrangements and chromosome aberrations from all malignant neoplasms [23].

With the rapid development of current technologies like high-throughput sequencing, our perceptions as to the origins of disease have revealed a critical involvement of chromosomal aberrations, in particular, the role of translocations and gene fusions in malignant development. With a better understanding of the role of these chromosomal aberrations, therapies designed to inhibit the molecular function of chimeric proteins have recently been developed and, like Imatinib, some have demonstrated a window of strong efficacy. Consequently, much hope has been generated by the potential for targeting existing and novel gene fusions that characterize specific cancer subtypes with rationally designed molecularly tailored therapies. Here, we review known genomic rearrangements in epithelial tumors that led to aberrant expression of chimeric transcripts and the emerging technologies that may lead to the identification of novel gene fusions.

## 2. Gene fusions in epithelial cancers

In order to highlight the number of genomic rearrangements leading to fusion genes that characterize epithelial cancers, we have surveyed some of the well-studied chimeras from several solid malignancies and describe the fusions in approximate chronological order (Fig. 1). In the ensuing sections, we will analyze concepts from a global view of epithelial gene fusions with a few case studies of rearrangements from leukemia and endometrial stromal tumors. Gene fusions will be categorized into three different types: (1) those which alter the transcriptional regulation, (2) those which alter mRNA regulation and (3) those which alter protein activity. This will be followed by a discussion of the potential reasons why gene fusions have not been in the limelight of solid tumor pathogenesis and the

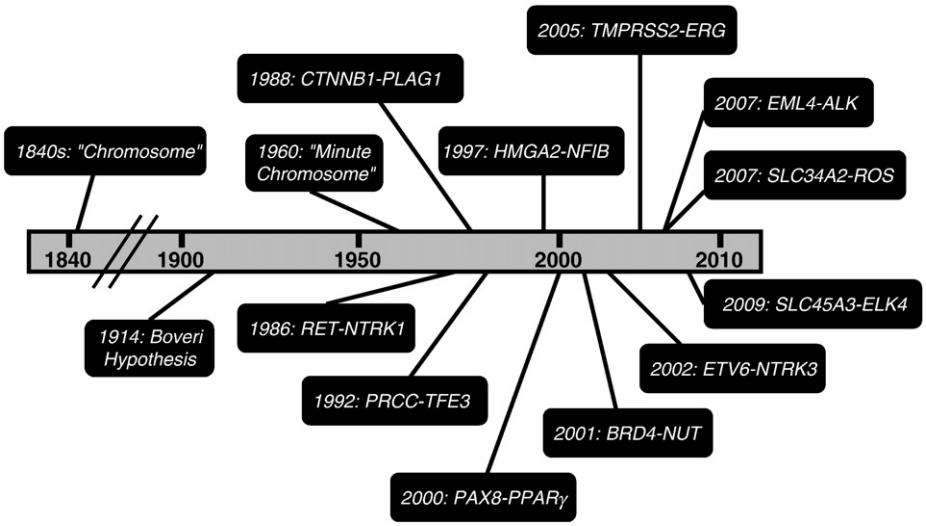



**Fig. 1.** Chronology of gene fusion discoveries in epithelial cancers.

developing technologies that are being used to find novel recurrent gene fusions in common epithelial tumors.

### 2.1. RET–NTRK1

The initial discovery of an epithelial gene fusion in mid-1989 comes directly from a novel screening technique used to identify transforming oncogenes. In this experimental approach, immortalized NIH3T3 cells were transfected with fragments of tumor cell genomic DNA, plated in soft agar. DNA is then isolated from cells and sequenced or sub-cloned to identify critical fragments. Using this approach, Martin-Zanca et al. identified the *RET–NTRK1* genomic translocation, providing some of the first insights into the possibility that recurrent genomic rearrangements were not specifically of hematologic phenomena [24].

*RET* (rearranged during transfection) encodes a tyrosine kinase [25,26] that was originally identified through transfection of DNA from a human T-cell lymphoma into NIH3T3 cells [27]. NTRK1 is a membrane-bound tyrosine kinase receptor that regulates neuronal cell growth, differentiation, and programmed cell death pathways [28]. Fusion of these two genes results in loss of the NTRK1 signal sequence giving rise to cytoplasmic localization and constitutive activation of the fusion protein [29]. Interestingly, although *NTRK1* was the first identified *RET* fusion partner, *RET* has several other N-terminal fusion partners including *H4* [30,31], *R1α* [32], *RFG5* [33] and *ELE1* [34,35]. One possible explanation for the diversity of genomic rearrangements observed in PTC is that the underling pathology is simply dependent on deregulation of either the *RET* or *NTRK1* tyrosine kinase domain (reviewed in [36]). Consequently, the important determining event in PTC carcinogenesis may be constitutive activation of the mitogen-activated protein kinase (MAPK) signaling pathway, which can be caused by rearrangement of either the *RET* and/or *NTRK1* gene. One reason for this hypothesis is that while the *RET–NTRK1* rearrangement appears to be the predominant gene fusion responsible for childhood PTC, in adult-onset populations activating point mutations in the *BRAF* gene or, controversially, the *RAS* gene [37–43], also lead to constitutive activation of the MAPK pathway without *RET* and/or *NTRK1* genomic rearrangement [44].

In addition to differences in the age-related molecular onset of PTC, the proportion of cases with either a *RET* or *NTRK1* rearrangement also appears to be based on the geographic area of origin [45–47], possibly because thyroid cancer is established to be associated with exposure to ionizing radiation [37,48]. Indeed, studies of patient populations exposed to either the Chernobyl nuclear power plant accident [49,50] or the atomic bombings [51] have demonstrated that genomic rearrangements occur at a higher frequency than mutations following extreme exposure to radiation [37,48], suggesting that under certain biological conditions exposure to high dose radiation may actually trigger specific DNA breaks leading to intentional genomic rearrangement. In fact, the fusion proteins that characterize PTC contain a number of different N-terminal partners fused the C-terminal tyrosine kinase domain of either *RET* or *NTRK1* [52] that may depend on the environmental cues leading to genomic rearrangement.

### 2.2. CTNNB1–PLAG1

Within a year of publication of the *RET–NTRK1* genomic rearrangement in PTC, another epithelial translocation was reported in pleomorphic adenoma (PA) [53], a slow-growing epithelial tumor that is responsible for more than 50% of salivary gland tumors [54], but less than 10% of tumors from the head and neck [55]. In contrast to *RET–NTRK1* which was discovered by a screening technique, rearrangements in PA were first identified by karyotypic analysis of primary tumors. In fact, before any of the breakpoint genes were identified, PAs were already divided into four cytogenetic groups (reviewed in [56]). Rearrangements of 8q12 account for about 40% of PAs with t(3;8)(p21;q12) comprising about half of rearrangements at this locus. Translocations of 12q14–15 account for about 8% of PAs with t(9;12)(p12–22;q13–15) or an ins(9;12)(p12–22;q13–15) responsible for these abnormalities [57,58]. Tumors with non-recurrent clonal changes comprise about 20% of PAs, and tumors with apparently normal karyotypes account for the remaining cases [56].

Almost 20 years after the initial karyotyping studies, Kas et al. used a comprehensive breakpoint mapping approach, southern blot analysis and 5′ rapid amplification of cDNA ends (5′ RACE) to identify the genes involved in the most prevalent PA translocation, t(3;8)(p21;q12) as *β-Catenin* (*CTNNB1*) and *PLAG1* (pleomorphic adenoma gene 1) [59]. Specifically, the t(3;8)(p21;q12) rearrangement fuses the *β-Catenin* (*CTNNB1*) promoter and exon 1 to *PLAG1* exon 2, resulting in a marked increase in PLAG1 expression (Fig. 2). As such, because the gene fusion results in altered DNA level regulation of PLAG1 transcript, this gene fusion is characterized as type 1. Interestingly, the reciprocal translocation links the *PLAG1* promoter and exon 1 to *β-Catenin* exon 2, reducing β-Catenin expression. As β-Catenin signals through several well-characterized oncogenic pathways (reviewed in [60]), the reduction in β-Catenin is curious. *PLAG1*, however, belongs to the PLAG family of proteins and encodes a zinc finger protein with two putative nuclear localization signals and can bind to either DNA or

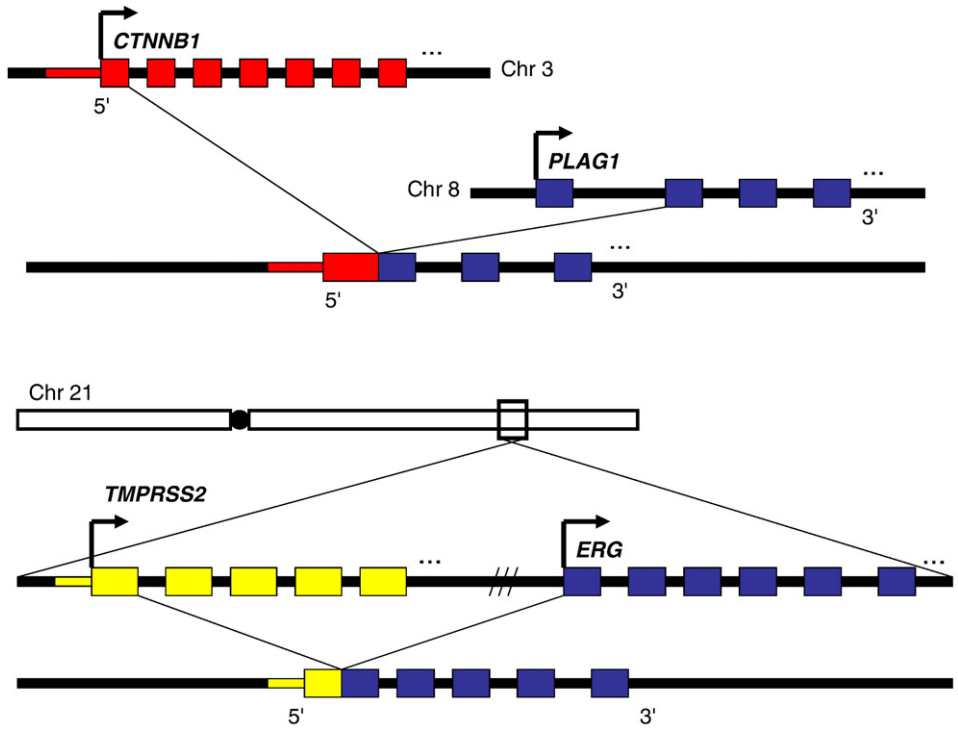



**Fig. 2.** Genomic structure of gene fusions with altered transcriptional regulation. The *CTNNB1–PLAG1* and *TMPRSS2–ERG* chimeras represent an important class of gene fusions in which the proto-oncogene remains largely intact, but the genomic rearrangement places a new promoter and 5′-UTR upstream of the main coding sequence, leading to aberrant expression of the proto-oncogene.

RNA. Forced expression of *PLAG1* in NIH3T3 cells has demonstrated that this protein can induce the standard characteristics of neoplastic transformation including loss cell–cell contact inhibition, anchorage-independent growth, and tumor formation in nude mice xenografts [61]. This suggests that the constitutive activity of the CTNNB1 promoter leads to sufficient PLAG1 expression for malignant transformation in PA.

### 2.3. PRCC–TFE3

As cloning and molecular strategies improved in the early 1999's, another recurrent gene fusion would soon be described in papillary renal cell carcinoma (PRCC), the second most common carcinoma of the renal tubules accounting for 15–20% of all renal cell carcinomas [62–66]. Karyotypic analysis as early as 1986 (de Jong et al.) led to the identification of abnormalities in the Xp11.2 region characterized by a genomic rearrangement, t(X;1)(p11.2;q21.2) [62–66]. Interestingly, before any of the genes surrounding the breakpoint were cloned a gene encoding TFE3, which was originally identified by their ability to bind to μE3 elements in the immunoglobin heavy chain intronic enhancer [67], was mapped to the Xp11.22 locus [68], and later shown to encode a member of the basic helix–loop–helix followed by a leucine zipper family (bHLHzip) of transcription factors. After the original genomic mapping, *TFE3* was soon identified at the translocation breakpoint by southern blot analysis [69]. Subsequent 5′-RACE identified *PRCC*; a ubiquitously expressed gene that encodes a protein with a high proportion of prolines and glycines — including three P-X-X-P motifs that are known to interact with SH3 domains [70,71]. Interestingly, the fusion event leading to the *PRCC–TFE3* rearrangement also results in a reciprocal *TFE3–PRCC* gene fusion [69,72].

To elucidate the properties of these reciprocal gene fusions, Weterman et al. introduced wild type *PRCC*, wild type *TFE3*, *PRCC–TFE3* and *TFE3–PRCC* expression vectors into COS cells and postulated that only the *PRCC–TFE3* gene fusion was responsible for tumor formation based on its ability to activate a generalized report assay [73]. Thus, the *PRCC–TFE3* genomic rearrangement is type 3 as the fusion protein gained a novel function through rearrangement. However, fusions of the *PSF* or *NonO* pre-mRNA splicing factors are also recurrently fused to *TFE3*, albeit at a much lower frequency than PRCC [69,72,74], suggesting that the *TFE3* portion of the fusion is responsible for malignant transformation. Subsequent transcriptional activation assays demonstrated that of the *PSF–TFE3*, *NonO–TFE3* and *PRCC–TFE3* chimeras, only the *PRCC–TFE3* fusion protein could activate the plasminogen activator inhibitor-1 (*PAI-1*) promoter [75], suggesting that only this gene fusion retains transcriptional activity. However, recent co-immunoprecipitation experiments demonstrated that antibodies against the pre-mRNA splicing factors SC35, PRL1, and CDC5 were able to immunoprecipitate wild type PRCC, and an anti-SM antibody was able to immunoprecipitate the PRCC–TFE3 fusion protein [75]. This data suggests that the fusion protein functions may partially function through transcriptional pathways, it may also function by altering pre-mRNA splicing, but more conclusive experiments need to be conducted to demonstrate this phenotype.

### 2.4. HMGA2, evading let-7

While most of the gene fusions discovered until this point including PRCC–TFE3 were thought to define specific epithelial tumor types, a new gene fusion that was associated with several different tumor types, including pleomorphic adenoma (PA) (see above), lipoma, uterine leiomyoma and some myeloid malignancies [76], would refute the notion. In fact, the discovery of translocations involving 12q15 had been established by karyotypic analysis in multiple tumor types before the rearranged genes were actually identified and one of the genes involved in the t(9;12)(p12–22;q13–15) PA translocation was first identified in both mesenchymal tumors [77] and lipomas [78]. This first gene to be described was the 5′ gene fusion partner, *HMGA2* (high mobility group AT-hook 2), belongs to the non-histone chromosomal high mobility group (HMG) protein family, which are small nuclear proteins (<30 kDa) that undergo

extensive post-translational modifications and contain nine amino acid segments that bind AT-rich DNA stretches in the minor groove (AT-hooks) (reviewed in [79]). Subsequent 3′ RACE of tumor samples revealed that HMGA2 has two different 3′ partners in PA, *FHIT* and *NFIB*, both of which contribute very little coding sequence to the resulting fusion gene. In fact, in one class of translocations, *HMGA2* exon 3 is fused to *FHIT* exon 9 or 10, resulting in retention of the C-terminal 26 amino acids of *FHIT* [80], and in the other set, *HMGA2* exon 3 or 4 fusion to *NFIB* exon 9 appends five amino acids (SWYLG) to the truncated *HMGA2* protein [81].

Surprisingly, transgenic mice overexpressing wild type *HMGA2* were observed to have similar phenotypes to mice expressing the truncated protein HMGA2 protein found in the PA gene fusions [82–84]. To complicate this observation, in hereditary renal cell carcinoma, *FHIT* was previously demonstrated to be fused to the patched related gene *TRC8* by t(3;8)(p14.2;q24.1) [85,86] and the (SWYLG) amino acid motif found in the *HMGA2–NFIB* gene fusion were shown to be essential for *NFIB* function [81]. Recent research, however, has shed light onto the importance of these translocations to neoplastic transformation.

The discovery that small RNAs called microRNAs can negatively regulate gene expression through direct binding to a gene's 3′-UTR has led to the hypothesis that certain microRNAs can function as tumor suppressors in cancer [87]. Bioinformatic analysis of the HMGA2 3′-UTR demonstrated that the mRNA contains seven conserved sites complementary to the *let-7* microRNA [88] (depicted in Fig. 3). To show that the let-7 microRNA negatively influences HMGA2 expression, Mayr et al. built a HMGA2 3′-UTR conjugated luciferase reporter and demonstrated that *let-7* represses its expression [89]. As such, although the genomic rearrangements between HMGA2 and *FHIT* or *NFIB* yield fusion proteins, replacement of a Let-7 regulated 3′-UTR seems to be the critical event because it leads to HMGA2 overexpression, which is sufficient for neoplastic transformation. Thus, the *HMGA2* genomic rearrangement represent the first of a novel class of gene fusions, type 2, in which fusion gene activity is enhanced by loss of mRNA level regulation (Fig. 3).

### 2.5. Pax8–PPARγ

In 2000, Kroll et al. employed fluorescence in situ hybridization (FISH), yeast artificial chromosome mapping and 3′ RACE to identify genes involved in a genomic rearrangement, t(2;3)(q13;p25) [90], that was originally identified by karyotype analysis of follicular thyroid carcinomas, a subset (10–20%) of all thyroid malignancies [91]. This translocation is thought to be specific to FTC as it has not been reported in other thyroid tumors or hyperplastic nodules [92]. In the resulting gene fusion, the *Pax8* (Paired box gene 8) gene is fused to *PPARγ* (Peroxisome proliferator-activated receptor-γ), a ubiquitously expressed transcription factor [90]. The Pax8 protein is involved in thyroid follicular cell development and regulation of thyroid-specific gene expression [93]. *PPARγ* plays a major role in a number of different diseases including obesity, atherosclerosis, diabetes as well as cancer (reviewed in [94]). Because *Pax8* is a thyroid-specific transcription factor and because its DNA binding domain is fused to the c-terminal domains of *PPARγ* [90], the resulting protein chimera is thought to have constitutive re-distribution of *PPARγ*-directed

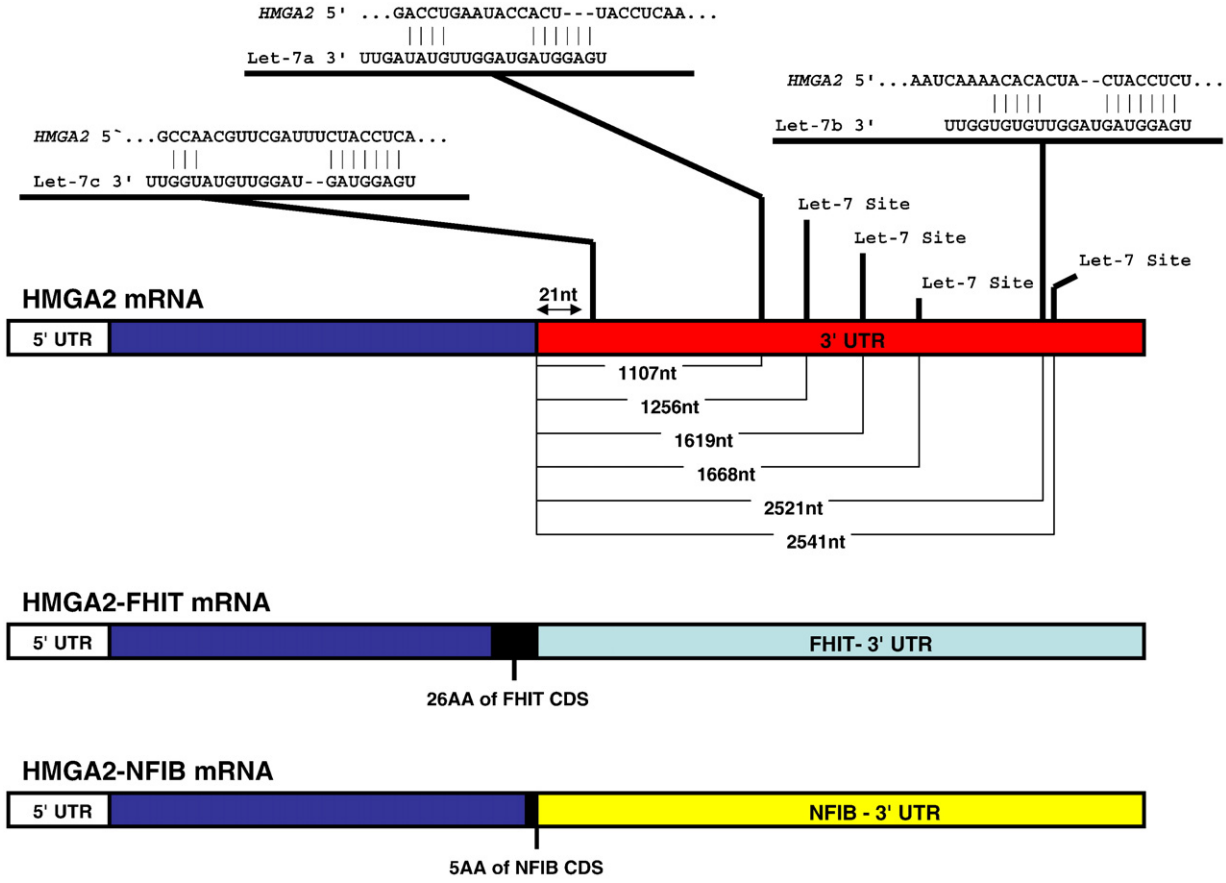



**Fig. 3.** *HMGA2* gene fusions elude the Let-7 family of microRNAs. The HMGA2 mRNA structure is shown along with putative Let-7 family binding sequences in the HMGA2 3′-UTR. Results were predicted by TargetScan [202] and three representative microRNAs are shown with there highest probability binding sites of the seven total predicted sites along the 3′ UTR. Distance to each predicted binding site is annotated as nucleotides from the start of the 3′UTR. Below the wild type HMGA2 mRNA are the HMGA2–FHIT and HMGA2–NFIB mRNAs that result from these two gene fusions. TargetScan did not predict any microRNA binding sites in these genes. As such, the *HMGA2* gene fusions represent a second class of gene fusions in which the recombination event allows the proto-oncogene mRNA to evade microRNA-mediated silencing.

transcription. In 2005, gene-expression microarray profiling revealed a distinct signature in follicular thyroid carcinomas harboring the *Pax8–PPARγ* gene fusion in which cell growth and chromatin remodeling pathways were over-represented and protein biosynthesis pathways were under-represented as compared to follicular thyroid carcinomas without the translocation [95], suggesting that *PPARγ*-transcription is indeed redefined by the gene fusion.

Interestingly, follicular thyroid carcinomas were originally thought to arise from disruption of distinct molecular pathways, either through the fusion of *Pax8* to *PPARγ*, or through the acquisition of point mutations leading to the constitutive activation of the G-protein RAS. In fact, one study reported that 16/33 (49%) of follicular carcinomas had *RAS* mutations, 12/33 (36%) had *Pax8–PPARγ* rearrangement, only 1/33 (3%) had both, and 4/33 (12%) had neither [96]. However, in 2006, quantitative reverse transcription PCR analysis of follicular carcinoma clinical samples demonstrated loss of the tumor suppressor *NORE1A* in samples harboring the *Pax8–PPARγ* rearrangement, but not in other samples [97]. Because *NORE1A* binds to the GTP bound (activated) RAS protein and suppresses *RAS* activity, this discovery suggested that activation of the RAS pathway is a critical event in pathogenesis of thyroid carcinoma that is altered either directly by activating mutation, or indirectly by the *Pax8–PPARγ* rearrangement.

### 2.6. BRD–NUT

Soon after the discovery of the *Pax8–PPARγ* rearrangement, the translocation t(15;19)(q13;p13.1) was identified in a rare, highly aggressive carcinoma arising in the midline organs and upper respiratory tract of young people now termed nuclear protein in testis (NUT) midline carcinomas (NMC) [98–100]. *BRD4*, which contains the chromosome 19 breakpoint, has two annotated transcripts encoding either short or long forms of the protein that both contain N-terminal bromodomains. The longer *BRD4* transcript encodes a ubiquitously expressed 200 kDa nuclear protein [101] with a c-terminal lysine rich region that is not found in the shorter transcript. The translocation resulting in fusion to the *NUT* gene (identified by southern blot analysis) only disrupts the longer *BRD4* transcript resulting in the loss of the lysine rich region in the fusion oncogene. Several studies of *BRD4* in both murine and human cell line models have demonstrated a critical role in cell cycle progression and cell proliferation [102,103]. In fact, Brd4 enhances cell growth by interacting with chromatin [104], replication factor C [102] and cyclinT1 and CDK1 that constitute core positive transcription elongation factor b (P-TEFb) [105]. Likewise, chromatin immunoprecipitation assays demonstrated that Brd4 is required to recruit P-TEFb to active promoters, and that increased Brd4 leads to increased P-TEFb-dependent phosphorylation of RNA polymerase and enhanced transcription from promoters *in vivo* [105].

More insight into the role of the BRD4–NUT fusion protein in NMC biology came from a screen for other NMC gene fusions. Because the *BRD4–NUT* translocation defines two-thirds of all NMCs, French et al. used a candidate gene approach to screen other NMC samples and discovered another recurrent translocation between *BRD3* and *NUT* that defined large portion of the remaining NMC cases [106]. The *BRD3–NUT* fusion gene encodes a protein highly similar to that encoded by the *BRD4–NUT* transcript. It is composed of two tandem chromatin-binding bromodomains, an extra-terminal domain, a bipartite nuclear localization sequence, and a significant portion of *NUT* coding sequence. As such, the conserved protein structure gave insight into the mechanism by which the chimeric protein induces neoplastic properties.

Wild type NUT, which is normally only expressed in the testis [99], contains both nuclear localization and export signal sequences and is shuttled between the nucleus and cytoplasm via a leptomycin-sensitive pathway [106]. Importantly, however, the Brd3–NUT and Brd4–NUT proteins are retained in the nucleus, suggesting that interactions between the Brd3 or Brd4 bromodomains and chromatin are essential to the fusion protein [106] (Fig. 4). Further evidence for this hypothesis comes from an siRNA experiment in which knockdown of Brd–NUT fusion transcripts in NMC cell lines resulted in squamous differentiation and cell cycle arrest [106]. This suggested that the nuclear retention of NUT, not the loss of the Brd C-terminal domain, is responsible for promoting NMC carcinogenesis [106]. The realization that *Brd–NUT* gene fusions define a class of translocations that fuse bromodomains to the NUT protein suggests that oncogenic translocations will arise from multiple partners when critical domains are present in more than one gene.

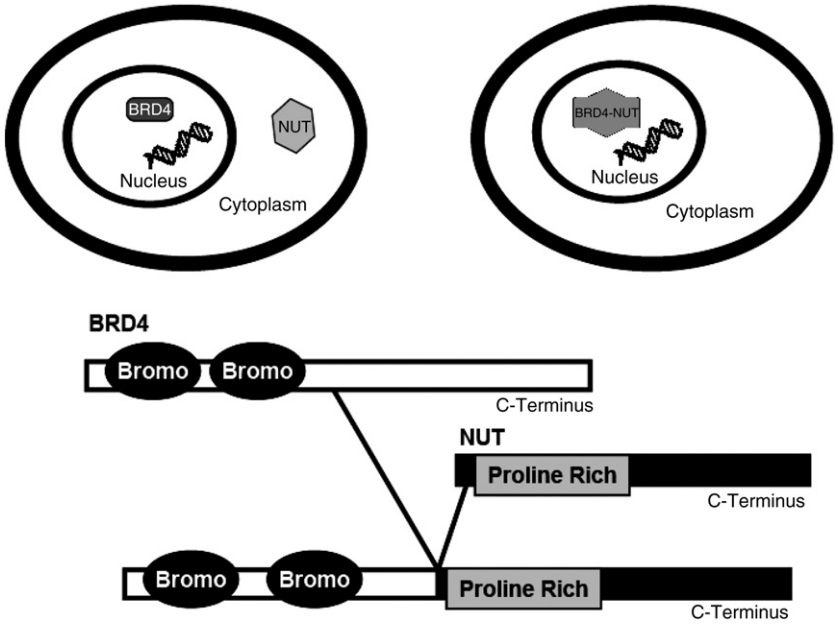



**Fig. 4.** Nuclear retention of NUT. The *BRD4–NUT* gene fusion represents a third class of rearrangements in which the resulting protein gains activity to become a proto-oncogene. In this case, the two bromodomains of BRD4 are fused to NUT. Although NUT usually cycles between the nucleus and cytoplasm in a highly controlled manner, appendage of the BRD4 bromodomains to the majority of the NUT protein lead to nuclear retention of the protein and aberrant activity.

### 2.7. ETV6–NTRK3

The first major example of a recurrent epithelial rearrangement that appeared not only in multiple tumor types, but had also been reported in a large subset of hematologic malignancies was detected in several cases of secretory breast carcinoma, a rare subtype of infiltrating ductal carcinoma affecting both children and adults [107]. Tognon et al. detected the *ETV6–NTRK3* fusion by comprehensive FISH analysis in 92% (12 of 13) secretory breast carcinoma cases [108]. *ETV6* (also *TEL*) is an ETS family member that is involved in a large number of fusions to either a transcription factor like *AML1* [109] or to a protein tyrosine kinase domain like that of *ABL* [110,111], *JAK2* [112–114], *ARG* [115,116], *PDGFRβ* [117] or *FGFR3* [118], each of which define a unique leukemia subtype (reviewed in [119]). *ETV6* contains a pointed oligomerization domain (PNT; also known as sterile alpha motif, SAM, or helix–loop–helix, HLH) and an ETS DNA binding domain, the expression of which is required for developmental processes such as hematopoiesis and yolk sac angiogenesis [120]. *NTRK3* is a transmembrane neurotrophin-3 surface receptor that contains a c-terminal protein tyrosine kinase domain and plays a role in growth, development, and cell survival of neural cells in the central nervous system (reviewed in [121]). The fusion of the N-terminal *ETV6* pointed domain to the C-terminal tyrosine kinase domain of *NTRK3* was first reported in congenital fibrosarcoma (CFS) [122], but has since been reported in multiple cell lineages including those that give rise to congenital mesoblastic nephroma (CMN), acute myelogenous leukemia, and secretory breast carcinoma [108] (reviewed in [123]).

Following the initial discovery, research focused on the transforming ability of the recombination product. By using retroviral gene delivery methods, the *ETV6–NTRK3* fusion gene was shown to be sufficient to induce the non-tumorigenic murine breast cell lines Eph4 (epithelial) and Scg6 (myoepithelial) as well as NIH-3T3 fibroblasts to form tumors, glandular structures and to express epithelial antigens [108]. This discovery suggested that the fusion gene acts as a dominant oncogene in secretory breast cancer. *ETV6–NTRK3* was also shown to inhibit TGF-β tumor suppressor activity in NIH3T3 cells [124], suggesting that it most likely regulates microRNA biogenesis indirectly [125], but this has not yet been explored. Although it is known that adults have a less favorable prognosis than children and distant metastases are rare [126], local recurrences and nodal metastases have been observed [127] suggesting that the gene fusion leads to an invasion-associated transcriptional program, but this also has not been explored. Despite this, it is known that constitutive activation of the fusion protein leads to activation of the Ras-mitogen-activated protein kinase (MAPK) pathway and the phosphoinositide-3-kinase (PI3K)–AKT pathway, the mechanism leading to activation of these pathways has remained elusive until recently, when the fusion protein was shown to associate with c-Src by immunoprecipitation from fusion-positive CFS and CMN human primary tumors [128]. More recently, however, a mouse knockin model was created by introducing the human *NTRK3* cDNA into exon 6 of the mouse *ETV6* locus, which induced a fully penetrant, multifocal breast cancer [129]. By using microarray analysis of unsorted and sorted tumors from this model, as well as NIH3T3 cells transduced with the fusion gene, the authors showed that *ETV6–NTRK3* enriches for WNT target genes through activation of the AP1 complex [129]. The requirement for AP1 activity in *ETV6–NTRK3*-mediated transformation was confirmed by showing that the co-expression of a dominant negative component of AP1 complex, c-JUN TAM67, with the gene fusion blocked tumorigenic properties both *in vitro* and *in vivo* [129]. The *ETV6–NTRK3* gene fusion represents one of the last gene fusions to be discovered by traditional biological techniques.

### 2.8. TMPRSS2–ETS

In 2005, advances in bioinformatics led to the discovery of rearrangements on chromosome 21 between *TMPRSS2* (transmembrane protease, serine 2) and *ERG* (v-ets erythroblastosis virus E26 oncogene homolog (avian)) resulting in the *TMPRSS2–ERG* gene fusion. Thus far, genomic rearrangements leading to an *ERG* gene fusion have been reported in approximately 50% of clinically localized prostate cancers published (reviewed in [130]). *TMPRSS2* is a prostate-specific, androgen-regulated gene [131–133] that has two annotated transcription variants, both of which are involved in the fusion with *ERG*; the annotated *TMPRSS2* in about 50% of the gene fusions, an alternative *TMPRSS2* variant in 10% of gene fusions, and both variants in slightly more than 40% of analyzed gene fusions [134]. *ERG* belongs to the ETS family of transcription factors and has two transcription variants that differ only slightly in the 5′-UTR (deleted in the gene fusion) and in the usage of an in-frame exon, the role of which remains undefined. The most common *TMPRSS2–ERG* gene fusion variants involve *TMPRSS2* exon 1 or 2 fused to *ERG* exon 2, 3, 4, or 5 [134–143] and less frequently rearrangements of *TMPRSS2* exon 4 or 5 fused to *ERG* exon 4 or 5 [141]. In line with the combinatorial complexity of *TMPRSS2–ERG* rearrangements, different fusions have correlated with slightly different phenotypic outcomes. For example, *TMPRSS2* exon 2 fused with *ERG* exon 4 is associated with aggressive disease, while others have been associated with seminal vesicle invasion and poor outcome [143].

Like *TMPRSS2*, the *TMPRSS2–ERG* gene fusion is androgen-regulated in an androgen-responsive cell line (VCAP) carrying the rearrangement [135], but not in an androgen-insensitive cell line harboring the fusion (NCI-H660) [144]. We have shown that VCaP cells and benign prostate cells forced to overexpress ERG drive components of the plasminogen activation pathway to mediate cellular invasion using transwell migration assays [145]. We have also reported that primary or immortalized benign prostate epithelial cells overexpressing *ERG* have a transcriptional program with high levels of several invasion-associated genes, but did not display phenotypic increases in cellular proliferation or anchorage-independent growth [145]. Despite this, one group recently identified *c-MYC* as a downstream target of ERG and demonstrated that ERG knockdown in *TMPRSS2–ERG* expressing CaP cells resulted in loss of cell growth *in vitro* and loss of tumorgenicity *in vivo*, with only 22% (2/9) mice developing detectable tumors at day 42 in siRNA treated cells as compared to 100% (5/5) in the control group [146]. Interestingly, transgenic mice expressing an androgen-regulated ERG fusion gene develop mouse prostatic intraepithelial neoplasia (PIN), a precursor lesion of prostate cancer, not prostate cancer. Taken together with our *in vitro* data, these results suggest that, without secondary molecular lesions such as loss of the tumor suppressors *PTEN* or *NKX3-1*, the *TMPRSS2–ERG* gene fusion may not be sufficient for transformation [145,147,148].

Although *ERG* clearly participates in the majority of ETS family gene fusions in prostate cancer, other ETS family members including *ETV1* [135], *ETV4* [149,150] and *ETV5* [151] also contribute to gene fusions in prostate cancer, albeit at a much lower frequency. In contrast to *TMPRSS2*, which is the only known 5′ partner to *ERG*, the other ETS family members may have a variety of 5′ partners including those with androgen-responsive promoters (*TMPRSS2*, *SLC45A3*, *KLK2*, *HERV-K_22q11.23* and *CANT1*), one with an androgen-insensitive promoter, but a constitutively active promoter (*HNRPA2B1*), and one with an androgen-repressed promoter (*C15orf21*) [135,149,151–153]. As in the case of *ERG*, forced expression of *ETV1* under the control of a CMV promoter did not enhance cell proliferation in benign prostate epithelial cell lines and did not lead to anchorage-independent colony formation in soft agar, but did lead to the enrichment of genes associated with invasion [145]. Consequently, knockdown of *ETV1* in LNCAP cells prevented transwell invasion through matrigel [145,154].

### 2.9. EML4–ALK

Recently, Soda et al. reported a retroviral-mediated transformation screen, in which they created a cDNA expression library from a

surgically resected lung adenocarcinoma [155]. Following transformation of NIH3T3 cells, cDNAs were recovered from cells by PCR amplification and sequenced. One of these sequenced transcripts contained a fusion between *EML4* (echinoderm microtubule-associated protein-like 4) and ALK (anaplastic lymphoma kinase) that was later confirmed as an inversion of chromosome 2p in 6.7% (5 of 75) NSCLC patients [155]. Wild type *EML4* is a member of the EMAP family of proteins and the amino-terminus (amino acids 1–249) were previously demonstrated to be essential for microtubule formation in HeLa cells [156]. *ALK* encodes a tyrosine kinase and a MAM domain (a domain frequently found on the extracellular side of the membrane on many receptors). Despite the apparent low frequency EML4–ALK gene fusions in NSCLC, the transforming ability of EML4–ALK gene fusion variant 1, 2, and 3b, but not a kinase inactive mutant (K589M) has been demonstrated by engrafting NIH-3T3 cells infected with retroviral expression vectors and showing that tumors arise in 8/8 mice from all groups except for the kinase dead mutant [157].

To corroborate the low frequency *EML4–ALK* rearrangements in NSCLC, careful PCR-based analysis was completed on NSCLC cases to identify novel in-frame *EML4–ALK* gene fusions that led to the identification of two novel fusion isoforms called variant 3a and 3b [157]. Even more recently, analysis of a cohort of 253 lung adenocarcinoma patient samples identified two new *EML4–ALK* fusions in which either exon 14 or exon 2 of *EML4* was fused to Exon 20 of *ALK* (variants 4 and 5, respectively), however, only 4.35% of patients were found to express any of the 5 known *EML4–ALK* genomic rearrangements [158]. A similarly low rate of the *ELM4–ALK* fusion was reported in a study of 104 lung cancer surgical specimens with only one fusion-positive case [159] and, in a study of different lung cancers, the fusion was identified in 3.4% (5 of 149) adenocarcinomas, but not in 48 squamous cell carcinomas, 3 large-cell neuroendocrine carcinomas, or 21 small-cell carcinomas [160]. However, this is to be expected, given the small sample size from non-adenocarcinomas. The *ALK* gene has previously been identified as the 3′ fusion partner of *NPM*- [161], *TPM3*- [162], *CLTC*- [163], *ATIC*- [164–166] and *TFG*- [167]. In light of this observation, RT-PCR analysis was used to screen all known hematologic *ALK* fusion partners in a cohort of 77 NSCLC samples, however, no redundant fusion partners were identified and only 2.6% (2 of 77) of NSCLC cases harbored the *EML4–ALK* fusion [168]. To supplement the existing RT-PCR data in the literature, our group developed a break-apart FISH assay to analyze *ELM4–ALK* fusion as well as the amplification of each gene. We reported the fusion occurred in less than 3% of NSCLC cases analyzed, and that, in most cases harboring the lesion, not all cells exhibited the fusion. We also found that *EML4* and/or *ALK* amplification occurred, indicating that other mechanisms of genomic rearrangement leading to amplification may arise [169].

### 2.10. SLC34A2–ROS

In 2007, a survey of phosphotyrosine signaling in lung cancer not only led to the re-identification of the *EML4–ALK* fusion, but also the discovery of a novel translocation between chromosomes 4p15 and 6q22, in which the transmembrane domain containing N-terminal region of the solute carrier family 32, member 2 (*SLC34A2*) is fused to an N-terminal transmembrane domain of the c-ros oncogenes 1 (*ROS*), respectively, in the lung cell line HCC78 [170]. SLC34A2 is encoded from a single transcription variant and ROS, which is a type I integral membrane-bound tyrosine kinase and is a known oncogene that is highly expressed in several tumor cell lines, and also encoded from a single transcript. Interestingly, while the authors did not identify *SLC34A2* rearrangements with *ROS* in patient samples, a gene fusion between *CD74*, located at 5q32, and *ROS* was observed, in which the tandem transmembrane domain structure was again observed [170]. This suggests not only that *ROS* is another promiscuous gene fusion partner, but the tandem transmembrane structure is one mechanism leading to constitutive activation of the tyrosine kinase. Indeed, forced expression of the *SLC34A2–ROS* chimera demonstrated constitutive kinase activity in the cellular membrane fraction [170].

### 2.11. SLC45A3–ELK4

With the recent advent of next generation sequencing technology (described below), our group has recently identified another recurrent gene fusion in prostate cancer [171]. Using this technology we identified the fusion of *SLC45A3* to *ELK4*, an ETS family member. Here exon 4 of *SLC45A3* is fused to exon 1 of *ELK4*. Interestingly, this novel gene fusion was identified from the RNA of a cell line harboring a known gene fusion involving another ETS family member gene, *ETV1*. Likewise this novel gene fusion involves *SLC45A3*, which is known to fuse with *ETV1* in other prostate cancer cases. Unlike other gene fusions described to this point, *SLC45A3–ELK4* seems to result from polymerase read-through and intergenic splicing rather than genomic rearrangement as no detectable alterations were detected on the DNA level by fluorescence in situ hybridization (FISH), array comparative hybridization (aCGH) or high-density single nucleotide polymorphism (SNP) arrays [171]. RNA level gene fusions were recently identified in endometrial stromal tumors and are discussed below.

## 3. Lessons from MLL translocations

While the list of epithelial derived gene fusions continues to expand, it is important to highlight unique mechanisms of oncogene formation through specific genomic rearrangements from the hematological malignancies. Translocations altering the mixed-lineage leukemia (*MLL*) gene on 11q23 frequently lead to fusions with over 40 different genes on different chromosomes with MLL-AF4 and MLL-AF9 among the most frequent chimeras (reviewed in [172,173]). Interestingly, different MLL fusions are highly associated with either acute myeloid leukemia (AML) or acute lymphoid leukemia (ALL, depending on the fusion partner [174]. *MLL* is the mammalian homologue of a *Drosophila* gene called *trithorax,* which was shown to play a critical role in axial morphogenesis and patterning during embryogenesis through the regulation of *HOX* genes (*HOM-C* in *Drosophila*) [175,176]. Multiple studies have suggested that deregulation of *HOX* gene expression contributes to leukemogenesis [177]. Additionally, retroviral transduction of a MLL fusion gene construct was able to transform wild type, but not the Hoxa9-deficient, bone marrow cells providing direct evidence that specific *HOX* gene expression may be required for leukemogenesis [178]. Because MLL chimeras often lose large fragments and different domains from either the N- or C-terminal regions, the seemingly critical role of MLL-associated *HOX* gene expression to leukemogenesis led to the question of whether the molecular mechanisms by which wild type MLL regulates gene expression are mutually exclusive from those employed by MLL chimeras [179].

As the molecular mechanisms of MLL target gene regulation continue to unravel, several studies have shed light on the fact that molecular function between wild type and fusion gene settings may be unique, though the outcome of gene activity is ultimately similar. Wild type MLL encodes a multi-domain protein with three AT-hooks used for binding AT-rich DNA sequences and a histone methyltransferase domain [180] and assembles into supercomplexes containing several different chromatin remodeling enzymes on target DNA motifs like those found in *HOX* genes [181]. Chimeric MLL proteins, on the other hand, appear to utilize different mechanisms to modulate *HOX* gene expression and initiate leukemogenesis. For example, fusion of coiled-coil domains from GAS7 or AF1p to MLL endow the chimeric protein with the ability to dimerize on the target gene promoters and have been suggested to stimulate transcription through the inappropriate recruitment of members of the MLL supercomplex [182].

This suggested that preventing dimerization of the coiled-coil domains with targeted small molecules could inhibit MLL activity in this subset of MLL fusions. In contrast, some MLL fusions lead to constitutive nuclear retention while maintaining similar binding patterns as the dimerizable MLL chimeras on the HoxA9 locus [183]. In the absence of a partner gene, MLL can acquire an in-frame partial tandem duplication (PTD) of exons 5 through 11 (occurring in approximately 4%–7% of AML cases) that causes overexpression of HoxA7, HoxA9, and HoxA10 in spleen, BM, and blood in a knockin mouse model [184]. As such, altering downstream *HOX* gene expression appears to be one critical role of MLL gene fusions and rearrangements.

Given that wild type and chimeric MLL proteins appear to accomplish at least one similar molecular function (*HOX* gene regulation), the question of how epithelial gene fusions will function in comparison to their wild type counterparts remains intriguing. For example, we have very little understanding of the normal molecular mechanisms utilized by ERG and ETV1 to control gene expression (prostate cancer gene fusions, discussed above), let alone the critical co-factors required for transcriptional regulation. Although we may expect the molecular mechanisms of ERG and ETV1 mediated gene regulation to be the same in the wild type and fusion settings (because the encoded proteins are nearly identical), this remains to be proven. Perhaps the ability to design rational drug targets against specific fusion proteins without obvious molecular susceptibilities (like the tyrosine kinase activity of BCR–ABL) will depend as much on our understanding of each fusion protein's function and critical co-factors as on their downstream targets.

## 4. Difficulty in identifying epithelial cancer gene fusions

With the discovery of the *TMPRSS2–ERG* gene fusion in prostate cancer, we look back on the history of cancer biology and wonder why gene fusions have not been identified in some of the most well-studied epithelial cancers? Part of the problem was methodological, as the chromosome quality in epithelial neoplasms is very poor when compared to hematologic neoplasms. However, cytogenetic techniques have improved dramatically since the discovery of the "minute" chromosome in 1960 [6]. In fact, in the 1960s, chromosome patterns in epithelial tumors were already being described as abnormal [185] and it was often thought that the degree of cytogenetic changes corresponded proportionally with clinical progression [186], making the identification of individual and recurrent translocations difficult. In fact, the idea that the induction of genomic instability is a critical and intended step in the malignant progression of solid tumors has gained considerable momentum [187,188]. Recently, it was demonstrated that overexpression of Separase, a protein that is overexpressed in a subset of breast cancers, leads to can induce chromosome instability and aneuploidy in the mutant p53 mouse mammary epithelial cell line FSK3 [189]. Likewise, deregulation of Mad2, which regulates separase activity, has been shown to promote chromosomal instability, induce aneuploidy and lead to tumorigenesis [190]. Interestingly, once Mad2-induced neoplastic transformation has occurred, Sotillo et. al. demonstrated that expression of Mad2 is no longer required for tumor progression suggesting that the induction of chromosomal instability could be a transient event in oncogenesis [190]. In fact, it is possible that specific gene fusions induce genomic instability through deregulation of normal mitotic events like separase or Mad2 activity or through novel mechanisms yet to be described. If induction of chromosomal instability was a mechanism of oncogenesis employed by a specific gene fusion, then induction of other secondary "carrier" chromosomal rearrangements would simply serve to mask the identification of the recurrent genetic rearrangement. Such a progression pattern in epithelial tumors could explain the complex heterogeneity often observed in such malignancies (Fig. 5). In contrast, leukemias, lymphomas and mesenchymal tumors

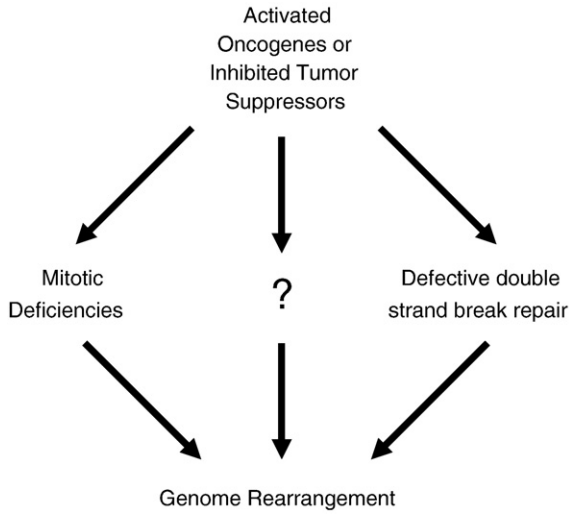



**Fig. 5.** Difficulty in discovering gene fusions. One possibility is that a critical function of oncogenes in epithelial cancers is to alter genomic structure and it has been suggested that such changes could lead to cancer progression. However, if such a model were true, it would give a reason for the genomic heterogeneity observed in epithelial cancers that has allowed recurrent gene fusions to go unnoticed in solid tumors.

are almost 95% clonal [191]. As such, the complexity and shear number of genomic rearrangements in epithelial malignancies has led to difficulty in defining primary aberrations in these neoplasms. This difficulty eventually led to the incorrect notion that genomic rearrangements leading to gene fusions were simply less common in epithelial tumors.

## 5. Mitelman hypothesis

In order to address this notion that fusion genes are almost exclusively a hematologic phenomena, Mitelman et al. completed a comprehensive study of all known cytogenetically abnormal neoplasms reported in the literature [192]. Importantly, data published by the group supported the counter-hypothesis that, in every tumor type, the numbers of recurrent balanced chromosome abnormalities, gene fusions and balanced rearrangements are a function of the total number of analyzed cases [192]. In this study, 271 gene fusions and 59 potential gene fusions (only one gene identified at the breakpoint) were catalogued, of which 275 unique genes were involved in the rearrangements [192]. This indicated that a substantial number of genes were present in more than one chimeric transcript (e.g., *MLL*, *ETV6* and *RET* as described above). In classifying each gene fusion by the class to which each member of the chimera belonged, the group demonstrated that the proportion of fusions belonging to each class was approximately equal in both hematologic and solid tumor malignancies, with the transcription factor class accounting for 38–44% and tyrosine kinase class tabulating 5–7% [192]. This study suggested that the occurrence of gene fusions is a general molecular event that has no fundamental tissue-specific differences. However, gene rearrangements must at least encourage function in specific genetic backgrounds such as the TMPRSS2–ERG fusion, which requires active androgen signaling, and thus encourages prostate specificity.

## 6. Tissue-specific gene fusions

The idea that genomic rearrangements are tissue-specific is an emerging concept in the field of gene fusion biology. For example, TMPRSS2 is a strongly androgen-regulated and prostate-specific gene that is fused to the ETS family members ERG and ETV1 in prostate cancer [135]. While other ETS family members form fusion genes

**Table 1**
Chromosomal rearrangements in epithelial cancers.

| Malignancy | Gene fusion | Chromosome rearrangement | Method of discovery | Study | Ref. |
|---|---|---|---|---|---|
| Follicular thyroid carcinoma | PAX8–PPARγ | t(2;3)(q13;p25) | Primary tumor karyotypic analysis/FISH/3′ RACE | Kroll et al. | [90] |
| Midline carcinoma | BRD3–NUT | t(9;15)(q34;q14) | Candidate gene FISH Screen | French et al. | [106] |
| | BRD4–NUT | t(15;19)(q14;p13) | Primary tumor karyotypic analysis/FISH/southern blot | French et al. | [98] |
| Non-small-cell lung cancer | EML4–ALK | inv(2p) | Transformation assay/direct sequencing | Soda et al. | [155] |
| | TFG–ALK | t(2;3)(p23;q12) | Tyrosine Kinase Activity Screen/5′ RACE | Rikova et al. | [170] |
| | SLC34A2–ROS | t(4;6)(p15;q22) | | | |
| Papillary renal cell carcinoma | PRCC–TFE3 | t(X;1)(p11;q23) | Primary tumor karyotypic analysis/southern blot/5′ RACE | Sidhar et al. | [69] |
| Papillary thyroid carcinoma | RET–NTRK1 | t(1;10)(q21;q11) | Transformation assay/direct sequencing | Martin-Zanca et al. | [24] |
| Pleomorphic adenoma | CTTNB1–PLAG1 | t(3;8)(p21;q12) | Primary tumor karyotypic analysis/ Breakpoint mapping/southern blot/5′ RACE | Kas et al. | [59] |
| | HMGA2–FHIT | t(3;12)(p14;q15) | Primary tumor karyotypic analysis/3′ RACE | Geurts et al. | [80] |
| | HMGA2–NFIB | t(9;12)(q24;q15) | Primary tumor karyotypic analysis/3′ RACE | Geurts et al. | [81] |
| Prostate cancer | TMPRSS2–ERG | del(21)(q22) | COPA/Exon walking/5′ RACE | Tomlins et al. | [135] |
| | TMPRSS2–ETV1 | t(7;21)(p21;q22) | | | |
| | TMPRSS2–ETV4 | t(17;21)(q21;q22) | | Tomlins et al. | [149] |
| | TMPRSS2–ETV5 | t(3;21)(p28;q22) | | Helgeson et al. | [151] |
| | SLC45A3–ELK4 | del(1)(q32) | Integrated high-throughput sequencing | Maher et al. | [171] |
| | DDX5–ETV4 | t(17)(q24;q21) | Candidate gene FISH Screen/5′ RACE | Han et al. | [150] |
| Secretory breast carcinoma | ETV6–NTRK3 | t(12;15)(q13;q25) | Primary tumor karyotypic analysis/FISH | Tognon et al. | [108] |

that give rise to other malignancies, chimeras between androgen-regulated genes and ETS genes have only been observed in prostate cancer [130]. Likewise, the ALK tyrosine kinase is frequently fused to multiple partners in hematopoietic (myelogenous leukemia), mesenchymal (congenital fibrosarcoma) and epithelial (secretory breast carcinoma) malignancies, but no redundant fusion partners have been identified across tissue types [159]. Retention of the TFE3 DNA binding domain in follicular thyroid carcinoma is another example of this, as TFE3 is a thyroid-specific transcription factor [93]. Importantly, little is understood about the molecular mechanisms leading to gene rearrangement and the underlying reasons that particular chimeras are formed recurrently. The idea that tissue-specific rearrangements occur by fusing highly transcribed genes holds promise and would at least partially explain the apparent tissue specificity observed in the formation of chimeric transcripts even between genes that are fused in multiple cancer types.

The idea that gene fusions are tissue-specific could have profound implications on the discovery of novel gene fusions. Clearly, however, gene fusions do not always confer tissue specificity. HMGA2 has a 3′-UTR that is negatively regulated by the Let7 microRNA and simply replaces its 3′-UTR through rearrangement with another gene (described above), therefore representing a gene fusion that most likely retains functionality in multiple tissue types. As such, while this concept may have its largest impact on underlying molecular mechanisms of newly discovered gene fusions, it will probably not alter the rate gene fusion discovery.

## 7. Discovery of novel gene fusions

Although the rate recurrent chromosomal rearrangement discovery in epithelial tumors has been modest, the recent discovery of gene fusions in prostate cancer has led to a renewed interest in gene fusions identification in other epithelial cancer subtypes. Perhaps the best explanation for the sudden increase in the characterization of recurrent gene fusions is the advent of novel technologies (Table 1). For example, the use of existing gene-expression data in the discovery of novel gene fusions was limited until the emergence of cancer outlier profile analysis (COPA), which ranks genes by normalizing expression values based on median absolute deviation of gene expression to accentuate outlier profiles (reviewed in [130]). When COPA was applied to gene-expression datasets in the Oncomine database [193–196], the analysis was able to identify several hallmark cancer related genes and led to the discovery of the ERG and ETV1 outlier profiles in prostate cancer [135]. Subsequent exon-walking quantitative PCR was used to demonstrate loss of the 5′ exons in both ERG and ETV1, giving rise to the notion that a gene fusion event was responsible for the outlier expression of these genes in prostate cancer. Finally, 5′-RNA ligase-mediated rapid amplification of cDNA ends (5′-RACE) was used to identify the 5′ untranslated region of TMPRSS2, a prostate-specific, androgen-regulated, transmembrane serine protease gene [131,132,197]. Fusion specific PCR and fluorescence in situ hybridization (FISH) were used to confirm the genomic rearrangement.

In contrast to using COPA and exon-walking quantitative PCR to identify fusion gene candidates, several labs are now employing next generation sequence methods wherein DNA or mRNA can be fragmented, sequenced and mapped to the genome in a matter of weeks to identify gene fusions. Various commercial platforms have been developed with the intent of sequencing as much of the genome or transcriptome as possible and are classified based on the length of the templates each platform sequences. Long read technologies, like 454, can sequence long templates (>1 kb) whereas short read technologies, like SOLEXA and SOLID, are currently capable of sequencing 35–50 nucleotide templates. At first glance, long read technologies may appear to have the advantage of making genome (or transcriptome) re-assembly much simpler than short read technologies. However, a major advantage of short read technologies is the depth of coverage, or the number of times a segment of the genome is sequenced, which is currently much higher for short read than long read technologies. As such, the choice of technology is still dependent on the scientific question.

If our question is to identify the best method for novel fusion gene discovery, we assume that sequencing the transcriptome space will be much efficient than sequencing cancer genomes. In theory, the discovery of gene fusions by long read technology will require sequencing across the actual gene fusion boundary of the chimeric transcript. In contrast, short read technologies may be able to identify gene fusions by two different methods. The first and most straight forward method is the identification of sufficient short reads that do not map directly to the transcriptome, but correspond to the gene fusion boundary; and these short reads should identify both contributing genes with high probability. Second, because transcripts are thought to be sequenced with a uniform distribution across the length of the transcript, except for at the extreme 5′ and 3′ ends, exon expression for each transcript can be analyzed. Genes involved in rearrangements, leading to chimeric transcripts, would be expected to lack any exon expression on one of the transcript ends. However, this method will need to be carefully developed, as mapping of short reads to duplicated sequences (or sequences that appear more than one time in the genome) remains challenging.

To test whether short or long read technology was better for the discovery of recurrent gene fusions, we recently sought to "re-discovered" the known gene fusions BCR–ABL1 and TMPRSS2–ERG by sequencing the RNA transcriptome of either the leukemia cell line K562 or the prostate cell line VCAP, respectively, with both short and long read platforms [171]. Initially both technologies were able to identify the known gene fusion from the sample, but were also able to identify several other candidate gene fusions. For example, the Illumina short read platform nominated 428 candidates from the VCAP cell line [171]. However, most of these candidates were likely to result from either trans-splicing [198], co-transcription of adjacent genes followed by intergenic splicing [199], or as a consequence of the sample preparation protocol. In order to reduce the list of potential candidate genes, we intersected the results of the two platforms to yield a much more condensed list. Indeed by integrating the short read and long read platforms rather than constraining the analysis to either short or long read technology, we were able to significantly reduce the percent of false positive gene fusions discovered [171].

In the future, an even newer adaptation of next generation sequencing will likely replace the current reliance on both short and long read technologies for fusion gene discovery. Paired end sequencing is a method in which short read technology is used to sequence nucleotides from both the 5′ and 3′ ends of 200–300 nucleotide fragments of the genome (or transcriptome). By sequencing both ends of a fragmented RNA, paired end sequencing enhances not only the reliability of mapping and assembly, but also maintains significant sequencing depth. In a manner similar to our recent integration of short and long read platforms, the use of paired end sequencing technology for gene fusion discovery should first be examined by comparing the ability of matched mate-pairs to identify known gene fusions from control samples. With paired end sequencing, a single sample preparation and individual sequencing run will hopefully provide sufficient coverage for gene fusion discovery and these improvements as well as other advancements in modern sequencing technologies will likewise lead to a dramatic improvement in our ability to identify novel, pathogenic gene fusions.

## 8. Lessons from the JAZF1–JJAZ1 chimera

Advances in sequencing technology will most likely lead to a rapid increase in the number of characterized gene fusions over the next few years. However, a much more pertinent question may address the reasons for chromosomal rearrangements leading to gene fusions. Could fusion transcripts be a part of normal cell biology? It is also plausible that tissue-specific fusions could impart growth advantages that allow a cell to survive traumatic stress. Nonetheless, while the underlying molecular mechanisms triggering genomic rearrangement are still unclear; we surmise that once a genomic rearrangement occurs, cells harboring favorable gene fusions will be selected over time.

Insight into the development of genomic rearrangements may come from fundamental observations made following the study of endometrial stromal (EMS) tumors. In 2001, a recurrent translocation t(7;17)(p15;q21) was demonstrated to occur in EMS tumors that led to expression of the chimeric JAZF1/JJAZ1 mRNA transcript [200]. Although the mechanism leading to this rearrangement remains unknown, a recent study demonstrated that trans-splicing of RNAs in normal human endometrial stromal cells can lead to the chimeric JAZF1/JJAZ1 RNA and protein independent of chromosomal rearrangement [201]. This observation suggests that certain gene fusions may be generated by trans-splicing of RNAs, which then lead to chromosomal rearrangement due to their pro-neoplastic nature. Interestingly, the group also demonstrated that the RNA trans-splicing event leading to the JAZF1/JJAZ1 chimera was inhibited at high concentrations of either estrogen or progesterone, further suggesting that certain RNA fusions may occur in a hormone-dependent manner. The question of whether or not other specific gene fusions arise due to abnormal exposure to specific hormones has not been studied.

## 9. Conclusions

A limited number of epithelial gene fusions have been described and the quest for novel recurrent gene fusions, like the discovery of *TMPRSS2–ERG* gene fusions in prostate cancer, may provide major advances in cancer research in the near future. Here, we have demonstrated that gene fusions lead to overexpression or constitutive activation of oncogenes by a variety of unique mechanisms including fusion of housekeeping or tissue-specific gene promoters to oncogenes, as in the case of *TMPRSS2* gene promoter and 5′-UTR to ERG or, as in the case of HMGA2, through evasion of a microRNA by replacement of an oncogene's 3′-UTR. Despite the multitude of mechanisms used by chimeric transcripts to drive malignancy, several important lessons can be taken from characterized epithelial gene fusions, studies of MLL translocations, as well as the very recent discovery of JAZF1–JJAZ1 RNA fusions, which precede genomic rearrangement in specific cell types.

As in the case of Imatinib and BCR–ABL1, perhaps the one of the best methods for interfering with the development of specific malignancies will be through inhibition of well-characterized, pathogenic fusion genes with rationally designed molecularly tailored therapies. In the future, the use of both COPA and high-throughput massively parallel sequencing will greatly increase the speed and reliability of fusion gene discovery on both the genomic and transcriptomic levels. We expect many more gene fusions to be reported over the next several years in various tumor types, many of which will hopefully serve as rational drug targets.

## Acknowledgements

We thank Jill Granger for critically reading the manuscript. A.M.C. is supported by a Burroughs Welcome Foundation Award in Clinical Translational Research. This work was supported in part by the National Institutes of Health Prostate SPORE P50CA69568 to A.M.C., and the Early Detection Research Network (UO1 CA113913 to A.M.C.) and the Department of Defense (PC051081 to A.M.C., BC083217 to J.C.B.).

## References

[1] D. Von Hansemann, Ueber asymmetrische zelltheilung in epithelhresbsen und deren biologische bedeutung. Virchows Arch., A Pathol. Anat. 119 (1890) 299–326.
[2] T. Boveri, Zur frage der entstehung maligner tumoren, Jena, Germany, 1914.
[3] J. Lejeune, M. Gautier, R. Turpin, [Study of somatic chromosomes from 9 mongoloid children.], C. R. Hebd. Seances Acad. Sci. 248 (1959) 1721–1722.
[4] J. Lejeune, R. Turpin, M. Gautier, [Mongolism; a chromosomal disease (trisomy)], Bull. Acad. Natl. Med. 143 (1959) 256–265.
[5] K.H. Rothfels, L. Siminovitch, An air-drying technique for flattening chromosomes in mammalian cells grown in vitro, Stain Technol. 33 (1958) 73–77.
[6] P. Nowell, D. Hungerford, A minute chromosome in human chronic granulocytic leukemia [Abstract], Science (New York, N.Y) 132 (1960) 1497.
[7] P.C. Nowell, D.A. Hungerford, Chromosome studies in human leukemia. II. Chronic granulocytic leukemia, J. Natl. Cancer Inst. 27 (1961) 1013–1035.
[8] I.M. Tough, W.M. Court Brown, A.G. Baikie, K.E. Buckton, D.G. Harnden, P.A. Jacobs, M.J. King, B.J. Mc, Cytogenetic studies in chronic myeloid leukaemia and acute leukaemia associated with monogolism, Lancet 1 (1961) 411–417.
[9] J.D. Rowley, Ph1-positive leukaemia, including chronic myelogenous leukaemia, Clin. Haematol. 9 (1980) 55–86.
[10] J. Groffen, J.R. Stephenson, N. Heisterkamp, A. de Klein, C.R. Bartram, G. Grosveld, Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22, Cell 36 (1984) 93–99.
[11] T.G. Lugo, A.M. Pendergast, A.J. Muller, O.N. Witte, Tyrosine kinase activity and transformation potency of bcr–abl oncogene products, Science (New York, N.Y) 247 (1990) 1079–1082.
[12] B.J. Druker, C.L. Sawyers, H. Kantarjian, D.J. Resta, S.F. Reese, J.M. Ford, R. Capdeville, M. Talpaz, Activity of a specific inhibitor of the BCR–ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome, N. Engl. J. Med. 344 (2001) 1038–1042.

[13] B.J. Druker, M. Talpaz, D.J. Resta, B. Peng, E. Buchdunger, J.M. Ford, N.B. Lydon, H. Kantarjian, R. Capdeville, S. Ohno-Jones, C.L. Sawyers, Efficacy and safety of a specific inhibitor of the BCR–ABL tyrosine kinase in chronic myeloid leukemia, N. Engl. J. Med. 344 (2001) 1031–1037.
[14] L. Zech, U. Haglund, K. Nilsson, G. Klein, Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas, Int. J. Cancer 17 (1976) 47–56.
[15] R. Dalla-Favera, M. Bregni, J. Erikson, D. Patterson, R.C. Gallo, C.M. Croce, Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 7824–7827.
[16] A. ar-Rushdi, K. Nishikura, J. Erikson, R. Watt, G. Rovera, C.M. Croce, Differential expression of the translocated and the untranslocated c-myc oncogene in Burkitt lymphoma, Science (New York, NY) 222 (1983) 390–393.
[17] J.M. Adams, A.W. Harris, C.A. Pinkert, L.M. Corcoran, W.S. Alexander, S. Cory, R.D. Palmiter, R.L. Brinster, The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice, Nature 318 (1985) 533–538.
[18] J.D. Rowley, Identification of a translocation with quinacrine fluorescence in a patient with acute leukemia, Ann. Genet. 16 (1973) 109–112.
[19] J. Gao, P. Erickson, K. Gardiner, M.M. Le Beau, M.O. Diaz, D. Patterson, J.D. Rowley, H.A. Drabkin, Isolation of a yeast artificial chromosome spanning the 8;21 translocation breakpoint t(8;21)(q22;q22.3) in acute myelogenous leukemia, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 4882–4886.
[20] G. Nucifora, J.D. Rowley, AML1 and the 8;21 and 3;21 translocations in acute and chronic myeloid leukemia, Blood 86 (1995) 1–14.
[21] D. Roulston, R. Espinosa III, G. Nucifora, R.A. Larson, M.M. Le Beau, J.D. Rowley, CBFA2(AML1) translocations with novel partner chromosomes in myeloid leukemias: association with prior therapy, Blood 92 (1998) 2879–2885.
[22] C. Meyer, B. Schneider, S. Jakob, S. Strehl, A. Attarbaschi, S. Schnittger, C. Schoch, M.W. Jansen, J.J. van Dongen, M.L. den Boer, R. Pieters, M.G. Ennas, E. Angelucci, U. Koehl, J. Greil, F. Griesinger, U. Zur Stadt, C. Eckert, T. Szczepanski, F.K. Niggli, B.W. Schafer, H. Kempski, H.J. Brady, J. Zuna, J. Trka, L.L. Nigro, A. Biondi, E. Delabesse, E. Macintyre, M. Stanulla, M. Schrappe, O.A. Haas, T. Burmeister, T. Dingermann, T. Klingebiel, R. Marschalek, The MLL recombinome of acute leukemias, Leukemia 20 (2006) 777–784.
[23] F. Mitelman, B. Johansson, F. Mertens, Mitelman Database of Chromosome Aberrations in Cancer, *http://cgap.nci.nih.gov/Chromosomes/Mitelman* (2008).
[24] D. Martin-Zanca, S.H. Hughes, M. Barbacid, A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences, Nature 319 (1986) 743–748.
[25] Y. Ishizaka, F. Itoh, T. Tahira, I. Ikeda, T. Sugimura, J. Tucker, A. Fertitta, A.V. Carrano, M. Nagao, Human ret proto-oncogene mapped to chromosome 10q11.2, Oncogene 4 (1989) 1519–1521.
[26] Y. Ishizaka, T. Tahira, M. Ochiai, I. Ikeda, T. Sugimura, M. Nagao, Molecular cloning and characterization of human ret-II oncogene, Oncogene Res. 3 (1988) 193–197.
[27] M. Takahashi, J. Ritz, G.M. Cooper, Activation of a novel human transforming gene, ret, by DNA rearrangement, Cell 42 (1985) 581–588.
[28] D.R. Kaplan, F.D. Miller, Signal transduction by the neurotrophin receptors, Curr. Opin. Cell Biol. 9 (1997) 213–221.
[29] A. Greco, L. Fusetti, C. Miranda, R. Villa, S. Zanotti, S. Pagliardini, M.A. Pierotti, Role of the TFG N-terminus and coiled-coil domain in the transforming activity of the thyroid TRK-T3 oncogene, Oncogene 16 (1998) 809–816.
[30] M. Grieco, M. Santoro, M.T. Berlingieri, R.M. Melillo, R. Donghi, I. Bongarzone, M.A. Pierotti, G. Della Porta, A. Fusco, G. Vecchio, PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas, Cell 60 (1990) 557–563.
[31] M.A. Pierotti, M. Santoro, R.B. Jenkins, G. Sozzi, I. Bongarzone, M. Grieco, N. Monzini, M. Miozzo, M.A. Herrmann, A. Fusco, et al., Characterization of an inversion on the long arm of chromosome 10 juxtaposing D10S170 and RET and creating the oncogenic sequence RET/PTC, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 1616–1620.
[32] G. Sozzi, I. Bongarzone, M. Miozzo, M.G. Borrello, M.G. Blutti, S. Pilotti, G. Della Porta, M.A. Pierotti, A t(10;17) translocation creates the RET/PTC2 chimeric transforming sequence in papillary thyroid carcinoma, Genes Chromosomes Cancer 9 (1994) 244–250.
[33] S. Klugbauer, E.P. Demidchik, E. Lengfelder, H.M. Rabes, Detection of a novel type of RET rearrangement (PTC5) in thyroid carcinomas after Chernobyl and analysis of the involved RET-fused gene *RFG5*, Cancer Res. 58 (1998) 198–203.
[34] L. Fugazzola, M.A. Pierotti, E. Vigano, F. Pacini, T.V. Vorontsova, I. Bongarzone, Molecular and biochemical analysis of RET/PTC4, a novel oncogenic rearrangement between *RET* and *ELE1* genes, in a post-Chernobyl papillary thyroid cancer, Oncogene 13 (1996) 1093–1097.
[35] M. Santoro, N.A. Dathan, M.T. Berlingieri, I. Bongarzone, C. Paulin, M. Grieco, M.A. Pierotti, G. Vecchio, A. Fusco, Molecular characterization of RET/PTC3; a novel rearranged version of the RETproto-oncogene in a human thyroid papillary carcinoma, Oncogene 9 (1994) 509–516.
[36] L. Alberti, C. Carniti, C. Miranda, E. Roccato, M.A. Pierotti, RET and NTRK1 proto-oncogenes in human diseases, J. Cell. Physiol. 195 (2003) 168–186.
[37] K. Hamatani, H. Eguchi, R. Ito, M. Mukai, K. Takahashi, M. Taga, K. Imai, J. Cologne, M. Soda, K. Arihiro, M. Fujihara, K. Abe, T. Hayashi, M. Nakashima, I. Sekine, W. Yasui, Y. Hayashi, K. Nakachi, RET/PTC rearrangements preferentially occurred in papillary thyroid cancer among atomic bomb survivors exposed to high radiation dose, Cancer Res. 68 (2008) 7176–7182.
[38] T. Fukushima, S. Suzuki, M. Mashiko, T. Ohtake, Y. Endo, Y. Takebayashi, K. Sekikawa, K. Hagiwara, S. Takenoshita, BRAF mutations in papillary carcinomas of the thyroid, Oncogene 22 (2003) 6455–6457.
[39] Y. Cohen, M. Xing, E. Mambo, Z. Guo, G. Wu, B. Trink, U. Beller, W.H. Westra, P.W. Ladenson, D. Sidransky, BRAF mutation in papillary thyroid carcinoma, J. Natl. Cancer Inst. 95 (2003) 625–627.
[40] M.N. Nikiforova, E.T. Kimura, M. Gandhi, P.W. Biddinger, J.A. Knauf, F. Basolo, Z. Zhu, R. Giannini, G. Salvatore, A. Fusco, M. Santoro, J.A. Fagin, Y.E. Nikiforov, BRAF mutations in thyroid tumors are restricted to papillary carcinomas and anaplastic or poorly differentiated carcinomas arising from papillary carcinomas, J. Clin. Endocrinol. Metab. 88 (2003) 5399–5404.
[41] H. Namba, M. Nakashima, T. Hayashi, N. Hayashida, S. Maeda, T.I. Rogounovitch, A. Ohtsuru, V.A. Saenko, T. Kanematsu, S. Yamashita, Clinical implication of hot spot BRAF mutation, V599E, in papillary thyroid cancers, J. Clin. Endocrinol. Metab. 88 (2003) 4393–4397.
[42] M. Frattini, C. Ferrario, P. Bressan, D. Balestra, L. De Cecco, P. Mondellini, I. Bongarzone, P. Collini, M. Gariboldi, S. Pilotti, M.A. Pierotti, A. Greco, Alternative mutations of BRAF, RET and NTRK1 are associated with similar but distinct gene expression patterns in papillary thyroid cancer, Oncogene 23 (2004) 7436–7440.
[43] C.T. Esapa, S.J. Johnson, P. Kendall-Taylor, T.W. Lennard, P.E. Harris, Prevalence of Ras mutations in thyroid neoplasia, Clin. Endocrinol. 50 (1999) 529–535.
[44] I. Bongarzone, L. Fugazzola, P. Vigneri, L. Mariani, P. Mondellini, F. Pacini, F. Basolo, A. Pinchera, S. Pilotti, M.A. Pierotti, Age-related activation of the tyrosine kinase receptor protooncogenes RET and NTRK1 in papillary thyroid carcinoma, J. Clin. Endocrinol. Metab. 81 (1996) 2006–2009.
[45] G. Tallini, S.L. Asa, RET oncogene activation in papillary thyroid carcinoma, Adv. Anat. Pathol. 8 (2001) 345–354.
[46] E.L. Chua, W.M. Wu, K.T. Tran, S.W. McCarthy, C.S. Lauer, D. Dubourdieu, N. Packham, C.J. O'Brien, J.R. Turtle, Q. Dong, Prevalence and distribution of ret/ptc 1, 2, and 3 in papillary thyroid carcinoma in New Caledonia and Australia, J. Clin. Endocrinol. Metab. 85 (2000) 2733–2739.
[47] D.L. Learoyd, M. Messina, J. Zedenius, A.I. Guinea, L.W. Delbridge, B.G. Robinson, RET/PTC and RET tyrosine kinase expression in adult papillary thyroid carcinomas, J. Clin. Endocrinol. Metab. 83 (1998) 3631–3635.
[48] K. Nakachi, T. Hayashi, K. Hamatani, H. Eguchi, Y. Kusunoki, Sixty years of follow-up of Hiroshima and Nagasaki survivors: current progress in molecular epidemiology studies, Mutat. Res. 659 (2008) 109–117.
[49] V.S. Kazakov, E.P. Demidchik, L.N. Astakhova, Thyroid cancer after Chernobyl, Nature 359 (1992) 21.
[50] L.N. Astakhova, L.R. Anspaugh, G.W. Beebe, A. Bouville, V.V. Drozdovitch, V. Garber, Y.I. Gavrilin, V.T. Khrouch, A.V. Kuvshinnikov, Y.N. Kuzmenkov, V.P. Minenko, K.V. Moschik, A.S. Nalivko, J. Robbins, E.V. Shemiakina, S. Shinkarev, S.I. Tochitskaya, M.A. Waclawiw, Chernobyl-related thyroid cancer in children of Belarus: a case-control study, Radiat. Res. 150 (1998) 349–356.
[51] M. Imaizumi, T. Usa, T. Tominaga, K. Neriishi, M. Akahoshi, E. Nakashima, K. Ashizawa, A. Hida, M. Soda, S. Fujiwara, M. Yamada, E. Ejima, N. Yokoyama, M. Okubo, K. Sugino, G. Suzuki, R. Maeda, S. Nagataki, K. Eguchi, Radiation dose–response relationships for thyroid nodules and autoimmune thyroid diseases in Hiroshima and Nagasaki atomic bomb survivors 55–58 years after radiation exposure, JAMA 295 (2006) 1011–1022.
[52] M.A. Pierotti, I. Bongarzone, M.G. Borello, A. Greco, S. Pilotti, G. Sozzi, Cytogenetics and molecular genetics of carcinomas arising from thyroid epithelial follicular cells, Genes Chromosomes Cancer 16 (1996) 1–14.
[53] J. Bullerdiek, G. Raabe, C. Boschen, S. Bartnitzke, Translocation (3;8;8)(p22 or p23;p23;q12) in a case of pleomorphic adenoma: similarity to a primary cytogenetic abnormality detected in an endometrial adenocarcinoma, Cancer Genet. Cytogenet. 27 (1987) 177–180.
[54] C. Waldron, Mixed tumor (pleomorphic adenoma and myoepithelioma), in: G.E.P.A.D. Gnepp (Ed.), Surgical Pathology of the Salivary Glands. W.B. Saunders Co., Philadelphia, 1991, pp. 165–184.
[55] J.P. Shah, J.K. Ihde, Salivary gland tumors, Curr. Probl. Surg. 27 (1990) 775–883.
[56] G. Stenman, Fusion oncogenes and tumor type specificity—insights from salivary gland tumors, Semin. Cancer Biol. 15 (2005) 224–235.
[57] J. Bullerdiek, J. Haubrich, K. Meyer, S. Bartnitzke, Translocation t(11;19)(q21; p13.1) as the sole chromosome abnormality in a cystadenolymphoma (Warthin's tumor) of the parotid gland, Cancer Genet. Cytogenet. 35 (1988) 129–132.
[58] J. Bullerdiek, G. Wobst, K. Meyer-Bolte, R. Chilla, J. Haubrich, B. Thode, S. Bartnitzke, Cytogenetic subtyping of 220 salivary gland pleomorphic adenomas: correlation to occurrence, histological subtype, and in vitro cellular behavior, Cancer Genet. Cytogenet. 65 (1993) 27–31.
[59] K. Kas, M.L. Voz, E. Roijer, A.K. Astrom, E. Meyen, G. Stenman, W.J. Van de Ven, Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations, Nat. Genet. 15 (1997) 170–174.
[60] W.J. Nelson, R. Nusse, Convergence of Wnt, beta-catenin, and cadherin pathways, Science (New York, N.Y) 303 (2004) 1483–1487.
[61] K. Hensen, I.C. Van Valckenborgh, K. Kas, W.J. Van de Ven, M.L. Voz, The tumorigenic diversity of the three PLAG family members is associated with different DNA binding capacities, Cancer Res. 62 (2002) 1510–1517.
[62] B. de Jong, I.M. Molenaar, J.A. Leeuw, V.J. Idenberg, J.W. Oosterhuis, Cytogenetics of a renal adenocarcinoma in a 2-year-old child, Cancer Genet. Cytogenet. 21 (1986) 165–169.
[63] A.M. Meloni, R.M. Dobbs, J.E. Pontes, A.A. Sandberg, Translocation (X;1) in papillary renal cell carcinoma. A new cytogenetic subtype, Cancer Genet. Cytogenet. 65 (1993) 1–6.
[64] A.M. Meloni, A.A. Sandberg, J.E. Pontes, R.M. Dobbs Jr., Translocation (X;1)(p11.2; q21). A subtype of renal adenocarcinomas, Cancer Genet. Cytogenet. 63 (1992) 100–101.

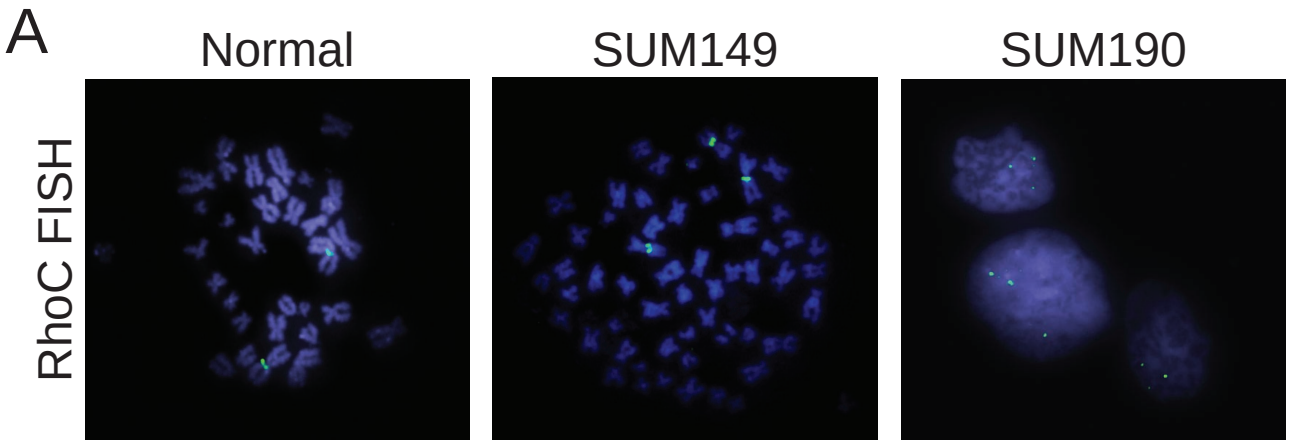
[65] J.M. Shipley, S. Birdsall, J. Clark, J. Crew, S. Gill, M. Linehan, J. Gnarra, S. Fisher, I.W. Craig, C.S. Cooper, Mapping the X chromosome breakpoint in two papillary renal cell carcinoma cell lines with a t(X;1)(p11.2;q21.2) and the first report of a female case, Cytogenet. Cell Genet. 71 (1995) 280–284.
[66] V. Tonk, K.S. Wilson, C.F. Timmons, N.R. Schneider, G.E. Tomlinson, Renal cell carcinoma with translocation (X;1). Further evidence for a cytogenetically defined subtype, Cancer Genet. Cytogenet. 81 (1995) 72–75.
[67] H. Beckmann, L.K. Su, T. Kadesch, TFE3: a helix–loop–helix protein that activates transcription through the immunoglobulin enhancer muE3 motif, Genes Dev. 4 (1990) 167–179.
[68] P.S. Henthorn, C.C. Stewart, T. Kadesch, J.M. Puck, The gene encoding human TFE3, a transcription factor that binds the immunoglobulin heavy-chain enhancer, maps to Xp11.22, Genomics 11 (1991) 374–378.
[69] S.K. Sidhar, J. Clark, S. Gill, R. Hamoudi, A.J. Crew, R. Gwilliam, M. Ross, W.M. Linehan, S. Birdsall, J. Shipley, C.S. Cooper, The t(X;1)(p11.2;q21.2) translocation in papillary renal cell carcinoma fuses a novel gene *PRCC* to the *TFE3* transcription factor gene, Hum. Mol. Genet. 5 (1996) 1333–1338.
[70] M. Sudol, From Src homology domains to other signaling modules: proposal of the 'protein recognition code', Oncogene 17 (1998) 1469–1474.
[71] R. Ren, B.J. Mayer, P. Cicchetti, D. Baltimore, Identification of a ten-amino acid proline-rich SH3 binding site, Science (New York, N.Y) 259 (1993) 1157–1161.
[72] M.A. Weterman, M. Wilbrink, A. Geurts van Kessel, Fusion of the transcription factor *TFE3* gene to a novel gene, *PRCC*, in t(X;1)(p11;q21)-positive papillary renal cell carcinomas, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 15294–15298.
[73] M.J. Weterman, J.J. van Groningen, A. Jansen, A.G. van Kessel, Nuclear localization and transactivating capacities of the papillary renal cell carcinoma-associated TFE3 and PRCC (fusion) proteins, Oncogene 19 (2000) 69–74.
[74] J. Clark, Y.J. Lu, S.K. Sidhar, C. Parker, S. Gill, D. Smedley, R. Hamoudi, W.M. Linehan, J. Shipley, C.S. Cooper, Fusion of splicing factor genes *PSF* and *NonO* (p54nrb) to the *TFE3* gene in papillary renal cell carcinoma, Oncogene 15 (1997) 2233–2239.
[75] Y.M. Skalsky, P.M. Ajuh, C. Parker, A.I. Lamond, G. Goodwin, C.S. Cooper, PRCC, the commonest TFE3 fusion partner in papillary renal carcinoma is associated with pre-mRNA splicing factors, Oncogene 20 (2001) 178–187.
[76] M.D. Odero, F.H. Grand, S. Iqbal, F. Ross, J.P. Roman, J.L. Vizmanos, J. Andrieux, J.L. Lai, M.J. Calasanz, N.C. Cross, Disruption and aberrant expression of HMGA2 as a consequence of diverse chromosomal translocations in myeloid malignancies, Leukemia 19 (2005) 245–252.
[77] E.F. Schoenmakers, S. Wanschura, R. Mols, J. Bullerdiek, H. Van den Berghe, W.J. Van de Ven, Recurrent rearrangements in the high mobility group protein gene, HMGI-C, in benign mesenchymal tumours, Nat. Genet. 10 (1995) 436–444.
[78] H.R. Ashar, M.S. Fejzo, A. Tkachenko, X. Zhou, J.A. Fletcher, S. Weremowicz, C.C. Morton, K. Chada, Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains, Cell 82 (1995) 57–65.
[79] A. Agresti, M.E. Bianchi, HMGB proteins and gene expression, Curr. Opin. Genet. Dev. 13 (2003) 170–178.
[80] J.M. Geurts, E.F. Schoenmakers, E. Roijer, G. Stenman, W.J. Van de Ven, Expression of reciprocal hybrid transcripts of HMGIC and FHIT in a pleomorphic adenoma of the parotid gland, Cancer Res. 57 (1997) 13–17.
[81] J.M. Geurts, E.F. Schoenmakers, E. Roijer, A.K. Astrom, G. Stenman, W.J. van de Ven, Identification of NFIB as recurrent translocation partner gene of HMGIC in pleomorphic adenomas, Oncogene 16 (1998) 865–872.
[82] M. Fedele, S. Battista, L. Kenyon, G. Baldassarre, V. Fidanza, A.J. Klein-Szanto, A.F. Parlow, R. Visone, G.M. Pierantoni, E. Outwater, M. Santoro, C.M. Croce, A. Fusco, Overexpression of the *HMGA2* gene in transgenic mice leads to the onset of pituitary adenomas, Oncogene 21 (2002) 3190–3198.
[83] S. Battista, V. Fidanza, M. Fedele, A.J. Klein-Szanto, E. Outwater, H. Brunner, M. Santoro, C.M. Croce, A. Fusco, The expression of a truncated *HMGI-C* gene induces gigantism associated with lipomatosis, Cancer Res. 59 (1999) 4793–4797.
[84] G. Baldassarre, M. Fedele, S. Battista, A. Vecchione, A.J. Klein-Szanto, M. Santoro, T.A. Waldmann, N. Azimi, C.M. Croce, A. Fusco, Onset of natural killer cell lymphomas in transgenic mice carrying a truncated *HMGI-C* gene by the chronic stimulation of the IL-2 and IL-15 pathway, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 7970–7975.
[85] M. Ohta, H. Inoue, M.G. Cotticelli, K. Kastury, R. Baffa, J. Palazzo, Z. Siprashvili, M. Mori, P. McCue, T. Druck, C.M. Croce, K. Huebner, The *FHIT* gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers, Cell 84 (1996) 587–597.
[86] R.M. Gemmill, J.D. West, F. Boldog, N. Tanaka, L.J. Robinson, D.I. Smith, F. Li, H.A. Drabkin, The hereditary renal cell carcinoma 3;8 translocation fuses *FHIT* to a patched-related gene, *TRC8*, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 9572–9577.
[87] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.
[88] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (2005) 15–20.
[89] C. Mayr, M.T. Hemann, D.P. Bartel, Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation, Science (New York, N.Y) 315 (2007) 1576–1579.
[90] T.G. Kroll, P. Sarraf, L. Pecciarini, C.J. Chen, E. Mueller, B.M. Spiegelman, J.A. Fletcher, PAX8–PPARgamma1 fusion oncogene in human thyroid carcinoma [corrected], Science (New York, N.Y) 289 (2000) 1357–1360.
[91] O. Gimm, Thyroid cancer, Cancer Lett. 163 (2001) 143–156.
[92] M.N. Nikiforova, Y.E. Nikiforov, Molecular genetics of thyroid cancer: implications for diagnosis, treatment and prognosis, Expert Rev. Mol. Diagn. 8 (2008) 83–95.
[93] C. Puppin, I. Presta, A.V. D'Elia, G. Tell, F. Arturi, D. Russo, S. Filetti, G. Damante, Functional interaction among thyroid-specific transcription factors: Pax8 regulates the activity of Hex promoter, Mol. Cell. Endocrinol. 214 (2004) 117–125.
[94] P. Tontonoz, B.M. Spiegelman, Fat and beyond: the diverse biology of PPARgamma, Ann. Rev. Biochem. 77 (2008) 289–312.
[95] W.O. Lui, T. Foukakis, J. Liden, S.R. Thoppe, T. Dwight, A. Hoog, J. Zedenius, G. Wallin, M. Reimers, C. Larsson, Expression profiling reveals a distinct transcription signature in follicular thyroid carcinomas with a PAX8–PPAR(gamma) fusion oncogene, Oncogene 24 (2005) 1467–1476.
[96] M.N. Nikiforova, R.A. Lynch, P.W. Biddinger, E.K. Alexander, G.W. Dorn II, G. Tallini, T.G. Kroll, Y.E. Nikiforov, RAS point mutations and PAX8–PPAR gamma rearrangement in thyroid tumors: evidence for distinct molecular pathways in thyroid follicular carcinoma, J. Clin. Endocrinol. Metab. 88 (2003) 2318–2326.
[97] T. Foukakis, A.Y. Au, G. Wallin, J. Geli, L. Forsberg, R. Clifton-Bligh, B.G. Robinson, W.O. Lui, J. Zedenius, C. Larsson, The Ras effector NORE1A is suppressed in follicular thyroid carcinomas with a PAX8–PPARgamma fusion, J. Clin. Endocrinol. Metab. 91 (2006) 1143–1149.
[98] C.A. French, I. Miyoshi, J.C. Aster, I. Kubonishi, T.G. Kroll, P. Dal Cin, S.O. Vargas, A.R. Perez-Atayde, J.A. Fletcher, BRD4 bromodomain gene rearrangement in aggressive carcinoma with translocation t(15;19), Am. J. Pathol. 159 (2001) 1987–1992.
[99] C.A. French, I. Miyoshi, I. Kubonishi, H.E. Grier, A.R. Perez-Atayde, J.A. Fletcher, BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma, Cancer Res. 63 (2003) 304–307.
[100] C.A. French, Molecular pathology of NUT midline carcinomas, J. Clin. Pathol. (2008).
[101] A. Dey, J. Ellenberg, A. Farina, A.E. Coleman, T. Maruyama, S. Sciortino, J. Lippincott-Schwartz, K. Ozato, A bromodomain protein, MCAP, associates with mitotic chromosomes and affects G(2)-to-M transition, Mol. Cell. Biol. 20 (2000) 6537–6549.
[102] T. Maruyama, A. Farina, A. Dey, J. Cheong, V.P. Bermudez, T. Tamura, S. Sciortino, J. Shuman, J. Hurwitz, K. Ozato, A mammalian bromodomain protein, brd4, interacts with replication factor C and inhibits progression to S phase, Mol. Cell. Biol. 22 (2002) 6509–6520.
[103] A. Farina, M. Hattori, J. Qin, Y. Nakatani, N. Minato, K. Ozato, Bromodomain protein Brd4 binds to GTPase-activating SPA-1, modulating its activity and subcellular localization, Mol. Cell. Biol. 24 (2004) 9059–9069.
[104] A. Dey, F. Chitsaz, A. Abbasi, T. Misteli, K. Ozato, The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 8758–8763.
[105] M.K. Jang, K. Mochizuki, M. Zhou, H.S. Jeong, J.N. Brady, K. Ozato, The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription, Mol. Cell 19 (2005) 523–534.
[106] C.A. French, C.L. Ramirez, J. Kolmakova, T.T. Hickman, M.J. Cameron, M.E. Thyne, J.L. Kutok, J.A. Toretsky, A.K. Tadavarthy, U.R. Kees, J.A. Fletcher, J.C. Aster, BRD-NUT oncoproteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells, Oncogene 27 (2008) 2237–2242.
[107] H.A. Oberman, Secretory carcinoma of the breast in adults, Am. J. Surg. Pathol. 4 (1980) 465–470.
[108] C. Tognon, S.R. Knezevich, D. Huntsman, C.D. Roskelley, N. Melnyk, J.A. Mathers, L. Becker, F. Carneiro, N. MacPherson, D. Horsman, C. Poremba, P.H. Sorensen, Expression of the *ETV6–NTRK3* gene fusion as a primary event in human secretory breast carcinoma, Cancer Cell 2 (2002) 367–376.
[109] T.R. Golub, G.F. Barker, S.K. Bohlander, S.W. Hiebert, D.C. Ward, P. Bray-Ward, E. Morgan, S.C. Raimondi, J.D. Rowley, D.G. Gilliland, Fusion of the *TEL* gene on 12p13 to the *AML1* gene on 21q22 in acute lymphoblastic leukemia, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 4917–4921.
[110] P. Papadopoulos, S.A. Ridge, C.A. Boucher, C. Stocking, L.M. Wiedemann, The novel activation of *ABL* by fusion to an *ets*-related gene, *TEL*, Cancer Res. 55 (1995) 34–38.
[111] T.R. Golub, A. Goga, G.F. Barker, D.E. Afar, J. McLaughlin, S.K. Bohlander, J.D. Rowley, O.N. Witte, D.G. Gilliland, Oligomerization of the ABL tyrosine kinase by the Ets protein TEL in human leukemia, Mol. Cell. Biol. 16 (1996) 4107–4116.
[112] V. Lacronique, A. Boureux, V.D. Valle, H. Poirel, C.T. Quang, M. Mauchauffe, C. Berthou, M. Lessard, R. Berger, J. Ghysdael, O.A. Bernard, A TEL–JAK2 fusion protein with constitutive kinase activity in human leukemia, Science (New York, NY) 278 (1997) 1309–1312.
[113] P. Peeters, S.D. Raynaud, J. Cools, I. Wlodarska, J. Grosgeorge, P. Philip, F. Monpoux, L. Van Rompaey, M. Baens, H. Van den Berghe, P. Marynen, Fusion of TEL, the ETS-variant gene 6 (*ETV6*), to the receptor-associated kinase JAK2 as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia, Blood 90 (1997) 2535–2540.
[114] J. Schwaller, J. Frantsve, J. Aster, I.R. Williams, M.H. Tomasson, T.S. Ross, P. Peeters, L. Van Rompaey, R.A. Van Etten, R. Ilaria Jr., P. Marynen, D.G. Gilliland, Transformation of hematopoietic cell lines to growth-factor independence and induction of a fatal myelo- and lymphoproliferative disease in mice by retrovirally transduced *TEL/JAK2* fusion genes, EMBO J. 17 (1998) 5321–5333.
[115] G. Cazzaniga, S. Tosi, A. Aloisi, G. Giudici, M. Daniotti, P. Pioltelli, L. Kearney, A. Biondi, The tyrosine kinase abl-related gene *ARG* is fused to *ETV6* in an AML-M4Eo patient with a t(1;12)(q25;p13): molecular cloning of both reciprocal transcripts, Blood 94 (1999) 4370–4373.

[116] Y. Iijima, T. Ito, T. Oikawa, M. Eguchi, M. Eguchi-Ishimae, N. Kamada, K. Kishi, S. Asano, Y. Sakaki, Y. Sato, A new *ETV6/TEL* partner gene, *ARG* (*ABL*-related gene or *ABL2*), identified in an AML-M3 cell line with a t(1;12)(q25;p13) translocation, Blood 95 (2000) 2126–2131.
[117] T.R. Golub, G.F. Barker, M. Lovett, D.G. Gilliland, Fusion of PDGF receptor beta to a novel *ets*-like gene, *tel*, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation, Cell 77 (1994) 307–316.
[118] F. Yagasaki, D. Wakao, Y. Yokoyama, Y. Uchida, I. Murohashi, H. Kayano, M. Taniwaki, A. Matsuda, M. Bessho, Fusion of ETV6 to fibroblast growth factor receptor 3 in peripheral T-cell lymphoma with a *t*(4;12)(p16;p13) chromosomal translocation, Cancer Res. 61 (2001) 8371–8374.
[119] S.K. Bohlander, ETV6: a versatile player in leukemogenesis, Semin. Cancer Biol. 15 (2005) 162–174.
[120] L.C. Wang, W. Swat, Y. Fujiwara, L. Davidson, J. Visvader, F. Kuo, F.W. Alt, D.G. Gilliland, T.R. Golub, S.H. Orkin, The *TEL/ETV6* gene is required specifically for hematopoiesis in the bone marrow, Genes Dev. 12 (1998) 2392–2402.
[121] D.R. Kaplan, F.D. Miller, Neurotrophin signal transduction in the nervous system, Curr. Opin. Neurobiol. 10 (2000) 381–391.
[122] S.R. Knezevich, D.E. McFadden, W. Tao, J.F. Lim, P.H. Sorensen, A novel *ETV6–NTRK3* gene fusion in congenital fibrosarcoma, Nat. Genet. 18 (1998) 184–187.
[123] C.L. Lannon, P.H. Sorensen, ETV6–NTRK3: a chimeric protein tyrosine kinase with transformation activity in multiple cell lineages, Semin. Cancer Biol. 15 (2005) 215–223.
[124] W. Jin, B.C. Kim, C. Tognon, H.J. Lee, S. Patel, C.L. Lannon, J.M. Maris, T.J. Triche, P.H. Sorensen, S.J. Kim, The ETV6–NTRK3 chimeric tyrosine kinase suppresses TGF-beta signaling by inactivating the TGF-beta type II receptor, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 16239–16244.
[125] B.N. Davis, A.C. Hilyard, G. Lagna, A. Hata, SMAD proteins control DROSHA-mediated microRNA maturation, Nature (2008).
[126] H. Herz, B. Cooke, D. Goldstein, Metastatic secretory breast cancer. Non-responsiveness to chemotherapy: case report and review of the literature, Ann. Oncol. 11 (2000) 1343–1347.
[127] E. de Bree, J. Askoxylakis, E. Giannikaki, N. Chroniaris, E. Sanidas, D.D. Tsiftsis, Secretory carcinoma of the male breast, Ann. Surg. Oncol. 9 (2002) 663–667.
[128] W. Jin, C. Yun, A. Hobbie, M.J. Martin, P.H. Sorensen, S.J. Kim, Cellular transformation and activation of the phosphoinositide-3-kinase-Akt cascade by the ETV6–NTRK3 chimeric tyrosine kinase requires c-Src, Cancer Res. 67 (2007) 3192–3200.
[129] Z. Li, C.E. Tognon, F.J. Godinho, L. Yasaitis, H. Hock, J.I. Herschkowitz, C.L. Lannon, E. Cho, S.J. Kim, R.T. Bronson, C.M. Perou, P.H. Sorensen, S.H. Orkin, ETV6–NTRK3 fusion oncogene initiates breast cancer from committed mammary progenitors via activation of AP1 complex, Cancer Cell 12 (2007) 542–558.
[130] C. Kumar-Sinha, S.A. Tomlins, A.M. Chinnaiyan, Recurrent gene fusions in prostate cancer, Nat. Rev. 8 (2008) 497–511.
[131] B. Lin, C. Ferguson, J.T. White, S. Wang, R. Vessella, L.D. True, L. Hood, P.S. Nelson, Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2, Cancer Res. 59 (1999) 4180–4184.
[132] D.E. Afar, I. Vivanco, R.S. Hubert, J. Kuo, E. Chen, D.C. Saffran, A.B. Raitano, A. Jakobovits, Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia, Cancer Res. 61 (2001) 1686–1692.
[133] M.H. Vaarala, K. Porvari, A. Kyllonen, O. Lukkarinen, P. Vihko, The *TMPRSS2* gene encoding transmembrane serine protease is overexpressed in a majority of prostate cancer patients: detection of mutated TMPRSS2 form in a case of aggressive disease, Int. J. Cancer 94 (2001) 705–710.
[134] J. Lapointe, Y.H. Kim, M.A. Miller, C. Li, G. Kaygusuz, M. van de Rijn, D.G. Huntsman, J.D. Brooks, J.R. Pollack, A variant TMPRSS2 isoform and ERG fusion product in prostate cancer with implications for molecular diagnosis, Mod. Path. 20 (2007) 467–473.
[135] S.A. Tomlins, D.R. Rhodes, S. Perner, S.M. Dhanasekaran, R. Mehra, X.W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kuefer, C. Lee, J.E. Montie, R.B. Shah, K.J. Pienta, M.A. Rubin, A.M. Chinnaiyan, Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer, Science (New York, N.Y) 310 (2005) 644–648.
[136] N. Cerveira, F.R. Ribeiro, A. Peixoto, V. Costa, R. Henrique, C. Jeronimo, M.R. Teixeira, *TMPRSS2–ERG* gene fusion causing ERG overexpression precedes chromosome copy number changes in prostate carcinomas and paired HGPIN lesions, Neoplasia (New York, N.Y) 8 (2006) 826–832.
[137] J. Clark, S. Merson, S. Jhavar, P. Flohr, S. Edwards, C.S. Foster, R. Eeles, F.L. Martin, D.H. Phillips, M. Crundwell, T. Christmas, A. Thompson, C. Fisher, G. Kovacs, C.S. Cooper, Diversity of TMPRSS2–ERG fusion transcripts in the human prostate, Oncogene 26 (2007) 2667–2673.
[138] K.G. Hermans, R. van Marion, H. van Dekken, G. Jenster, W.M. van Weerden, J. Trapman, TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer, Cancer Res. 66 (2006) 10658–10663.
[139] K. Iljin, M. Wolf, H. Edgren, S. Gupta, S. Kilpinen, R.I. Skotheim, M. Peltola, F. Smit, G. Verhaegh, J. Schalken, M. Nees, O. Kallioniemi, TMPRSS2 fusions with oncogenic ETS factors in prostate cancer involve unbalanced genomic rearrangements and are associated with HDAC1 and epigenetic reprogramming, Cancer Res. 66 (2006) 10242–10246.
[140] A.B. Rajput, M.A. Miller, A. De Luca, N. Boyd, S. Leung, A. Hurtado-Coll, L. Fazli, E.C. Jones, J.B. Palmer, M.E. Gleave, M.E. Cox, D.G. Huntsman, Frequency of the *TMPRSS2:ERG* gene fusion is increased in moderate to poorly differentiated prostate cancers, J. Clin. Pathol. 60 (2007) 1238–1243.
[141] M.J. Soller, M. Isaksson, P. Elfving, W. Soller, R. Lundgren, I. Panagopoulos, Confirmation of the high frequency of the *TMPRSS2/ERG* fusion gene in prostate cancer, Genes Chromosomes Cancer 45 (2006) 717–719.
[142] M. Yoshimoto, A.M. Joshua, S. Chilton-Macneill, J. Bayani, S. Selvarajah, A.J. Evans, M. Zielenska, J.A. Squire, Three-color FISH analysis of TMPRSS2/ERG fusions in prostate cancer indicates that genomic microdeletion of chromosome 21 is associated with rearrangement, Neoplasia (New York, N.Y) 8 (2006) 465–469.
[143] J. Wang, Y. Cai, C. Ren, M. Ittmann, Expression of variant TMPRSS2/ERG fusion messenger RNAs is associated with aggressive prostate cancer, Cancer Res. 66 (2006) 8347–8351.
[144] K.D. Mertz, S.R. Setlur, S.M. Dhanasekaran, F. Demichelis, S. Perner, S. Tomlins, J. Tchinda, B. Laxman, R.L. Vessella, R. Beroukhim, C. Lee, A.M. Chinnaiyan, M.A. Rubin, Molecular characterization of *TMPRSS2–ERG* gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model, Neoplasia (New York, NY) 9 (2007) 200–206.
[145] S.A. Tomlins, B. Laxman, S. Varambally, X. Cao, J. Yu, B.E. Helgeson, Q. Cao, J.R. Prensner, M.A. Rubin, R.B. Shah, R. Mehra, A.M. Chinnaiyan, Role of the *TMPRSS2-ERG* gene fusion in prostate cancer, Neoplasia (New York, NY) 10 (2008) 177–188.
[146] C. Sun, A. Dobi, A. Mohamed, H. Li, R.L. Thangapazham, B. Furusato, S. Shaheduzzaman, S.H. Tan, G. Vaidyanathan, E. Whitman, D.J. Hawksworth, Y. Chen, M. Nau, V. Patel, M. Vahey, J.S. Gutkind, T. Sreenath, G. Petrovics, I.A. Sesterhenn, D.G. McLeod, S. Srivastava, TMPRSS2–ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation, Oncogene 27 (2008) 5348–5353.
[147] S.A. Tomlins, B. Laxman, S.M. Dhanasekaran, B.E. Helgeson, X. Cao, D.S. Morris, A. Menon, X. Jing, Q. Cao, B. Han, J. Yu, L. Wang, J.E. Montie, M.A. Rubin, K.J. Pienta, D. Roulston, R.B. Shah, S. Varambally, R. Mehra, A.M. Chinnaiyan, Distinct classes of chromosomal rearrangements create oncogenic *ETS* gene fusions in prostate cancer, Nature 448 (2007) 595–599.
[148] D.R. Shaffer, P.P. Pandolfi, Breaking the rules of cancer, Nat. Med. 12 (2006) 14–15.
[149] S.A. Tomlins, R. Mehra, D.R. Rhodes, L.R. Smith, D. Roulston, B.E. Helgeson, X. Cao, J.T. Wei, M.A. Rubin, R.B. Shah, A.M. Chinnaiyan, *TMPRSS2:ETV4* gene fusions define a third molecular subtype of prostate cancer, Cancer Res. 66 (2006) 3396–3400.
[150] B. Han, R. Mehra, S.M. Dhanasekaran, J. Yu, A. Menon, R.J. Lonigro, X. Wang, Y. Gong, L. Wang, S. Shankar, B. Laxman, R.B. Shah, S. Varambally, N. Palanisamy, S.A. Tomlins, C. Kumar-Sinha, A.M. Chinnaiyan, A fluorescence in situ hybridization screen for E26 transformation-specific aberrations: identification of DDX5–ETV4 fusion protein in prostate cancer, Cancer Res. 68 (2008) 7629–7637.
[151] B.E. Helgeson, S.A. Tomlins, N. Shah, B. Laxman, Q. Cao, J.R. Prensner, X. Cao, N. Singla, J.E. Montie, S. Varambally, R. Mehra, A.M. Chinnaiyan, Characterization of *TMPRSS2:ETV5* and *SLC45A3:ETV5* gene fusions in prostate cancer, Cancer Res. 68 (2008) 73–80.
[152] K.G. Hermans, A.A. Bressers, H.A. van der Korput, N.F. Dits, G. Jenster, J. Trapman, Two unique novel prostate-specific and androgen-regulated fusion partners of ETV4 in prostate cancer, Cancer Res. 68 (2008) 3094–3098.
[153] R. Mehra, S.A. Tomlins, J. Yu, X. Cao, L. Wang, A. Menon, M.A. Rubin, K.J. Pienta, R.B. Shah, A.M. Chinnaiyan, Characterization of *TMPRSS2–ETS* gene aberrations in androgen-independent metastatic prostate cancer, Cancer Res. 68 (2008) 3584–3590.
[154] C. Cai, C.L. Hsieh, J. Omwancha, Z. Zheng, S.Y. Chen, J.L. Baert, L. Shemshedini, ETV1 is a novel androgen receptor-regulated gene that mediates prostate cancer cell invasion, Mol. Endocrinol. (Baltimore, Md) 21 (2007) 1835–1846.
[155] M. Soda, Y.L. Choi, M. Enomoto, S. Takada, Y. Yamashita, S. Ishikawa, S. Fujiwara, H. Watanabe, K. Kurashina, H. Hatanaka, M. Bando, S. Ohno, Y. Ishikawa, H. Aburatani, T. Niki, Y. Sohara, Y. Sugiyama, H. Mano, Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer, Nature 448 (2007) 561–566.
[156] M. Pollmann, R. Parwaresch, S. Adam-Klages, M.L. Kruse, F. Buck, H.J. Heidebrecht, Human EML4, a novel member of the EMAP family, is essential for microtubule formation, Exp. Cell Res. 312 (2006) 3241–3251.
[157] Y.L. Choi, K. Takeuchi, M. Soda, K. Inamura, Y. Togashi, S. Hatano, M. Enomoto, T. Hamada, H. Haruta, H. Watanabe, K. Kurashina, H. Hatanaka, T. Ueno, S. Takada, Y. Yamashita, Y. Sugiyama, Y. Ishikawa, H. Mano, Identification of novel isoforms of the EML4–ALK transforming gene in non-small cell lung cancer, Cancer Res. 68 (2008) 4971–4976.
[158] K. Takeuchi, Y.L. Choi, M. Soda, K. Inamura, Y. Togashi, S. Hatano, M. Enomoto, S. Takada, Y. Yamashita, Y. Satoh, S. Okumura, K. Nakagawa, Y. Ishikawa, H. Mano, Multiplex reverse transcription-PCR screening for EML4–ALK fusion transcripts, Clin. Cancer Res. 14 (2008) 6618–6624.
[159] Y. Fukuyoshi, H. Inoue, Y. Kita, T. Utsunomiya, T. Ishida, M. Mori, EML4–ALK fusion transcript is not found in gastrointestinal and breast cancers, Br. J. Cancer 98 (2008) 1536–1539.
[160] K. Inamura, K. Takeuchi, Y. Togashi, K. Nomura, H. Ninomiya, M. Okui, Y. Satoh, S. Okumura, K. Nakagawa, M. Soda, Y.L. Choi, T. Niki, H. Mano, Y. Ishikawa, EML4–ALK fusion is linked to histological characteristics in a subset of lung cancers, J. Thorac. Oncol. 3 (2008) 13–17.
[161] S.W. Morris, M.N. Kirstein, M.B. Valentine, K.G. Dittmer, D.N. Shapiro, D.L. Saltman, A.T. Look, Fusion of a kinase gene, *ALK*, to a nucleolar protein gene, *NPM*, in non-Hodgkin's lymphoma, Science (New York, N.Y) 263 (1994) 1281–1284.
[162] L. Lamant, N. Dastugue, K. Pulford, G. Delsol, B. Mariame, A new fusion gene *TPM3–ALK* in anaplastic large cell lymphoma created by a (1;2)(q25;p23) translocation, Blood 93 (1999) 3088–3095.
[163] C. Touriol, C. Greenland, L. Lamant, K. Pulford, F. Bernard, T. Rousset, D.Y. Mason, G. Delsol, Further demonstration of the diversity of chromosomal changes

involving 2p23 in ALK-positive lymphoma: 2 cases expressing ALK kinase fused to CLTCL (clathrin chain polypeptide-like), Blood 95 (2000) 3204–3207.
[164] G.W. Colleoni, J.A. Bridge, B. Garicochea, J. Liu, D.A. Filippa, M. Ladanyi, ATIC–ALK: A novel variant *ALK* gene fusion in anaplastic large cell lymphoma resulting from the recurrent cryptic chromosomal inversion, inv(2)(p23q35), Am. J. Pathol. 156 (2000) 781–789.
[165] Z. Ma, J. Cools, P. Marynen, X. Cui, R. Siebert, S. Gesk, B. Schlegelberger, B. Peeters, C. De Wolf-Peeters, I. Wlodarska, S.W. Morris, Inv(2)(p23q35) in anaplastic large-cell lymphoma induces constitutive anaplastic lymphoma kinase (ALK) tyrosine kinase activation by fusion to ATIC, an enzyme involved in purine nucleotide biosynthesis, Blood 95 (2000) 2144–2149.
[166] M. Trinei, L. Lanfrancone, E. Campo, K. Pulford, D.Y. Mason, P.G. Pelicci, B. Falini, A new variant anaplastic lymphoma kinase (ALK)-fusion protein (ATIC–ALK) in a case of ALK-positive anaplastic large cell lymphoma, Cancer Res. 60 (2000) 793–798.
[167] L. Hernandez, M. Pinyol, S. Hernandez, S. Bea, K. Pulford, A. Rosenwald, L. Lamant, B. Falini, G. Ott, D.Y. Mason, G. Delsol, E. Campo, TRK-fused gene (*TFG*) is a new partner of *ALK* in anaplastic large cell lymphoma producing two structurally different *TFG–ALK* translocations, Blood 94 (1999) 3265–3268.
[168] K. Shinmura, S. Kageyama, H. Tao, T. Bunai, M. Suzuki, T. Kamo, K. Takamochi, K. Suzuki, M. Tanahashi, H. Niwa, H. Ogawa, H. Sugimura, EML4–ALK fusion transcripts, but no NPM-, TPM3-, CLTC-, ATIC-, or TFG–ALK fusion transcripts, in non-small cell lung carcinomas, Lung Cancer (Amsterdam, Netherlands) (2008).
[169] S. Perner, P.L. Wagner, F. Demichelis, R. Mehra, C.J. Lafargue, B.J. Moss, S. Arbogast, A. Soltermann, W. Weder, T.J. Giordano, D.G. Beer, D.S. Rickman, A.M. Chinnaiyan, H. Moch, M.A. Rubin, EML4–ALK fusion lung cancer: a rare acquired event, Neoplasia (New York, N.Y) 10 (2008) 298–302.
[170] K. Rikova, A. Guo, Q. Zeng, A. Possemato, J. Yu, H. Haack, J. Nardone, K. Lee, C. Reeves, Y. Li, Y. Hu, Z. Tan, M. Stokes, L. Sullivan, J. Mitchell, R. Wetzel, J. Macneill, J.M. Ren, J. Yuan, C.E. Bakalarski, J. Villen, J.M. Kornhauser, B. Smith, D. Li, X. Zhou, S.P. Gygi, T.L. Gu, R.D. Polakiewicz, J. Rush, M.J. Comb, Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer, Cell 131 (2007) 1190–1203.
[171] C.A. Maher, C. Kumar-Sinha, X. Cao, S. Kalyana-Sundaram, B. Han, X. Jing, L. Sam, T. Barrette, N. Palanisamy, A.M. Chinnaiyan, Transcriptome sequencing to detect gene fusions in cancer, Nature (2009).
[172] P.M. Ayton, M.L. Cleary, Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins, Oncogene 20 (2001) 5695–5707.
[173] E.C. Collins, T.H. Rabbitts, The promiscuous *MLL* gene links chromosomal translocations to cellular differentiation and tumour tropism, Trends Mol. Med. 8 (2002) 436–442.
[174] A. Daser, T.H. Rabbitts, Extending the repertoire of the mixed-lineage leukemia gene *MLL* in leukemogenesis, Genes Dev. 18 (2004) 965–974.
[175] A. Schumacher, T. Magnuson, Murine polycomb- and trithorax-group genes regulate homeotic pathways and beyond, Trends Genet. 13 (1997) 167–170.
[176] M. Maconochie, S. Nonchev, A. Morrison, R. Krumlauf, Paralogous *Hox* genes: function and regulation, Annu. Rev. Genet. 30 (1996) 529–556.
[177] A.T. Look, Oncogenic transcription factors in the human acute leukemias, Science (New York, NY) 278 (1997) 1059–1064.
[178] P.M. Ayton, M.L. Cleary, Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9, Genes Dev. 17 (2003) 2298–2307.
[179] K. Hsu, A.T. Look, Turning on a dimer: new insights into MLL chimeras, Cancer Cell 4 (2003) 81–83.
[180] N.J. Zeleznik-Le, A.M. Harden, J.D. Rowley, 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (*MLL*) gene, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 10610–10614.
[181] T. Nakamura, T. Mori, S. Tada, W. Krajewski, T. Rozovskaia, R. Wassell, G. Dubois, A. Mazo, C.M. Croce, E. Canaani, ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation, Mol. Cell 10 (2002) 1119–1128.
[182] C.W. So, M. Lin, P.M. Ayton, E.H. Chen, M.L. Cleary, Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias, Cancer Cell 4 (2003) 99–110.
[183] T.A. Milne, M.E. Martin, H.W. Brock, R.K. Slany, J.L. Hess, Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications, Cancer Res. 65 (2005) 11367–11374.
[184] A.M. Dorrance, S. Liu, W. Yuan, B. Becknell, K.J. Arnoczky, M. Guimond, M.P. Strout, L. Feng, T. Nakamura, L. Yu, L.J. Rush, M. Weinstein, G. Leone, L. Wu, A. Ferketich, S.P. Whitman, G. Marcucci, M.A. Caligiuri, Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations, J. Clin. Invest. 116 (2006) 2707–2716.
[185] A.A. Sandberg, The chromosomes and causation of human cancer and leukemia, Cancer Res. 26 (1966) 2064–2081.
[186] J. DeGrouchy, Chromosomes in neoplastic tissues, in: J.C.J. Neel (Ed.), Proceedings of the Third International Congress of Human Genetics, Johns Hopkins Univ. Press, Chicago, 1967.
[187] A. Aguilera, B. Gomez-Gonzalez, Genome instability: a mechanistic view of its causes and consequences, Nat. Rev., Genet. 9 (2008) 204–217.
[188] T.D. Halazonetis, V.G. Gorgoulis, J. Bartek, An oncogene-induced DNA damage model for cancer development, Science (New York, N.Y) 319 (2008) 1352–1355.
[189] N. Zhang, G. Ge, R. Meyer, S. Sethi, D. Basu, S. Pradhan, Y.J. Zhao, X.N. Li, W.W. Cai, A.K. El-Naggar, V. Baladandayuthapani, F.S. Kittrell, P.H. Rao, D. Medina, D. Pati, Overexpression of Separase induces aneuploidy and mammary tumorigenesis, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 13033–13038.
[190] R. Sotillo, E. Hernando, E. Diaz-Rodriguez, J. Teruya-Feldstein, C. Cordon-Cardo, S.W. Lowe, R. Benezra, Mad2 overexpression promotes aneuploidy and tumorigenesis in mice, Cancer Cell 11 (2007) 9–23.
[191] L. Gorunova, M. Hoglund, A. Andren-Sandberg, S. Dawiskiba, Y. Jin, F. Mitelman, B. Johansson, Cytogenetic analysis of pancreatic carcinomas: intratumor heterogeneity and nonrandom pattern of chromosome aberrations, Genes Chromosomes Cancer 23 (1998) 81–99.
[192] F. Mitelman, B. Johansson, F. Mertens, Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer, Nat. Genet. 36 (2004) 331–334.
[193] D.R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, A.M. Chinnaiyan, ONCOMINE: a cancer microarray database and integrated data-mining platform, Neoplasia (New York, NY) 6 (2004) 1–6.
[194] D.R. Rhodes, S. Kalyana-Sundaram, V. Mahavisno, T.R. Barrette, D. Ghosh, A.M. Chinnaiyan, Mining for regulatory programs in the cancer transcriptome, Nat. Genet. 37 (2005) 579–583.
[195] D.R. Rhodes, S. Kalyana-Sundaram, V. Mahavisno, R. Varambally, J. Yu, B.B. Briggs, T.R. Barrette, M.J. Anstet, C. Kincead-Beal, P. Kulkarni, S. Varambally, D. Ghosh, A. M. Chinnaiyan, Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles, Neoplasia (New York, NY) 9 (2007) 166–180.
[196] D.R. Rhodes, S. Kalyana-Sundaram, S.A. Tomlins, V. Mahavisno, N. Kasper, R. Varambally, T.R. Barrette, D. Ghosh, S. Varambally, A.M. Chinnaiyan, Molecular concepts analysis links tumors, pathways, mechanisms, and drugs, Neoplasia (New York, NY) 9 (2007) 443–454.
[197] E. Jacquinet, N.V. Rao, G.V. Rao, W. Zhengming, K.H. Albertine, J.R. Hoidal, Cloning and characterization of the cDNA and gene for human epitheliasin, Eur. J. Biochem./FEBS 268 (2001) 2687–2699.
[198] T. Takahara, B. Tasic, T. Maniatis, H. Akanuma, S. Yanagisawa, Delay in synthesis of the 3′ splice site promotes trans-splicing of the preceding 5′ splice site, Mol. Cell 18 (2005) 245–251.
[199] D. Communi, N. Suarez-Huerta, D. Dussossoy, P. Savi, J.M. Boeynaems, Cotranscription and intergenic splicing of human *P2Y11* and *SSF1* genes, J. Biol. Chem. 276 (2001) 16561–16566.
[200] J.I. Koontz, A.L. Soreng, M. Nucci, F.C. Kuo, P. Pauwels, H. van Den Berghe, P. Dal Cin, J.A. Fletcher, J. Sklar, Frequent fusion of the *JAZF1* and *JJAZ1* genes in endometrial stromal tumors, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 6348–6353.
[201] H. Li, J. Wang, G. Mor, J. Sklar, A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells, Science (New York, NY) 321 (2008) 1357–1361.
[202] B.P. Lewis, I.H. Shih, M.W. Jones-Rhoades, D.P. Bartel, C.B. Burge, Prediction of mammalian microRNA targets, Cell 115 (2003) 787–798.

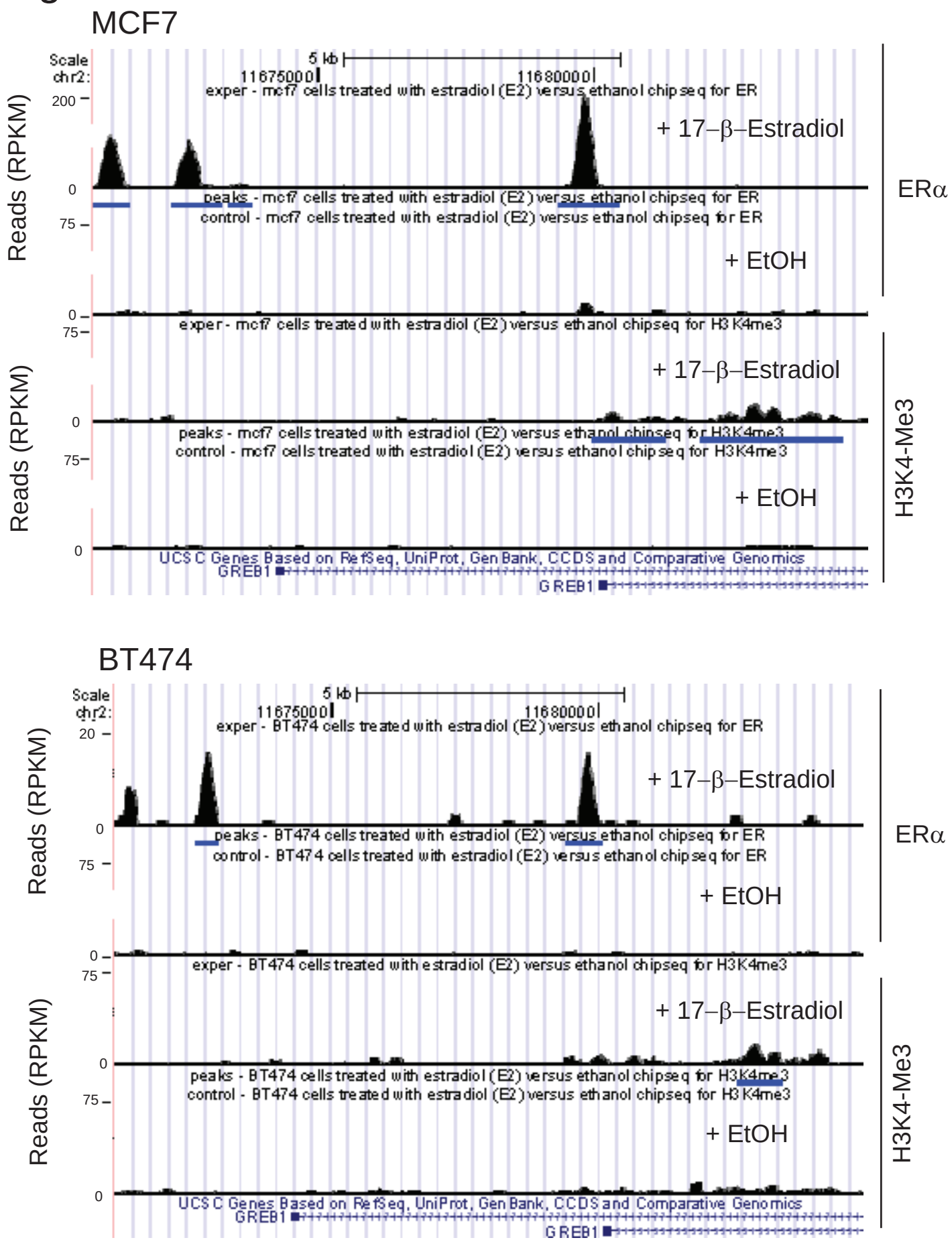
Figure 1

UNPUBLISHED DATA



UNPUBLISHED DATA

# Figure 2



# Figure 3

UNPUBLISHED DATA

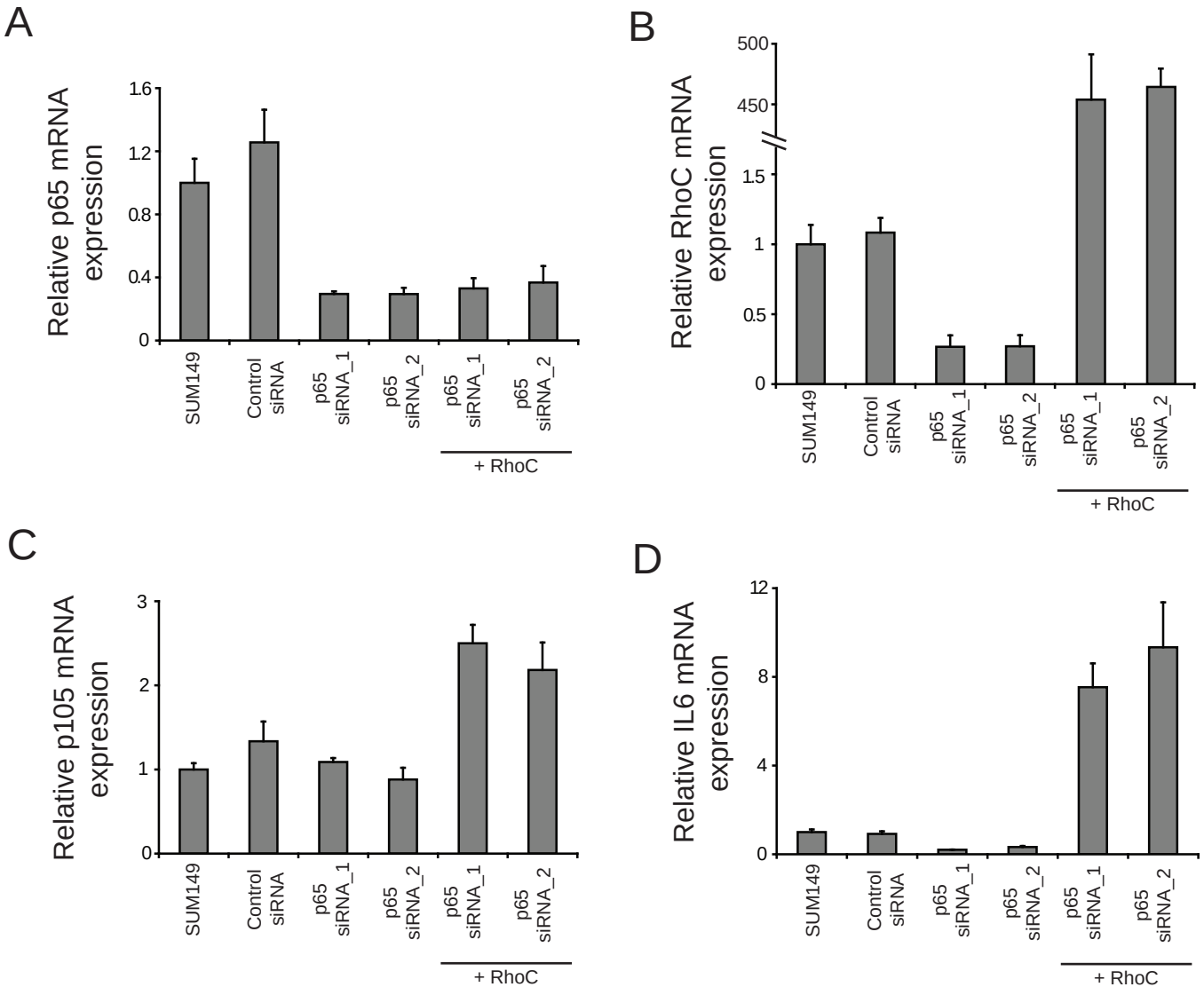
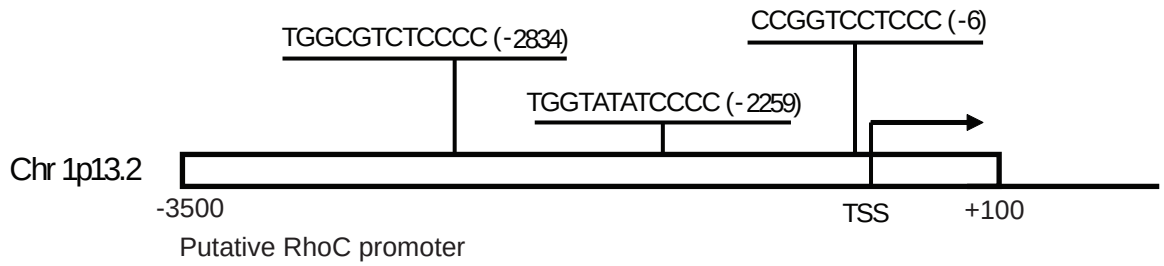


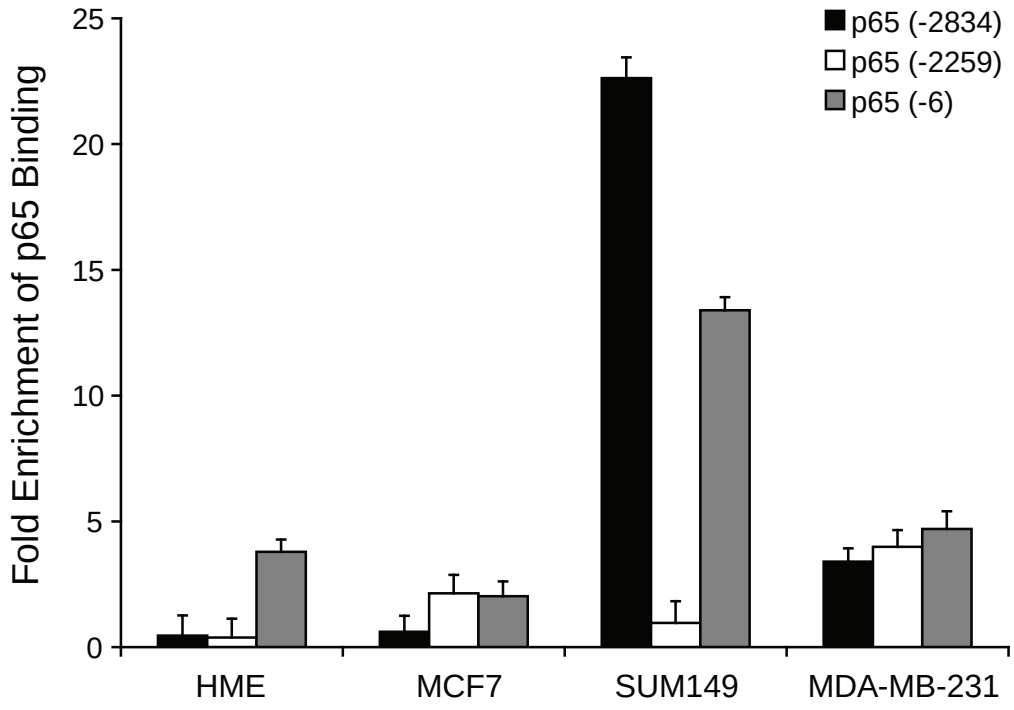
Figure 4

UNPUBLISHED DATA

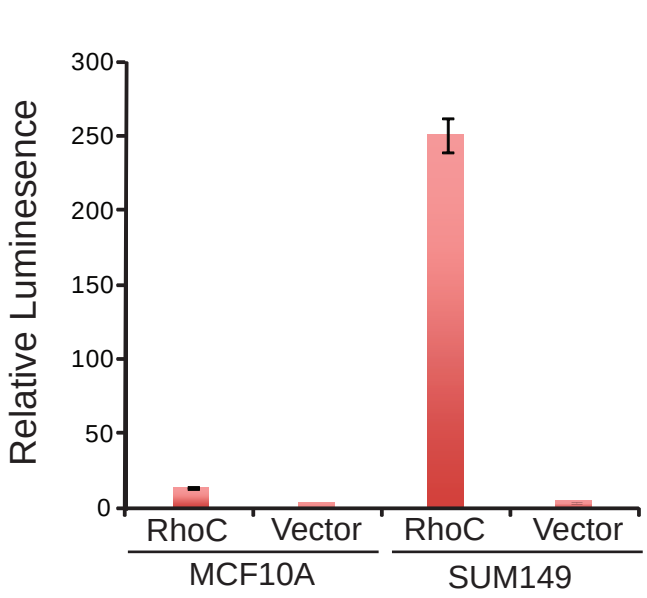
A




B



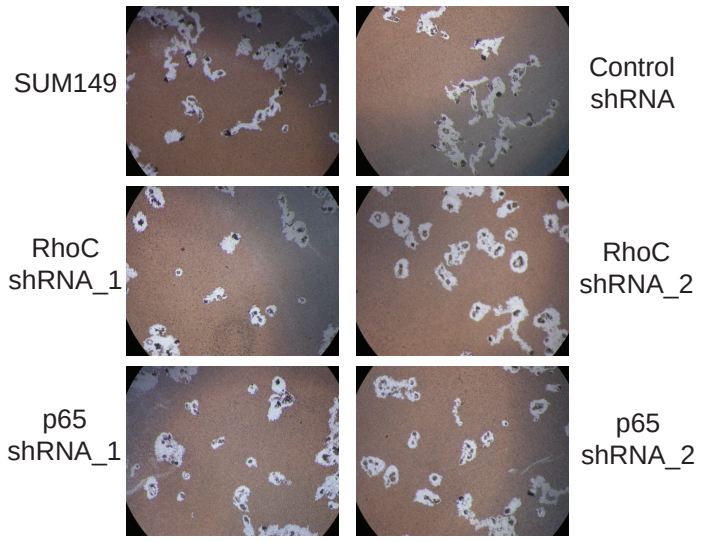

C



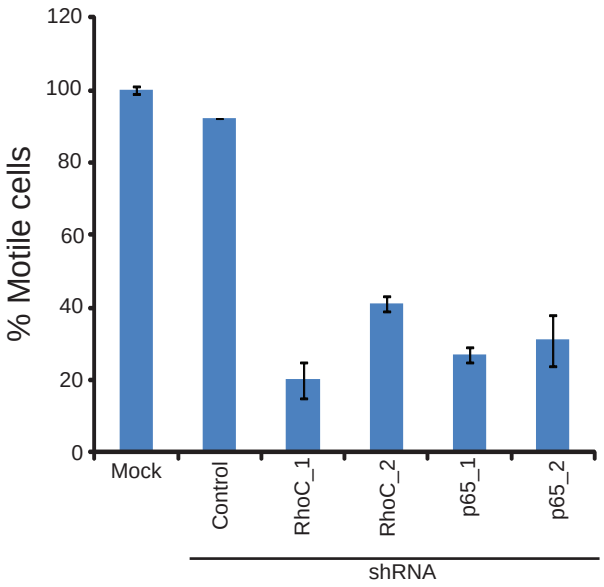
# Figure 5

UNPUBLISHED DATA

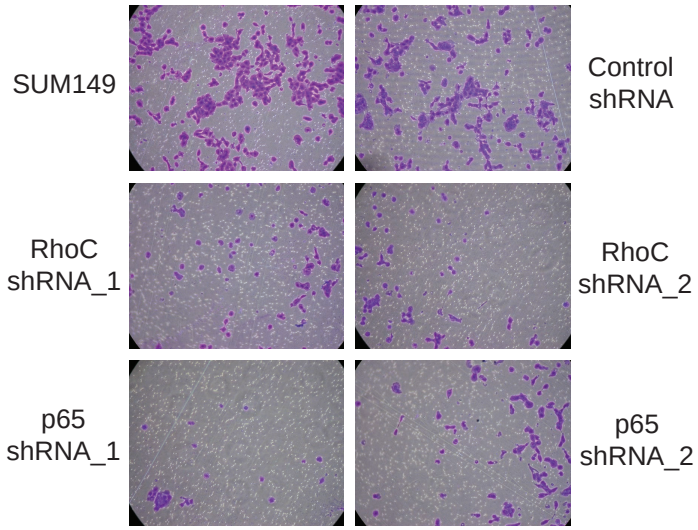
A



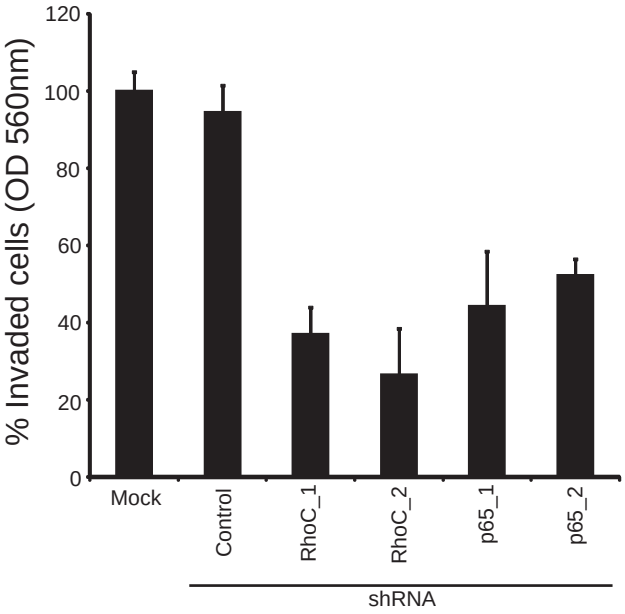

B




C




D



# Table1

| | Prostate | | | | Breast | | | | |
|---|---|---|---|---|---|---|---|---|---|
| TargetID NORM: HME | PrEC | RWPE | PC3 | DU145 | MCF10A | SUM149 | SUM190 | BT20 | HCC1937 |
| hsa-miR-296-5p | 0.832594 | 4.446693 | 0.282801 | 2.938353 | 0.954421 | 0.228308 | 0.168524 | 0.192569 | 0.192236 |
| hsa-miR-708 | 0.597511 | 1.740731 | 0.02235 | 0.023809 | 0.786263 | 0.027347 | 0.024853 | 0.023295 | 0.025006 |
| hsa-miR-663 | 1.742116 | 0.405908 | 0.15265 | 0.387818 | 0.66306 | 0.175342 | 0.078522 | 0.475149 | 0.106728 |
| hsa-miR-31* | 1.43628 | 1.84149 | 0.659765 | 0.909713 | 0.627522 | 0.028181 | 0.021377 | 0.595517 | 0.986864 |
| hsa-miR-31 | 1.105523 | 1.126393 | 0.910725 | 0.981528 | 0.614428 | 0.031018 | 0.023398 | 0.758384 | 0.972423 |

Table 1: Summarized Illumina microRNA array bead station data. microRNAs that were greater than 2-fold down or upregulated in both IBC cell lines SUM149 and SUM190, but not other breast (MCF10A, BT20 and HCC1937) or prostate (PrEC, RWPE, PC3 and DU145) cells as compared to HME cells were identified. Validations were performed by qPCR using Taqman probes specific for both miR-31 and miR-31*.