

GWAS Meets Microarray: Are the Results of Genome-Wide Association Studies and Gene-Expression Profiling Consistent? Prostate Cancer as an Example

Ivan P. Gorlov^{1*}, Gary E. Gallick¹, Olga Y. Gorlova², Christopher Amos², Christopher J. Logothetis¹

1 Department of Genitourinary Medical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas, United States of America, **2** Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas, United States of America

Abstract

Background: Genome-wide association studies (GWASs) and global profiling of gene expression (microarrays) are two major technological breakthroughs that allow hypothesis-free identification of candidate genes associated with tumorigenesis. It is not obvious whether there is a consistency between the candidate genes identified by GWAS (GWAS genes) and those identified by profiling gene expression (microarray genes).

Methodology/Principal Findings: We used the Cancer Genetic Markers Susceptibility database to retrieve single nucleotide polymorphisms from candidate genes for prostate cancer. In addition, we conducted a large meta-analysis of gene expression data in normal prostate and prostate tumor tissue. We identified 13,905 genes that were interrogated by both GWASs and microarrays. On the basis of P values from GWASs, we selected 1,649 most significantly associated genes for functional annotation by the Database for Annotation, Visualization and Integrated Discovery. We also conducted functional annotation analysis using same number of the top genes identified in the meta-analysis of the gene expression data. We found that genes involved in cell adhesion were overrepresented among both the GWAS and microarray genes.

Conclusions/Significance: We conclude that the results of these analyses suggest that combining GWAS and microarray data would be a more effective approach than analyzing individual datasets and can help to refine the identification of candidate genes and functions associated with tumor development.

Citation: Gorlov IP, Gallick GE, Gorlova OY, Amos C, Logothetis CJ (2009) GWAS Meets Microarray: Are the Results of Genome-Wide Association Studies and Gene-Expression Profiling Consistent? Prostate Cancer as an Example. PLoS ONE 4(8): e6511. doi:10.1371/journal.pone.0006511

Editor: Eshel Ben-Jacob, Tel Aviv University, Israel

Received: May 11, 2009; **Accepted:** June 29, 2009; **Published:** August 4, 2009

Copyright: © 2009 Gorlov et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by the David Koch Center for Applied Research in Genitourinary Cancer. Partial support for this study has been provided by US National Institute of Health grants R01CA121197-01A2 to CA and AR055258 subcontract to OG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ipgorlov@mdanderson.org

Introduction

Microarray technology allows simultaneous assessment of the expression of virtually all genes in the genome. This approach has been widely used to identify candidate genes associated with cancer development and progression [1–3]. Genome-wide association studies (GWASs) have recently emerged as a powerful tool to identify genetic polymorphisms associated with cancer risk [4,5]. In a GWAS, hundreds of thousands of single nucleotide polymorphisms (SNPs) are genotyped in a large number of cases and controls. A difference in allelic or genotype frequencies between cases and controls suggests an association between cancer risk and the SNP and a linked gene or regulatory region.

Whether these two approaches produce comparable results has not been examined. Recently Chen et al. [6] identified the genes that tend to be differentially expressed across various experiential conditions and states using gene-expression data from the Gene Expression Omnibus (GEO). They found that differentially expressed genes are more likely to be detected as disease variants in association studies.

In this study, we undertook a more direct approach to link GWAS and microarray data. We performed functional annotations

of the top genes identified in prostate cancer GWASs and the same number of the top candidate genes identified in a meta-analysis of the gene-expression data for normal prostate and prostate tumor. The results of our analyses indicate that these two approaches yield similar results at the functional level.

Materials and Methods

Several prostate cancer GWASs were recently conducted, and a number of candidate genes were identified (Table 1) [7–10]. Though only a few SNPs with the genome-wide level of significance, 10^{-7} , were identified in these studies, a number of SNPs were significant at the level of individual tests but nonsignificant after correction for multiple testing. Such SNPs likely indicate enrichment with causal SNPs that do not reach the genome-wide significance level because of their small effect size or low allele frequency [11].

The GWAS data for this analysis were retrieved from the Cancer Genetic Markers Susceptibility (CGEMS) database, http://cgems.cancer.gov/about/executive_summary.asp. We used the OncoPrint database <http://www.oncoPrint.org/main/index.jsp> to conduct a

Table 1. Candidate genes identified in prostate cancer GWAS.

Gene	GeneID	Comparison	PMID	Microarray Ps
<i>CTBP2</i>	1488	nonaggressive vs. aggressive prostate cancer	18264096	4.6×10^{-6}
<i>EHBP1</i>	23301	case control study	18264098	0.63
<i>SLC22A3</i>	6581	case control study	18264097	1.1×10^{-4}
<i>LMTK2</i>	22853	case control study	18264097	0.38
<i>KLK2</i>	3817	case control study	18264097	1.4×10^{-7}
<i>KLK3</i>	354	case control study	18264097	2.1×10^{-9}
<i>NUDT10</i>	170685	case control study	18264097	5.6×10^{-11}
<i>DAB2IP</i>	153090	nonaggressive vs. aggressive prostate cancer	18073375	8.0×10^{-7}
<i>HNF1B</i>	6928	nonaggressive vs. aggressive prostate cancer	18264096	0.31
<i>HPC1(RNASEL)</i>	6041	controls vs. sporadic prostate cancer	17876339	1.1×10^{-3}
<i>JAZF1</i>	221895	nonaggressive vs. aggressive prostate cancer	18264096	1.5×10^{-10}
<i>MSMB</i>	4477	nonaggressive vs. aggressive prostate cancer	18264096	4.7×10^{-4}
<i>MYC</i>	4609	controls vs. cases	17401366	2.6×10^{-40}

doi:10.1371/journal.pone.0006511.t001

meta-analysis of the number of studies comparing gene expression in normal prostate tissue with that of localized prostate tumor tissue [12]. The complete list of the studies used in the meta-analysis can be found in the supplementary materials (Table S1). We used an extension of Stouffer's method [13] for the meta-analysis. This approach is based on estimating the standard normal deviation, ζ , and is similar to the approach recently proposed by Ochsner et al. [14]. The meta-analysis identified a number of genes differentially expressed between normal prostate and prostate tumor.

Results

As an initial validation of our hypothesis that GWASs and microarrays tend to identify the same genes, we used a meta-analysis of the Oncomine gene-expression data to assess the expression of the GWAS-identified genes (Table 1). We found that all but three (*HNF1B*, *EHBP1*, and *LMTK2*) of the genes were differentially expressed between the normal and tumorous prostate. Therefore, 10 of 13 (77%) of the GWAS genes were differentially expressed in the transition from normal prostate to prostate cancer that is higher than one can expect to detect among randomly chosen 13 genes ($\chi^2 = 20.9$, $df = 1$, $P < 0.0001$).

The prostate cancer GWAS data from CGEMS Phase 1A and Phase 1B, were used in the analysis. We limited our analysis to the gene-associated SNPs to make GWAS and microarray results comparable. We followed the CGEMS designation of the gene-associated SNPs. A total 63,831 gene-associated SNPs belonging to 16,550 unique genes were identified. For each gene, a SNP with the smallest P value was used to characterize an association. If a given SNP was associated with multiple genes, all those associations were included in our analysis. Because in many cases aliases rather than official gene names were used in GWAS, we linked various gene identifiers to the official gene names and EntrezIDs using the latest version of the NCBI gene database (accessed January 17, 2009). Overlapping of the unique GWAS and microarray genes demonstrated that 13,905 genes were assessed in both GWAS and microarray analyses. The list of the genes with corresponding GWAS and microarray P values is shown in Table S1.

To assess whether the GWAS and microarray analyses tend to identify similar sets of genes we assess a correlation between $-\log(P)$

values based on GWAS data and $-\log(P)$ values based on the analysis of gene expression. We found a small but significant (because of the large sample size) positive correlation between GWAS and microarray $-\log(P)$ s (Figure 1).

The Database for Annotation, Visualization and Integrated Discovery (DAVID) [15] was used for the functional annotation of GWAS and microarray genes. We selected genes with GWAS P values ≤ 0.01 . A total of 1,649 genes were identified. We used exactly the same number of the top genes identified in the meta-analysis of the gene-expression data. To control for possible biases in gene selection, we used the list of 13,905 genes as background. Functional annotation charts were used to retrieve an extended annotation coverage that included more than 40 annotation categories [15]. A functional chart for the top GWAS genes can be found in Table S2. Many cell adhesion-related categories are among the top annotation categories. Clustering of the terms of functional annotations summarized all types of the functional description used by DAVID, identifying cell adhesion as the top cluster, followed by plasma membrane and fibronectin.

Functional annotation of the top differentially expressed genes identified cytoskeleton, focal adhesion, extracellular matrix, and cell adhesion as the top annotation terms (Table S3). Clustering of the terms of functional annotation demonstrated cytoskeleton, actin cytoskeleton, extracellular matrix, and cell adhesion among the top identified clusters. Figure 2 shows the results of the clustering of functional terms by DAVID based on the analysis of the top GWAS and differentially expressed genes (see also Table S4). In both lists, most of the top functional clusters derived from GWAS and microarray data are directly or indirectly related to cell adhesion.

We next looked for an overlap between the top 1,649 GWAS and the top 1,649 differentially expressed genes. We identified 248 appearing in both lists genes (see supplementary materials for the list of the genes). This number is higher than would be expected by chance. If we randomly sample 1,649 genes from among the 13,905 genes, the expected number of the genes found in two independent samples would be $(1,649/13,905) \times 2 \times 1,649 = 23.2$. The functional annotation of these 248 genes identified cytoskeleton, focal adhesion, and actin binding as top functional categories. Functional clustering of the genes identified cell migration, cell motility, cytoskeleton, and cell adhesion as the top clusters.

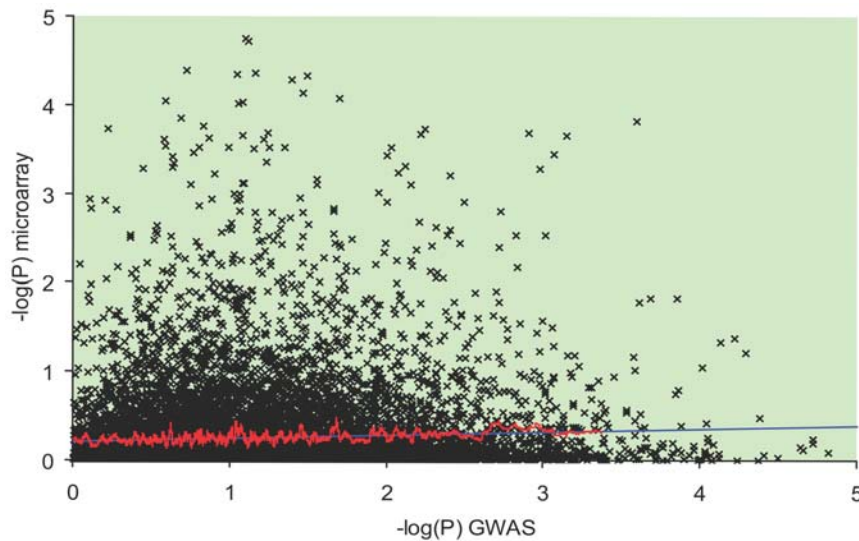


Figure 1. The plot of GWAS and microarray $-\log(P)$ s. Black line shows the linear regression curve, red line – moving average computed using a sliding window of 100 points. Spearman's rank-order correlation coefficient: $r=0.043$, $N=13905$, $P=0.0000001$. doi:10.1371/journal.pone.0006511.g001

Discussion

GWAS and microarray analyses both allow unbiased identification of candidate genes and pathways associated with cancer development. These two approaches each have advantages and drawbacks. By combining data from multiple expression studies, analyses of gene expressions have the statistical power to detect even small differences in gene expression between normal and tumor tissues. On the other hand, because genes in the human genome are involved in multiple interactions, modulation of the expression of a single gene may cause a “ripple effect” on multiple downstream targets, making it difficult to separate causal and induced changes in gene expression. This is unlikely to be an issue in GWASs. GWASs, however, are often statistically underpowered to detect SNPs with small effect size.

When we compared candidate genes for prostate cancer identified by GWAS with those identified by microarray, we noted a significant positive correlation between the GWAS and microarray $-\log(P)$ s. The correlation was small, with the Pearson rank correlation coefficient being only 0.04, but positive correlation between two ranks is expected to be driven by a relatively small number of causal genes. Not all causal genes will be detected by GWAS. Even if the gene is mechanistically linked to prostate tumorigenesis, it can be detected by GWAS only if it carries genetic variants that modulate its function. On the other hand, genes identified by microarray analysis are expected to be a mix of causal genes and the genes that are differentially expressed because of the ripple effect of the causal genes. This suggests that only a fraction of the genes significant in both analyses are causal genes.

We found that the top GWAS and differentially expressed candidates were enriched in cell adhesion genes. If we consider all known cell adhesion genes in the genome, only 74 genes or 10% of them were among the top differentially expressed genes. If the cell adhesion pathway is associated with prostate tumorigenesis, one can expect that other cell adhesion genes—those that did not make it to the top 1,649 genes—also will tend to be significantly positively associated. We found that the average GWAS-derived P value for the cell adhesion genes that failed to reach the top 1649 was lower than the average value for the GWAS genes (t test = 2.9, $df=13,902$, $P=0.001$). A similar result was obtained for the P

values derived from the analysis of the gene expression: the absolute Z score was higher among cell adhesion genes (excluding those among the top 1649 genes) than was the average Z score (t test = 1.81, $df=17811$, $P=0.07$ on the two-tailed test and $P=0.03$ on the one-tailed test). This suggests that cell adhesion function as a whole is associated with prostate tumorigenesis.

Both GWAS and microarray genes form functional clusters related to different aspects of cell adhesion, including cell adhesion itself, cell junction, extracellular matrix glycoproteins, fibronectin, actin cytoskeleton, and cell motility. Several other clusters also show a mechanistic association with cell adhesion. For example, cadherin uptake from the cell surface by endocytosis regulates the level of the free cadherins on the cell surface and therefore cell adhesion [16]. Also, zinc finger proteins with the LIM domain are important for focal adhesion and cell adhesion to fibronectin [17,18]. The modulation of the cell adhesion function seems not to be limited to any specific adhesion type but includes cadherins, integrins, and selectins as well as adhesion molecules associated with tight junctions.

The results of a number of studies suggested the involvement of the cell adhesion system in prostate cancer development. Cadherins play a role in regulating tumor cell proliferation through cyclins and cyclin-dependent kinases [19]. Integrins are involved in different aspects of prostate tumorigenesis, including cell proliferation, cell motility, and apoptosis [20–22]. Modulation of cell adhesion can play an important role in epithelial-to-mesenchymal transition that is believed to be a key step in malignant transformation [23–25]. Also the results of a number of studies suggested an involvement of cell adhesion in angiogenesis [26–28].

GWAS-identified genes are considered to be cancer susceptibility genes that are mainly associated with tumor initiation. We believe, however, that genes identified by GWAS are also likely to include genes important for tumor progression. Indeed, the detection of tumor is usually symptomatic: the tumor needs to reach a certain size to be detected. This suggests that genes involved in tumor progression will be among GWAS-detected candidate genes. Therefore, GWAS and gene expression analysis may target essentially the same set of genes, providing the theoretical basis for the joint analysis of GWAS and microarray data.

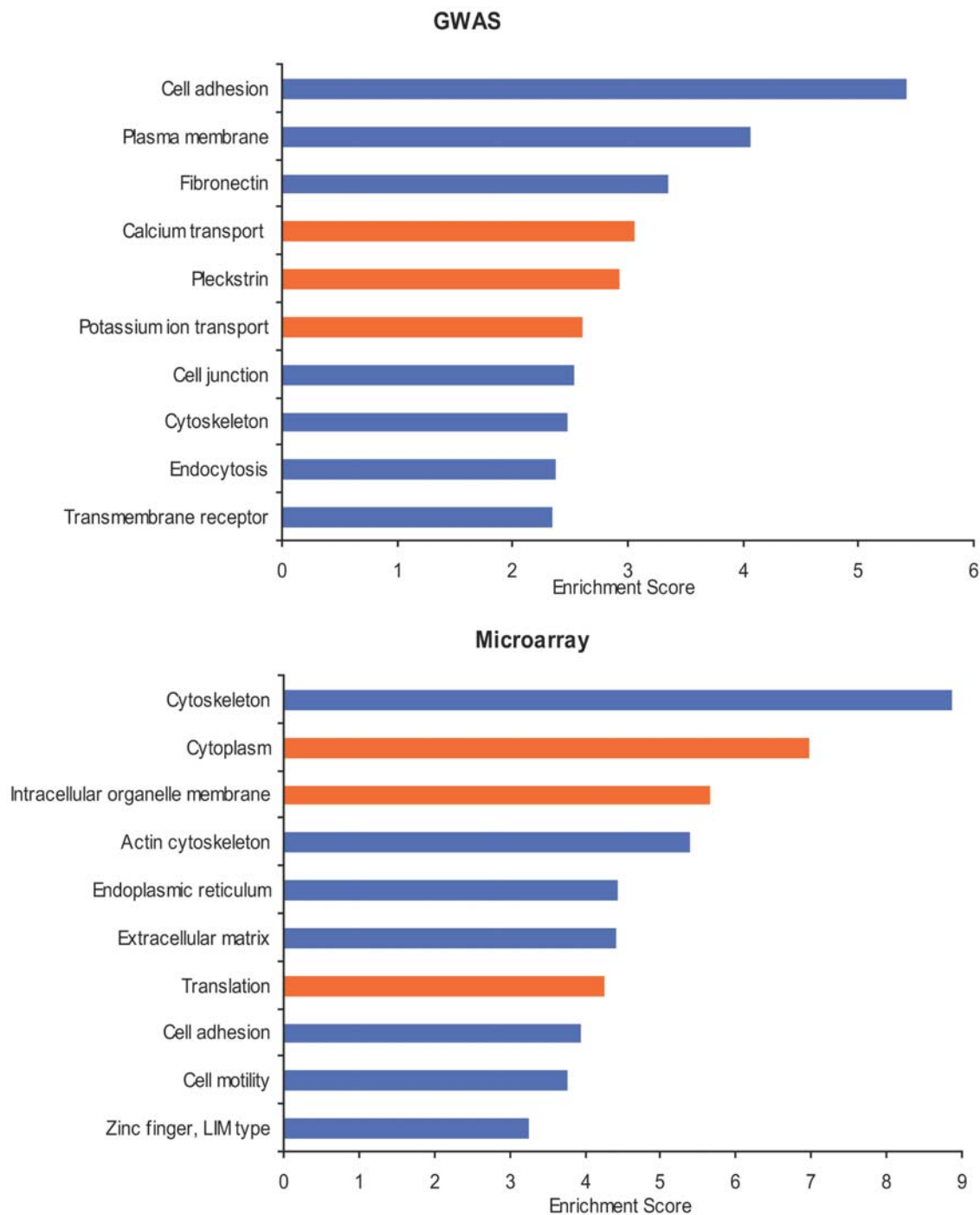


Figure 2. Clustering of the functional annotation terms based on GWAS- (upper panel) and microarray-derived genes (lower panel). Functional clusters related to cell adhesion are shown in blue. Detailed information on the composition of clusters can be found in Table S4.

doi:10.1371/journal.pone.0006511.g002

In summary, our analysis found a considerable overlap between prostate cancer genes identified by GWAS and those identified by global profiling of the gene expression. We identified cell adhesion as a biological function associated with prostate tumorigenesis. The results of this study also suggest that combining GWAS and microarray data might be a more effective approach than using just the analysis of the individual datasets, and can help to refine the identification of candidate genes and/or functions involved in tumor development.

Supporting Information

Table S1

Found at: doi:10.1371/journal.pone.0006511.s001 (0.83 MB PDF)

Table S2

Found at: doi:10.1371/journal.pone.0006511.s002 (0.07 MB XLS)

Table S3

Found at: doi:10.1371/journal.pone.0006511.s003 (0.33 MB XLS)

Table S4

Found at: doi:10.1371/journal.pone.0006511.s004 (0.33 MB XLS)

References

- Lacroix L, Commo F, Soria JC (2008) Gene expression profiling of non-small-cell lung cancer. *Expert Rev Mol Diagn* 8: 167–178.
- Konstantinopoulos PA, Spentzos D, Cannistra SA (2008) Gene-expression profiling in epithelial ovarian cancer. *Nat Clin Pract Oncol* 5: 577–587.
- Bao T, Davidson NE (2008) Gene expression profiling of breast cancer. *Adv Surg* 42: 249–260.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9: 356–369.
- Browning SR (2008) Missing data imputation and haplotype phase inference for genome-wide association studies. *Hum Genet* 124: 439–450.
- Chen R, Morgan AA, Dudley J, Deshpande T, Li L, et al. (2008) FitSNPs: highly differentially expressed genes are more likely to have variants associated with disease. *Genome Biol* 9: R170–173.
- Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, et al. (2007) Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst* 99: 1836–1844.
- Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, et al. (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39: 631–637.
- Nam RK, Zhang WW, Loblaw DA, Klotz LH, Trachtenberg J, et al. (2008) A genome-wide association screen identifies regions on chromosomes 1q25 and 7p21 as risk loci for sporadic prostate cancer. *Prostate Cancer Prostatic Dis* 11: 241–246.
- Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, et al. (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 40: 310–315.
- Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI (2008) Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 82: 100–112.
- Gorlov IP, Gorlova OY, Efsthathiou E, Logothetis CJ (2009) Candidate pathways and genes for prostate cancer: a meta-analysis of gene expression data. *BMC Cancer*, In press.
- Rosenthal R (1979) The file drawer problem and tolerance for null results. *Psychol Bull* 86: 638–641.
- Ochsner SA, Steffen DL, Hilsenbeck SG, Chen ES, Watkins C, et al. (2009) GEMS (Gene Expression MetaSignatures), a Web resource for querying meta-analysis of expression microarray datasets: 17beta-estradiol in MCF-7 cells. *Cancer Res* 69: 23–26.
- Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44–57.
- Delva E, Kowalczyk AP (2009) Regulation of cadherin trafficking. *Traffic* 10: 259–267.
- Brown MC, Perrotta JA, Turner CE (1998) Serine and threonine phosphorylation of the paxillin LIM domains regulates paxillin focal adhesion localization and cell adhesion to fibronectin. *Mol Biol Cell* 9: 1803–1816.
- Hansen MD, Beckerle MC (2006) Opposing roles of zyxin/LPP ACTA repeats and the LIM domain region in cell-cell adhesion. *J Biol Chem* 281: 16178–16188.
- Mason MD, Davies G, Jiang WG (2002) Cell adhesion molecules and adhesion abnormalities in prostate cancer. *Crit Rev Oncol Hematol* 41: 11–28.
- Moschos SJ, Drogowski LM, Reppert SL, Kirkwood JM (2007) Integrins and cancer. *Oncology (Williston Park)* 21(9 Suppl 3): 13–20.
- Fornaro M, Manes T, Languino LR (2001) Integrins and prostate cancer metastases. *Cancer Metastasis Rev* 20(3-4): 321–331.
- Goel H, Li J, Kogan S, Languino L (2008) Integrins in prostate cancer progression. *Endocr Relat Cancer* 15: 657–664.
- Ke XS, Qu Y, Goldfinger N, Rostad K, Hovland R, et al. (2008) Epithelial to mesenchymal transition of a primary prostate cell line with switches of cell adhesion modules but without malignant transformation. *PLoS ONE* 3(10): e3368.
- Takkunen M, Ainola M, Vainionpaa N, Grenman R, Patarroyo M, et al. (2008) Epithelial-mesenchymal transition downregulates laminin alpha5 chain and upregulates laminin alpha4 chain in oral squamous carcinoma cells. *Histochem Cell Biol* 130: 509–525.
- Etienne-Manneville S (2008) Polarity proteins in migration and invasion. *Oncogene* 27: 6970–6980.
- Piao M, Mori D, Satoh T, Sugita Y, Tokunaga O (2006) Inhibition of endothelial cell proliferation, in vitro angiogenesis, and the down-regulation of cell adhesion-related genes by genistein. Combined with a cDNA microarray analysis. *Endothelium* 13: 249–266.
- Bazas VM, Lukyanova NY, Demash DV, Galakhin KO, Myasoedov DV (2008) Relation between cell-to-cell adhesion and angiogenesis and clinico-morphological prognostic factors in patients with gastric cancer. *Exp Oncol* 30: 235–239.
- Ramjaun AR, Hodivala-Dilke K (2009) The role of cell adhesion pathways in angiogenesis. *Int J Biochem Cell Biol* 41: 521–530.

Author Contributions

Conceived and designed the experiments: IPG OYG CL. Performed the experiments: IPG CIA. Analyzed the data: IPG GG OYG. Wrote the paper: IPG GG CIA CL.