

Research article

Open Access

## Correlation of mRNA and protein levels: Cell type-specific gene expression of cluster designation antigens in the prostate

Laura E Pascal<sup>\*1,2,3</sup>, Lawrence D True<sup>2,4</sup>, David S Campbell<sup>3</sup>,  
Eric W Deutsch<sup>3</sup>, Michael Risk<sup>1</sup>, Ilsa M Coleman<sup>5</sup>, Lillian J Eichner<sup>1,3</sup>,  
Peter S Nelson<sup>5</sup> and Alvin Y Liu<sup>1,2,3</sup>

Address: <sup>1</sup>Department of Urology, University of Washington, Seattle WA 98195, USA, <sup>2</sup>Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle WA 98195, USA, <sup>3</sup>Institute for Systems Biology, Seattle WA 98103, USA, <sup>4</sup>Department of Pathology, University of Washington, Seattle WA 98195, USA and <sup>5</sup>Fred Hutchinson Cancer Research Center, Seattle WA 98109, USA

Email: Laura E Pascal\* - lpascal@systemsbiology.org; Lawrence D True - ltrue@u.washington.edu;  
David S Campbell - dcampbell@systemsbiology.org; Eric W Deutsch - edeutsch@systemsbiology.org; Michael Risk - mrisk@u.washington.edu;  
Ilsa M Coleman - icoleman@fhcrc.org; Lillian J Eichner - lillian.eichner@mail.mcgill.ca; Peter S Nelson - pnelson@fhcrc.org;  
Alvin Y Liu - aliu@u.washington.edu

\* Corresponding author

Published: 23 May 2008

Received: 7 December 2007

BMC Genomics 2008, 9:246 doi:10.1186/1471-2164-9-246

Accepted: 23 May 2008

This article is available from: <http://www.biomedcentral.com/1471-2164/9/246>

© 2008 Pascal et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Expression levels of mRNA and protein by cell types exhibit a range of correlations for different genes. In this study, we compared levels of mRNA abundance for several cluster designation (CD) genes determined by gene arrays using magnetic sorted and laser-capture microdissected human prostate cells with levels of expression of the respective CD proteins determined by immunohistochemical staining in the major cell types of the prostate – basal epithelial, luminal epithelial, stromal fibromuscular, and endothelial – and for prostate precursor/stem cells and prostate carcinoma cells. Immunohistochemical stains of prostate tissues from more than 50 patients were scored for informative CD antigen expression and compared with cell-type specific transcriptomes.

**Results:** Concordance between gene and protein expression findings based on 'present' vs. 'absent' calls ranged from 46 to 68%. Correlation of expression levels was poor to moderate (Pearson correlations ranged from 0 to 0.63). Divergence between the two data types was most frequently seen for genes whose array signals exceeded background (> 50) but lacked immunoreactivity by immunostaining. This could be due to multiple factors, e.g. low levels of protein expression, technological sensitivities, sample processing, probe set definition or anatomical origin of tissue and actual biological differences between transcript and protein abundance.

**Conclusion:** Agreement between these two very different methodologies has great implications for their respective use in both molecular studies and clinical trials employing molecular biomarkers.

## Background

Immunostaining and microarray analysis are techniques frequently used to characterize tissue phenotypes. Immunohistochemistry (IHC) is a method of assessing protein levels of gene expression that is based on the ability of antibodies to bind proteins expressed by cells in sections of frozen or formalin-fixed, paraffin-embedded tissues. IHC enables one to detect and localize a specific antigen to specific cell types. Gene arrays determine expression levels for thousands of genes simultaneously by detecting sequence segments or partial segments of mRNA in a sample. To fully understand the underlying mechanisms of biological processes, it is essential to determine whether observed changes in mRNA can also be seen in the translated protein, and to pinpoint what cell types are exhibiting these changes. Gene array analysis and immunostaining are powerful tools for determining gene and protein expression patterns in health and diseases. Establishing the extent of agreement between semi-quantitative immunostaining data and gene array data obtained from sorted cell populations and tissue specimens is important to account for possible discrepancies between these two very different methods.

Determining a direct relationship between protein and mRNA levels can be problematic, and previous efforts to find correlations have found variable success. A study comparing yeast proteomic and transcriptomic data showed that correlation was insufficient to predict protein expression levels from mRNA except for the most abundant proteins, suggesting that protein abundance may be a factor that influences the correlation between mRNA and protein [1]. However, a relationship between mRNA/protein correlation coefficient and protein abundance was not observed in a study of human lung adenocarcinomas [2] or in a study of MMP-2, MMP-9 and TIMP-1 in human prostate cancers [3]. For some genes, such as HER2/neu, expression levels assayed by RT-PCR, IHC and fluorescence in situ hybridization (FISH) in breast tumors show highly significant correlation among these techniques [4]. However, studies evaluating the overall concordance between protein and RNA expression levels have found wide variability. For example, transcript and protein concordance in the LNCaP prostate cancer cell line has been reported to vary from 32% [5] to 83.5% [6]. Highly significant correlations in mRNA changes and protein expression levels were found by Orntoft *et al.* in human carcinomas [7]. Studies such as these suggest that external factors as well as actual biological differences between mRNA and protein abundance might affect the relationships between the two data types.

The assessment of CD24 as a potential prostate cancer biomarker through RNA expression profiling and IHC analysis in previous studies further illustrates the difficul-

ties in directly comparing gene and protein expression levels. Normalized CD24 transcript levels showed an average 2.69-fold increase in prostate cancer as determined by qPCR [8] and an increase in staining intensity as determined by IHC [9]. Two reports by Kristiansen *et al.* found differential CD24 gene expression in 38.5% of tumor cases as determined by Affymetrix GeneChip analysis [10] and 48% as examined by IHC [11]. It is unclear if the 10% disparity between these studies has biological significance, or if a component of the measured difference is due to technical attributes of the assays. Although, based on these studies, CD24 could potentially be an important prognostic prostate cancer tissue marker, the relationship between mRNA levels and resulting protein expression remains unclear.

Our laboratory has previously characterized benign and neoplastic prostate tissues by CD phenotype [9,12] and isolated the constituent cell types with magnetic cell sorting (MACS) for gene expression analysis [13]. In the present study, we compared the expression levels of a panel of 58 informative CD antigens in prostate tissue scored by immunostaining with expression levels of the respective mRNA by microarray-based quantitation of MACS-sorted and laser capture microdissected (LCM) cell populations. Our objective was to characterize the relationship between protein and mRNA levels measured by these two techniques.

## Results

### **Immunolocalization and transcriptome data summary**

To determine correlations between disparate methods of assessing gene expression levels, we selected a cohort of CD antigens as a reference gene set and compared transcript abundance measurements acquired by microarray hybridization with protein abundance measurements determined by immunohistochemistry in specific cell types found in the human prostate gland. The microarray and immunostaining results for each cell type are summarized in Figure 1. All annotations are deposited in a public database [14]. The sample data conform to the Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE) standards [15], and describe the tissue, distribution of reaction product in the tissue, localization pattern within cell types, and provide an assessment of the level of protein expression based on the immunostaining data. Our initial criteria for assessing expression centered on using 'present' *vs.* 'absent' call for both Affymetrix and Agilent array analyses. Based on this qualification, CD antigen expression for MACS-sorted cell populations agreed with the following frequency to IHC data by cell type: endothelial = 55%, luminal = 64%, stromal = 46%, basal epithelial = 63%, stem = 64%, cancer = 68%. CD antigen expression for LCM cells analyzed by Affymetrix array agreed 39% for

Cell type:		Endothelial			Luminal			Stromal			Basal		Progenitor		Cancer (Gleason Pattern 3 + 3 = 6 tumor)
CD	HUGO Gene Symbol	IHC	Affymetrix mRNA	IHC	Affymetrix mRNA	Agilent LCM mRNA	IHC	Affymetrix mRNA	Affymetrix LCM mRNA	Agilent LCM mRNA	IHC	Affymetrix mRNA	IHC	Affymetrix mRNA	
CD1a	CD1A	0.00	35	0.00	32	31	0.10	28	171	133	0.00	29	0.00	24	22
CD6	CD6	0.00	32	0.04	31	58	0.00	26	92	33	0.00	27	0.00	26	0.35
CD9	CD9	0.00	6161	1.00	7062	60150	0.22	2714	467	11577	1.00	6899	1.00	8802	0.63
CD10	MME	0.00	55	0.61	1416	549	0.01	223	687	32	0.00	562	0.00	40	0.17
CD13	ANPEP	0.23	221	0.99	361	12623	0.23	97	32	269	0.24	79	0.24	103	0.19
CD24	CD24	0.00	219	0.19	1254	1162	0.01	405	142	212	0.00	548	0.00	120	0.96
CD26	DPP4	0.58	40	0.75	759	31647	0.00	48	147	275	0.00	124	0.00	29	0.80
CD29	ITGB1	0.05	9	0.00	9	14926	0.64	9	2033	18957	0.00	10	0.00	10	0.33
CD30	TNFRSF8	0.00	25	0.05	24	25	0.03	23	217	40	0.00	23	0.00	26	0.00
CD32	FCGR2B	0.00	38	0.00	20	315	0.00	37	46	255	0.07	20	0.07	36	0.00
CD38	CD38	0.05	143	0.62	670	2652	0.00	101	86	40	0.29	163	0.29	51	0.56
CD40	CD40	0.00	402	0.00	182	117	0.00	170	177	84	0.18	157	0.18	398	0.31
CD41	ITGA2B	0.00	27	0.00	29	274	0.00	28	24	139	0.00	28	0.00	29	0.00
CD43	SPN	0.00	37	0.00	39	3034	0.00	35	183	2233	0.00	36	0.00	39	0.00
CD44	CD44	0.00	1009	0.00	606	5483	0.00	780	224	2206	0.30	701	0.30	565	0.19
CD45	PTPRC	0.00	310	0.00	219	231	0.00	227	363	159	0.00	23	0.00	100	0.00
CD46	CD46	0.00	318	0.00	477	3650	0.00	222	399	992	0.00	398	0.00	372	0.66
CD47	CD47	0.00	279	0.08	814	13379	0.31	376	1826	2484	0.63	482	0.63	282	0.00
CD49a	ITGA1	0.00	889	0.00	320	48	0.94	2360	1331	1590	0.00	303	0.00	1038	0.00
CD49b	ITGA2	0.30	1533	0.02	2115	817	0.00	610	142	166	0.00	2327	0.51	1231	0.62
CD49c	ITGA3	0.71	84	0.00	68	4393	0.31	50	86	920	0.00	138	0.00	103	0.00
CD49f	ITGA6	0.90	371	0.00	257	266	0.00	151	89	39	0.95	717	0.95	304	0.00
CD50	ICAM3	0.00	158	0.00	192	2605	0.00	151	84	551	0.00	108	0.00	183	0.00
CD54	ICAM1	0.62	2675	0.00	1197	3508	0.00	1046	218	2452	0.00	924	0.00	2529	0.00
CD56	NCAM1	0.00	38	0.00	30	76	0.00	36	152	1477	0.00	36	0.00	31	0.00
CD57	B3GAT1	0.00	30	0.51	53	404	0.00	27	23	47	0.03	33	0.03	34	0.67
CD58	CD58	0.33	560	0.03	343	292	0.06	230	175	61	0.01	280	0.01	509	0.01
CD59	CD59	0.38	2800	0.00	1681	1258	0.55	1247	1144	1858	0.68	2418	0.68	3658	0.00
CD62L	SELL	0.00	151	0.00	29	22	0.00	52	168	43	0.00	30	0.00	36	0.00
CD63	CD63	0.00	6719	0.81	4255	12095	0.00	4524	5380	17548	0.00	4872	0.00	7777	0.91
CD64	FCGR1A	0.28	52	0.53	28	0.00	0.04	24	56	0.00	0.00	10	0.00	21	0.00
CD66f	PSG1	0.00	9	0.00	10	33	0.00	9	75	34	0.00	35	0.00	10	0.10
CD70	CD70	0.00	124	0.00	94	0.00	0.00	70	303	0.00	0.00	75	0.00	111	0.00
CD73	NT5E	0.52	122	0.00	30	2162	0.00	103	669	5959	0.00	94	0.00	103	0.00
CD74	CD74	0.76	441	0.00	145	3731	0.00	161	209	1505	0.38	175	0.38	374	0.00
CD75s	ST6GAL1	0.00	712	1.00	743	2095	0.01	185	145	80	0.04	490	0.04	976	0.50
CD84	CD84	0.00	29	0.00	21	112	0.00	19	82	208	0.00	20	0.00	24	0.00
CD88	CSAR1	0.00	150	0.00	33	687	0.00	59	174	276	0.00	52	0.00	118	0.00
CD90	THY1	0.71	255	0.00	105	498	0.15	372	277	962	0.00	83	0.00	183	0.00
CD91	LRP1	0.00	54	0.00	50	1687	0.05	94	69	3719	0.00	51	0.00	54	0.00
CD92	SLC44A1	0.00	220	0.02	115	1195	0.00	397	312	1060	0.30	227	0.30	243	0.00
CD94	KLRD1	0.00	33	0.01	25	22	0.02	26	42	29	0.00	24	0.00	24	0.00
CD95	FAS	0.02	230	0.18	133	482	0.00	536	260	561	0.43	378	0.43	121	0.03
CD98	SLC3A2	0.00	1561	0.37	1717	1456	0.00	904	142	503	0.02	1386	0.02	1694	0.14
CD103	ITGAE	0.00	472	0.00	399	875	0.07	564	406	249	0.00	444	0.00	670	0.00
CD104	ITGB4	1.00	23	0.02	21	2665	0.00	19	58	1051	0.71	23	0.71	22	0.00
CD107a	LAMP1	1.00	1775	1.00	2056	2248	0.36	2117	12017	654	0.23	2053	0.23	2158	0.19
CD107b	LAMP2	0.45	914	0.99	3101	26473	0.03	1219	1382	10028	0.49	1253	0.49	1077	0.67
CD138	SDC1	0.00	83	0.01	144	1180	0.05	113	119	286	0.62	547	0.62	120	0.00
CD144	CDH5	1.00	917	0.00	221	39	0.00	224	77	40	0.00	527	0.00	1281	0.00
CD147	BSG	0.00	179	0.55	225	1056	0.05	302	151	524	0.45	251	0.45	232	0.37
CD153	TNFSF8	0.00	47	0.00	33	108	0.03	30	108	57	0.00	31	0.00	31	0.02
CD164	CD164	0.00	510	1.00	557	14331	0.05	435	2333	5228	0.01	453	0.01	525	0.03
CD166	ALCAM	0.00	795	0.16	1607	15170	0.00	376	1038	5373	0.03	815	0.03	493	0.21
CD179a	VPREB1	0.00	11	0.00	11	148	0.03	10	51	194	0.00	11	0.00	12	0.01
CD226	CD226	0.00	13	0.00	14	45	0.03	12	102	48	0.00	12	0.00	13	0.01
CD271	NGFR	0.00	22	0.00	23	145	0.00	22	36	45	0.61	24	0.61	23	0.00

Affymetrix mRNA data: \*Genes in the dataset with an average raw fluorescence signal <50 or > 50% of probesets measured absent by Affymetrix GCOS Software analysis were considered to be undetected by the experiment

Expression level color scale: None Low Mid High

Immunohistochemistry (IHC) data: CD's in the dataset with scores of 0-0.250 were considered to be no more than background or no protein expression

Expression level color scale: 0 - 0.25 0.251 - 0.500 0.501 - 0.750 0.751 - 1.000

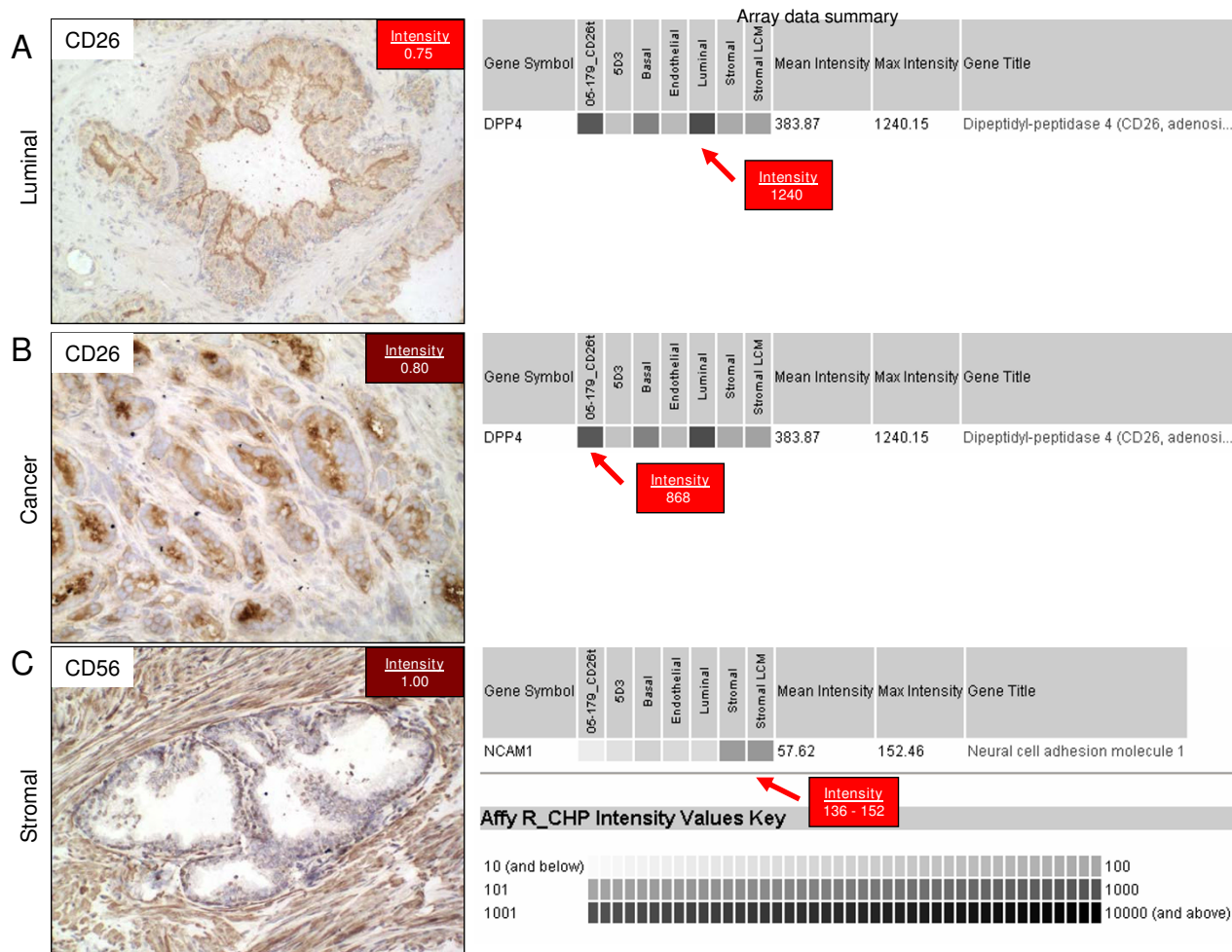
Agilent LCM array data: Genes in the dataset with an average raw fluorescence signal <100 were considered absent

Expression level color scale: 0-99 100-999 1000-9999 10000-99999

**Figure 1**  
**CD expression summary.** CD expression levels of prostate tissue as scored by immunostaining (IHC) and gene array (mRNA) for each constituent cell type. Overall concordance between IHC and MACS-sorted cell gene array ranged from 46 to 67% and was determined as the percentage of agreement for each cell type based on 'present' or 'absent' call for the total number of IHC and mRNA determinations. Concordance between IHC and LCM Affymetrix gene array for stromal cells was 39%. Concordance between IHC and LCM Agilent gene array was 25% for stromal cells and 36% for luminal cells.

stromal cells with IHC data, and analyzed by Agilent array agreed 25% for stromal cells and 36% for luminal cells.

We next sought to establish a more quantitative relationship between CD immunoreactivity of prostate cells and corresponding gene expression data (Figure 2). For exam-

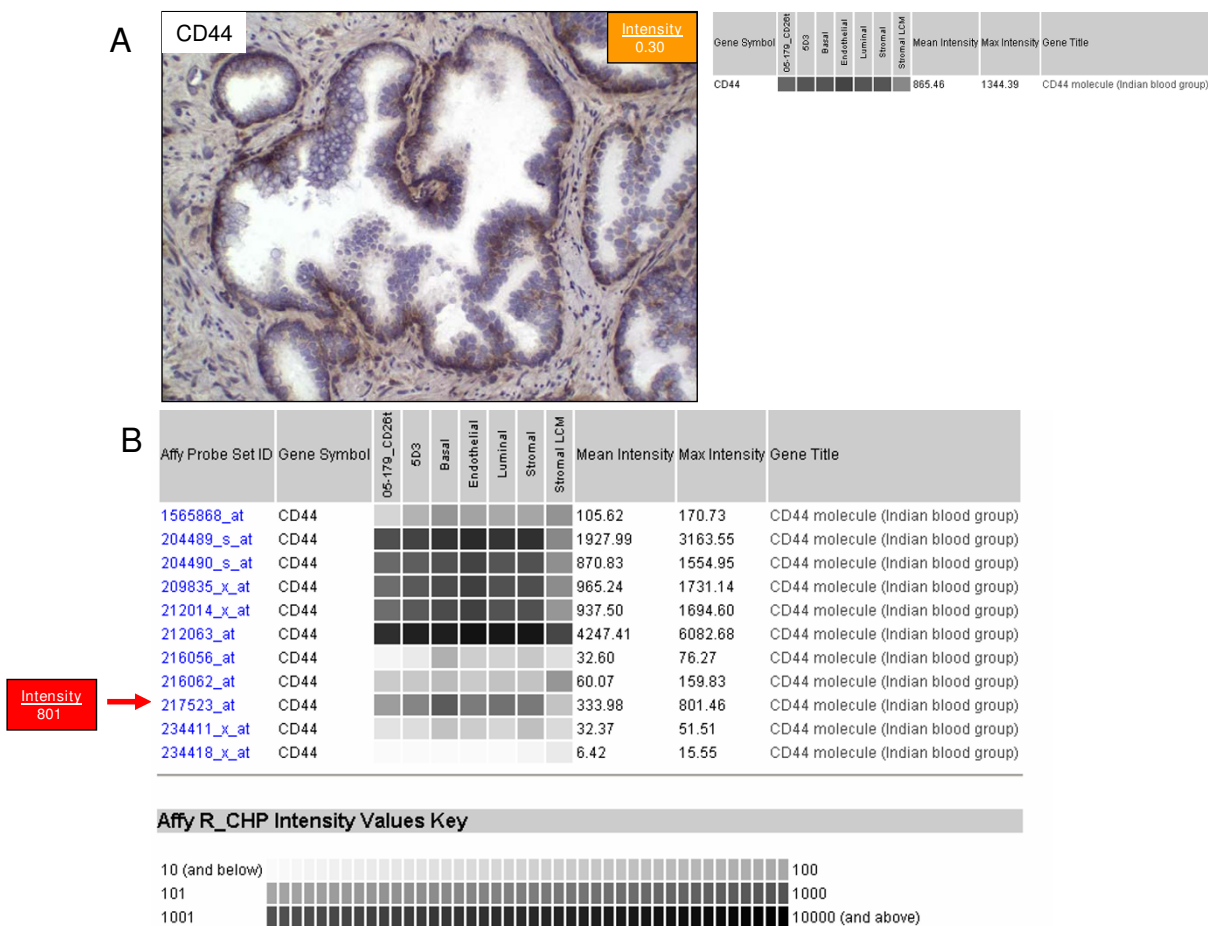


**Figure 2**  
**Data-type concordance for CD26 and CD56.** CD immunoreactivity of prostate cells (immunoreaction product red-brown; pale blue hematoxylin nuclear counterstain) and corresponding gene expression data (darker shading of the boxes indicates higher mRNA levels). **A.** CD26 staining in normal prostate tissue was confined to luminal cells with a staining intensity of 0.75, this correlated well with array data summary where a maximum gene expression level of 1240 was found for luminal cells. **B.** CD26 staining in prostate cancer was restricted to cancer cells with a staining intensity of 0.80, this correlates with array data where a maximum gene expression level of 868 was found for the cancer cells. **C.** CD56 staining in normal prostate tissue was confined to stromal cells with a staining intensity of 1.00, this agreed with array data where maximum gene expression level of 136 was found for stromal cells. Original magnification: 200x.

ple, staining for CD26 (DPP4) in normal (A) and cancer (B) prostate tissue shows clear luminal staining, which corresponds with intense gene expression measured by array in the luminal and cancer (05-179\_CD26t) cell populations. The greyscale gradient indicates robust multi-array average (RMA) normalized Affymetrix signal intensity. Signals of 10 or less are represented as white and signals greater than or equal to 10,000 are represented as black. Higher Affymetrix signal (more black) indicates higher levels of gene expression. CD26 expression determined by Agilent array in LCM luminal cells was also in concordance (data shown in Figure 1). Immunostaining

with CD56 (NCAM1) showed clear stromal staining and corresponding gene expression levels in both the MACS-sorted and LCM stromal cell populations. Agilent array analysis results for CD56 were 'absent' in luminal and 'present' in stromal cells as well (data shown in Figure 1).

Data type discordance is typified by CD44 (Figure 3). Immunostaining with CD44 (G44-26, BD PharMingen, San Diego, CA), was confined to the basal cells (A), however, measured gene expression was high in all cell types (B). This antibody recognizes the epitope 1 of the standard variant of CD44 antigen. Affymetrix has 11 probe sets

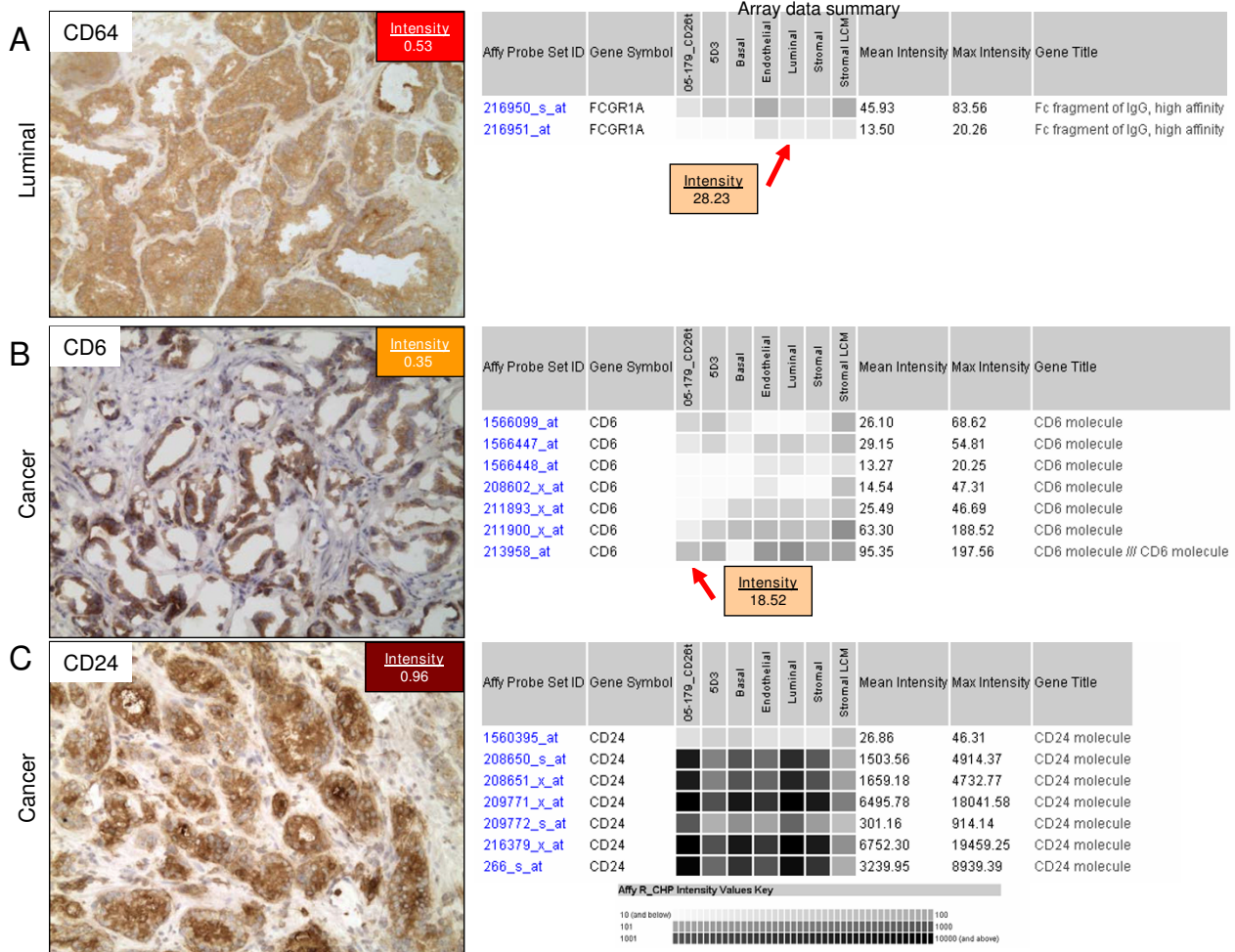


**Figure 3**  
**Data-type discordance for CD44.** CD44 immunoreactivity of prostate cells and corresponding gene expression data. **A.** CD44 staining in normal prostate was confined to basal cells with a staining intensity of 0.30, this contrasted with array data where gene expression levels were similar across all cell types. Original magnification: 200×. **B.** Examination of individual probesets showed that 217523\_at was most similar to immunostaining results with a maximum gene expression level of 683 for basal cells.

for CD44, showing variable expression profiles, of which only probe set 217523\_at showed agreement with IHC results. CD44 has a wide range of splice variants [16] which may account for the discrepancy. Although microarray experiments detect mRNA, and it is mRNA that has the direct relationship with protein, the methodology and algorithms for Affymetrix data analysis are gene-based. Gene level annotation fails to discriminate multiple mRNAs transcribed from the same gene [17] such as is the case with CD44 [18]. Affymetrix probesets for CD44 were analyzed with Basic Local Alignment Search Tool (BLAST). Probesets 204489\_s\_at, 204490\_s\_at, 209835\_x\_at, and 212063\_at map directly to multiple variants of CD44, whereas the other probesets map to individual variants. It is likely that gene level annotation

compared to protein immunostaining will be less straightforward in cases such as CD44, where multiple protein isoforms are generated from a single gene.

Further examples of data discordance for CD64 (FCGR1A), CD6 and CD24 expression are shown in Figure 4. For CD64, immunostaining identified the luminal cell population, but measured gene expression was low or absent for all cell types regardless of probeset (A). This was also the case for CD6 (B). CD6 shows equivocal staining of luminal cells. Examination of CD24 immunostaining identifies the cancer cells, however measured gene expression was high across all cell types for some probe sets (i.e. 209771\_x\_at) and absent for others (i.e. 1560395\_at).



**Figure 4**  
**Data-type discordance for CD64, CD6 and CD24.** CD immunoreactivity of prostate cells and corresponding gene expression data. **A.** CD64 staining of normal prostate was confined to luminal cells with a staining intensity of 0.53, contrasting with the array data where gene expression levels were absent across all cell types and for all probe sets. **B.** CD6 staining in prostate cancer was confined to cancer cells with a staining intensity of 0.35, in contrast with the array data where gene expression levels were absent across all cell types and for all probe sets with the exception of probe set 213958\_at, where cancer, endothelial and luminal expression were present. **C.** CD24 staining in prostate cancer was confined to luminal cells with a staining intensity of 0.96, this contrasts with the array data where gene expression levels were present across all cell types and for all probesets with the exception of probeset 1560395\_at, where expression levels were absent for all cell types. Original magnification: 200x.

Gene expression in LCM samples was also discordant (data shown in Figure 1).

**Correlation of immunolocalization with array data**

Correlations between IHC staining intensity and gene expression levels determined by Affymetrix array for MACS-sorted cell populations are summarized in Table 1. Statistical Pearson correlation ranged from -0.01 – 0.55 and Spearman Coefficients ranged from 0.00 – 0.51. There was no positive overall correlation of immunolocalization data with array data for endothelial, stromal or

progenitor cells, and moderate positive correlation for luminal, basal and cancer. Scatter plots are shown in Additional File 1. It has been suggested that under conditions where expression differences are not dramatic, problems related to probe and probe set identity could lead to significant errors [19]. Using updated probe set definitions provided by Dai et al. [19], statistical correlations ranged from 0.09 – 0.57 and Spearman Coefficients ranged from 0.03 – 0.63 and are summarized in Table 2. Updated probe set definitions increased correlation for all cell types except for cancer cells (0.45 – 0.42). The greatest

**Table 1: Correlation between immunostaining intensity and mRNA level for each prostate cell type**

Cell type	Median		Mean	
	Spearman coefficient	Pearson	Spearman coefficient	Pearson
CD31 Endothelial	0.25	0.28	0.26	0.28
CD26 Luminal	0.46	0.47	0.51	0.55
CD49a Stromal	-0.00	-0.01	0.06	0.07
CD104 Basal	0.35	0.31	0.42	0.34
ABCG2 Progenitor	0.22	0.17	0.34	0.23
CD26 Cancer	0.45	0.49	0.49	0.50

improvement was a 2-fold increase for progenitor (Spearman 0.22 – 0.41). Scatter plots are shown in Additional File 2. In summary, for either probe set definition, there was no positive correlation of immunolocalization with transcriptome data for endothelial or stromal cells, and moderate positive correlation for luminal, basal, and cancer. Updated probe set definitions increased correlation from none to moderate positive correlation for progenitor cells.

**LCM confirmation**

Correlations between IHC staining intensity and gene expression determined by Affymetrix and Agilent array for LCM cells are summarized in Table 3. Statistical Pearson correlation ranged from 0.26 – 0.57 and Spearman Coefficients ranged from 0.25 – 0.43. There was no positive correlation of immunolocalization data with array data for LCM stromal cells assayed by either Agilent or Affymetrix. Pearson correlation for LCM luminal cells was 0.57 and Spearman coefficient was 0.43, which were the same values as that for MACS-sorted luminal cells using updated probe set definitions. Pearson correlation between LCM stromal cell data and MACS-sorted stromal cell data assayed by Affymetrix array was 0.68, Spearman coefficient was 0.78. Pearson correlation between LCM stromal cell data analyzed by Agilent array and Affymetrix MACS-sorted stromal cell data was 0.65, Spearman coefficient was 0.55. Pearson correlation between LCM stromal cell data analyzed by Agilent array and Affymetrix LCM

stromal cell data was 0.33, Spearman coefficient was 0.59. Scatter plots are shown in Additional File 3.

**Discussion**

In this study, we analyzed the gene and protein expression of 58 informative CD antibodies in prostate tissue to determine the degree of concordance between immunohistochemical assessments of protein abundance and microarray-based measures of transcript levels. Based on previous characterization of prostate cancer by CD phenotype [13,20], we were interested in the correlation with corresponding cell type-specific transcriptome data and the relationship between MACS-sorted cell and laser capture microdissected populations. The CD antibodies used for comparison were selected based on their previously determined cell type specificity in prostate tissue [20]. For this study, there was good agreement between CD phenotype characterized by immunostaining and gene expression level based on 'absent' or 'present' call. Quantitative comparison of expression levels mirrored these results with correlation being lowest for stromal cells and highest for luminal cells and cancer cells. There was moderate positive correlation between measured gene expression level and immunostaining intensity for luminal, basal and cancer. There was moderate positive correlation for progenitor cells using updated probe set definitions.

Previous work in our lab has indicated that Affymetrix 'present' calls are approximately 90% accurate, however,

**Table 2: Correlation between immunostaining intensity and mRNA level using updated probeset definitions for each prostate cell type**

Cell type	Median		Mean	
	Spearman coefficient	Pearson	Spearman coefficient	Pearson
CD31 Endothelial	0.43	0.57	0.28	0.27
CD26 Luminal	0.57	0.61	0.57	0.63
CD49a Stromal	0.09	0.03	0.09	0.03
CD104 Basal	0.49	0.45	0.49	0.45
ABCG2 Progenitor	0.41	0.33	0.41	0.34
CD26 Cancer	0.42	0.49	0.42	0.49

**Table 3: Correlation between immunostaining intensity and mRNA level for LCM samples.**

Cell type	Mean	
	Spearman coefficient	Pearson
CD26 Luminal Agilent	0.43	0.57
CD49a Stromal Affymetrix	0.26	0.26
CD49a Stromal Agilent	0.25	0.28

'absent' calls are only 40% accurate [21]. Most often genes classified by array analysis as absent were, in fact, expressed when measured by more sensitive RT-PCR methods, but were of very low abundance. With respect to characterizing CD expression by immunostaining, there were few cases where CD expression was scored by immunostaining and not by gene arrays. In fact, it was more likely that CD genes classified as present by array were not immunostained by that CD antibody. There are several possible explanations for these discrepancies. Biological reasons for poor correlation include post-transcriptional and post-translational modifications, as well as the possibility that proteins have very different half-lives [22,23]. Recent analyses indicate that protein concentrations correlate with the corresponding mRNA levels by only 20 – 40%, and that mRNA abundances are only a weak surrogate for corresponding protein concentration [24,25]. Further complicating matters is the potential for errors and noise in protein and mRNA experiments. Although we have previously documented the advantages of MACS-sorted cell over whole tissue transcriptomes [13] gene expression may be affected by collagenase tissue digestion and cell isolation processes. To examine this possibility we compared array results from MACS-sorted cells with LCM cells. Results for luminal cells were virtually the same

for LCM and MACS-sorted cells, however for stromal cells, correlation increased about 3-fold (from 0.08 in MACS-sorted cells to 0.26 in LCM indicating that for stromal cells the non-specific detection is due to low levels of contaminating cells in the MACS-sort cells. Gene expression analysis is much more sensitive than immunohistochemistry, but it may also be that genes expressed are at levels not high enough for translated protein expression. Satisfactory immunostaining of surface membrane antigens requires preservation of protein tertiary structure in the region of the relevant epitopes [26] and it may be that some of these CD antibodies are less immunoreactive than others. Additionally, some mRNAs may cross-hybridize probes in the array that are designed to detect another mRNA, or probes that are designed to detect the mRNA of a particular gene may be relying on genomic expressed sequence tag (EST) information that is incorrectly associated with that gene. Clearly there are difficulties associated with some antigens such as those we found in CD44, CD64 (FCGR1A), CD6 and CD24 where probe set variability indicates that IHC is a better method for analysis at this time. The use of updated probe set definitions slightly increased correlations for all cell types indicating that inaccurate probe sets may account for some of the disparities found between the two techniques in this study. For studies involving the prostate, suggested targets for CD profiling using the combination of DNA array and IHC technologies are listed in Table 4.

Many cross-platform studies have been conducted on DNA microarrays. These previous studies have discovered several factors that effect data comparison across different microarray platforms. In addition to our findings, updated annotation through RefSeq or UniGene databases has repeatedly been shown to increase cross-platform content and consistency [27-31]. Biological variability, and inter-laboratory variation have also been previously implicated as sources of cross-platform data

**Table 4: Suggested CD profiling targets with good agreement between data types for prostate characterizations.**

Cell type	CD antigens 'present'	CD antigens 'absent'
CD31 Endothelial	31, 49f, 54, 58, 59, 73, 74, 90, 107a, 107b, 144	1a, 6, 10, 30, 32, 41, 43, 56, 66f, 70, 75s, 84, 91, 94, 179a, 226, 271
CD26 Luminal	9, 10, 13, 26, 38, 63, 75s, 98, 107a, 107b, 147, 164	1a, 6, 30, 32, 41, 43, 49c, 50, 56, 66f, 70, 74, 84, 88, 91, 94, 104, 153, 179a, 226, 271
CD49a Stromal	29, 47, 49a, 56, 59	1a, 6, 26, 30, 32, 41, 43, 50, 64, 66f, 70, 75s, 84, 91, 94, 104, 153, 179a, 271
CD104 Basal	9, 44, 47, 49b, 49f, 59, 74, 92, 95, 104, 107b, 138, 147, 271	1a, 6, 30, 32, 41, 43, 49c, 50, 56, 64, 66f, 70, 75s, 84, 91, 94, 153, 179a, 226
ABC2 Progenitor	9, 44, 47, 49b, 49f, 59, 74, 92, 95, 138, 147	1a, 6, 30, 10, 13, 26, 30, 32, 41, 43, 49c, 50, 56, 64, 66f, 70, 84, 91, 94, 153, 179a, 226
CD26 Cancer	9, 24, 26, 29, 38, 40, 46, 49b, 63, 75s, 147	10, 13, 30, 32, 41, 43, 49c, 50, 56, 62L, 66f, 84, 88, 90, 91, 94, 104, 144, 153, 179a, 226, 271

'Present' denotes the list of CD antigens expressed by the specific cell type and detected as by both array and IHC. 'Absent' cites the list of CD antigens undetected by both array and IHC.



discrepancies [32-34]. For this study, although conducted by two different laboratories, the correlation between the two platform types (Affymetrix and Agilent) was only different for prostate stromal cells. Therefore, it is probable that these discrepancies were due to varying sampling methods (e.g. MACS-sorting of digested tissue *vs.* LCM of frozen tissue sections). Additionally, discrepancy could be due to the differing tissue histology of the prostate and therefore possible differing purity for stromal *vs.* luminal cell types. The prostate stroma is fairly heterogenous, consisting of smooth muscle cells and fibroblasts, with embedded blood vasculature, peripheral nerves and ganglia, and tissue infiltrating white blood cells; whereas the prostate luminal layer contains luminal secretory cells and rare neuroendocrine cells.

## Conclusion

In summary, CD molecules can be used to isolate cell populations from prostate tissue. These cell populations retain to a high degree their CD phenotype as determined by immunostaining in intact tissue. Observed discrepancies were mostly in expression levels detected by microarray and not immunohistochemistry; these differences are possibly due in part to varying sampling methods, problems related to probe and probe set identity as well as actual biological differences. It is hoped that continued improvements in cell sorting techniques and probe set development could further elucidate the relationship between mRNA levels and corresponding protein expression. In addition to comparison of immunohistochemistry and microarray data, there is further need for cross-comparison of results utilizing techniques such as *in situ* hybridization, western blotting and high throughput proteome analysis. Future studies analyzing mRNA and protein correlation will be required to determine the potential bias of various experimental methodologies as well as how mRNA expression differences are related to protein expression.

## Methods

### Prostate tissue

The methods of tissue collection, immunostaining and expression data collection and analysis used in this study have been published previously [13,20,35]. Briefly, tissue samples for IHC and Affymetrix array analysis consisted of both cancer-enriched and cancer-free samples obtained from 55 radical prostatectomies. Upon receipt of prostatectomy specimens, 3-mm thick transverse sections were made of the prostate after inking the exterior surface (the surgical margin). The same approach was used for both cancer-free and cancer-enriched (where at least 85% of the cells in the corresponding frozen section were cancer cells) samples. Between 4 and 6 blocks of tissue from the posterior aspect (the peripheral zone) of each prostate were frozen. Sections from the frozen blocks of tissue pro-

vided a histological template for characterizing the cell composition of the adjacent non-frozen tissue blocks. Cancer-free samples, weighing between 2 and 10 g, were harvested primarily from the anterior aspect of the prostate (transition zone) as described [12,13,35]. Cancer-enriched samples, weighing at least 100 mg, were dissected from the opposing aspect of the non-fixed section adjacent to the block of tissue that had been frozen. Tissue specimens for cell sorting were digested by collagenase [36,37].

As an external control, IHC and array data from MACS-sorted cells were compared to LCM cells from prostate stromal and luminal cells analyzed by hybridization to long-oligonucleotide microarrays (Agilent Corporation). Additionally, LCM stromal cells were analyzed by Affymetrix array to control for potential platform variation. Tissue specimens for LCM were obtained from prostate needle biopsies from 14 men identified from a biospecimen repository with no evidence of cancer as well as a second or third set of biopsies that also showed no evidence of carcinoma. Each biopsy specimen was immediately embedded in OCT and flash frozen in isopentane immersed in liquid nitrogen. All protocols for tissue acquisition and processing were under approval by the University of Washington Institutional Review Board following a standard protocol.

### Immunohistochemistry

Blocks of unfixed prostate tissue were harvested at surgery. Serial 5- $\mu$ m sections were prepared from randomly selected frozen blocks, fixed in cold acetone, and processed for immunohistochemistry. We used frozen sections since a majority of the anti-CD antibodies we used do not immunoreact with antigens in formalin-fixed, paraffin-embedded tissue. Immunohistochemistry was performed as described previously [20]. The primary antibodies used were mouse monoclonal CD antibodies (BD-PharMingen, San Diego, CA) diluted to 8 ng/ $\mu$ l or less. Antigen was localized using biotinylated anti-mouse IgG or IgM (Vector Labs, Burlingame, CA) as the secondary antibody and diaminobenzidine tetrahydrochloride as the chromogen. The sections were counterstained with hematoxylin.

The percentage of cells of a specific histological phenotype that expressed the antigen was estimated in five randomly selected fields at a final magnification of 40 $\times$  (ocular 10 $\times$ ; objective 4 $\times$ ). Staining intensity was evaluated by two parameters (staining intensity and percentage of cells exhibiting each level of intensity). Intensity of reaction product was based on a 3-point scale – none, faint/equivocal and intense. These categories of staining intensity were defined as follows. *Intense*: immunoreaction deposit is distinctly more optically dense than background and than tissue that does not express the antigen. *Equivocal*:

immunoreaction deposit is either similar enough in optical density to the background and/or to tissue that does not express the antigen, or is so focal, i.e. < 5% of cells, that there is reasonable uncertainty regarding whether the cells express the antigen. *None*: there is either no immunoreaction deposit or reaction product is no more optically dense than background.

A single value *A* was calculated for each immunostain by cell type using the following formula:

$$A = \frac{0(\% \text{ no stain}) + WF(\% \text{ faint/equivocal stain}) + 1(\% \text{ convincing stain})}{100}$$

WF = weighting factor, usually 0.2

The value of the weighting factor can be changed on our web site by independent observers of the images. We did this since we thought that observers of the images should be able to alter the value assigned to the weighting factor based on whether they thought that faint/equivocal immunostained cells either did or did not truly represent immunolocalization of each respective CD antigen. For the present study we used a weighting factor of 0.2.

An average score for each cell type for each CD antigen was calculated using the following equation:

$$\text{Score} = \frac{1}{n} \sum_{i=1}^n A_i$$

where:

*n* = number of immunostains for each CD by cell type

*A* = staining value for each immunostain by cell type

Average scores in the range of 0.751 to 1.0 were categorized as *intense*, 0.251 to 0.750 as *equivocal* or 0 to 0.250 as *none*. Immunostained sections were imaged with an Olympus BX41 microscope (Olympus, Melville, NY) equipped with a MicroFire digital camera (Optronics, Goleta, CA). Composite images were constructed with Photoshop CS (Adobe Systems, San Jose, CA) and are publicly available online [38].

#### **Affymetrix gene expression profiling of MACS-sorted cell populations and laser captured stromal cells**

Transcriptomes of the following prostate cell types have been previously determined: CD104<sup>+</sup> basal epithelial cells, CD26<sup>+</sup> luminal epithelial cells, CD49a<sup>+</sup> stromal fibromuscular cells, and CD31<sup>+</sup> endothelial cells [13], ABCG2<sup>+</sup> stem cells [39] and CD26<sup>+</sup> cancer cells [in preparation]. The CD molecules cited above were the ones used for sorting each specific cell type for gene expression profiling. The expression of genes documented in the litera-

ture as cell-type specific was compared to Affymetrix datasets; analyses showed that the transcriptomes were representative of their respective cell type (*p* < 0.05) [13]. Five separate biological replicates of each MACS-sorted cell populations (except the CD26<sup>+</sup> cancer cells which comprised of one biological sample) and three separate biological replicates of laser captured stromal cells were assayed to produce a dataset using the Human Genome U133 Plus 2.0 GeneChips (Affymetrix, Santa Clara, CA). The U-133 Plus 2.0 array contains probesets representing 54,675 genes, splice variants, and ESTs. The GeneChips were prepared, hybridized, and scanned according to the protocols provided by Affymetrix. Briefly, 200 ng of RNA was reverse transcribed with poly (dT) primer containing a T7 promoter, and the cDNA was made double-stranded. In vitro transcription was performed to produce unlabeled cRNA. Next, first-strand cDNA was produced with random primer. cDNA was made double-stranded with poly (dT) primer/T7 promoter. Finally, in vitro transcription was performed with biotinylated ribonucleotides. The biotin-labeled cRNA was hybridized to the GeneChips. The chips were washed and stained with streptavidin-PE using an Affymetrix FS-450 fluidics station. Data was collected with an Affymetrix GeneChip Scanner 3000.

#### **LCM and RNA isolation of luminal and stromal cells for Agilent array analysis**

Approximately two thousand luminal epithelial or stromal cells from histologically benign glands were captured using the Veritas Microdissection System (Molecular Devices Corporation, Sunnyvale, CA) as previously described [40]. Frozen sections (8 μm) were cut from tissue embedded in OCT (Miles, Inc. Diagnostic Division, Elkhart, IN) and fixed in cold 95% ethanol. After brief (5–10 seconds) staining with hematoxylin using the Histogene staining solution (Arcturus Engineering, Mountain View, CA), the sections were dehydrated in 100% ethanol, followed by sequential ethanol-xylene solutions. Three stromal cell samples and 14 luminal cell samples were captured. Digital photographs were taken of tissue sections before, during, and after LCM and assessed independently by two investigators to confirm that the laser-captured cells were stromal and luminal cells, respectively.

Captured individual specimens were lysed and RNA was isolated using Arcturus PicoPure RNA Isolation. Samples were treated with RNase-free DNase (Qiagen, Inc., Valencia, CA). This RNA was converted to a cDNA library and amplified using the TransPlex WTA (Rubicon Genomics, Ann Arbor, MI) using Titanium Taq polymerase (Clontech Laboratories, Inc., Mountain View, CA) in the presence of amino-allyl deoxyuridine triphosphate (dUTP) for postamplification labeling. Real-time PCR amplifications were terminated at plateau phase, as measured by fluores-

cence incorporation, to preserve maximum representation. Concentration and purity was determined using the ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE). To provide a reference standard RNA for use on 2-color microarrays, we isolated total RNA from LNCaP, DU145, PC3, and CWR22 cell lines (American Type Culture Collection, Manassas, VA) growing at log phase in dye-free RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS; Life Technologies, Rockville, MD). RNA was purified using Trizol (Life Technologies). The RNA was then further purified by Qiagen RNeasy maxi and treated with Qiagen RNase-free DNase. This total RNA was then amplified using the TransPlex WTA similar to the experimental samples.

#### **Gene expression profiling on Agilent microarrays**

Two µg of WTA-amplified amino-allyl cDNA from each of 14 luminal and 3 stromal samples were labeled with Cy3 fluorescent dye (cell line reference cDNA was labeled with Cy5) as described previously [40], and hybridized to Agilent 44 K whole human genome expression oligonucleotide microarray slides (Agilent Technologies, Inc., Santa Clara, CA) following the manufacturer's suggested protocols. Fluorescence array images were collected for both Cy3 and Cy5 using an Agilent fluorescent scanner, and Feature Extraction software (Agilent Technologies, Santa Clara, CA, USA) was used to grid, extract image intensities, and normalize data. Spots of poor quality, as determined by the software, were removed from further analysis. Data from the 14 luminal or 3 stromal samples were combined and the average signal values were used for comparative analyses. For genes with multiple probes on the array, the data from the probe with the highest sequence specificity was chosen.

#### **Data comparison and statistical analysis**

Immunohistochemistry data was compared to Affymetrix array data for the following prostate cell types: MACS-sorted basal cells, luminal cells, stromal fibromuscular cells, endothelial cells, stem cells and cancer cells from a primary tumor of Gleason 3+3 = 6. Additionally IHC data was compared to Affymetrix array data for LCM stromal and Agilent array data for LCM stromal and epithelial cells. Data was analyzed for concordance and correlation of expression levels. Data concordance was defined as agreement between data types based on 'present' or 'absent' calls, i.e., if both techniques detected a gene, they were considered to be in agreement. Genes in the Affymetrix array dataset with an average raw fluorescence signal < 50 were considered to be undetected by the experiment and were classified as 'absent' by GCOS version 1.0 software (Affymetrix) [13]. Genes were also classified as undetected if the majority (> 50%) of individual probe sets were classified as 'absent'. All other hybridization patterns were classified as 'present'. Genes with signal 51–500 were

considered to have 'low' expression, those with signal 501–5000 were considered to have 'medium' expression, and those with signal > 5000 were considered to have 'high' expression. Genes in the Agilent array dataset with an average raw fluorescence signal < 100 were considered to be 'absent', those with signal 100–999 were considered to have 'low' expression, those with signal 1000–9999 were considered to have 'medium' expression, and those with signal > 9999 were considered to have 'high' expression. Immunohistochemistry data having a staining level of *intense* (score range 0.751 – 1.0) or *equivocal* (score range 0.251 – 0.750) were classified as in agreement with genes classified as 'present'. Immunostaining results of *none* (score range 0 – 0.250) were considered equivalent to 'absent'.

The correlation between mRNA expression levels and immunostaining intensity for each CD antibody was described by Pearson product-moment and Spearman nonparametric correlation measures and tests. Correlations were calculated for mean and median values. Median was included as a more robust measurement in the presence of outlier or skewed values especially for the array data. Individual Affymetrix probes within a probe set were originally designated to hybridize with the same unique mRNA transcript. However, increasing knowledge of genomic sequences has shown that a substantial number of the manufacturer's original probe groupings and mappings are inaccurate and must be corrected [41]. To address the possibility that disparity in expression level was due to probe set definition, correlation was additionally determined using updated custom CDF definitions [19]. Dai et al., have reorganized probe set definitions based on the latest genome and transcriptome information [17]. These updated definitions based on Version 8 Entrez Gene mappings were downloaded and used in addition to the Affymetrix probe set definitions. To control for possible gene expression changes induced by the cell sorting process, LCM data for stromal cells were also compared for correlation. Additionally, LCM data for luminal and stromal cells assayed by Agilent array analysis was also used for correlation calculations. Correlations were calculated using both original and updated probe set definitions and were categorized as follows: no positive correlation (0.00 to 0.30), moderate positive correlation (0.31 to 0.79), and strong positive correlation (0.80 to 0.99).

#### **Authors' contributions**

LEP participated in study design, experiments, and drafting of the manuscript. Data analysis and database design and programming were performed by DSC with input from EWD. MR, LJE, IMC and PSN carried out LCM experiments and data analysis. AYL conceived of study and

assisted in study design and coordination. All authors have read and approved the final manuscript.

## Additional material

### Additional file 1

Scatter plot of IHC staining intensity and gene expression levels determined by Affymetrix array for MACS-sorted cell populations. Statistical Pearson correlation ranged from -0.01 – 0.55 and Spearman Coefficients ranged from 0.00 – 0.51. There was no positive overall correlation of immunolocalization data with array data for endothelial, stromal or progenitor cells, and moderate positive correlation for luminal, basal and cancer.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-246-S1.pdf>]

### Additional file 2

Scatter plot using updated probe set definitions. Statistical Pearson correlations ranged from 0.09 – 0.57 and Spearman Coefficients ranged from 0.03 – 0.63 using updated probe set definitions. Updated probe set definitions increased correlation from none to moderate positive correlation for progenitor cells.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-246-S2.pdf>]

### Additional file 3

Scatter plot of IHC staining intensity and gene expression levels determined by Affymetrix and Agilent array for LCM cells. Statistical Pearson correlation ranged from 0.26 – 0.57 and Spearman Coefficients ranged from 0.25 – 0.43. There was no positive correlation of immunolocalization data with array data for LCM stromal cells assayed by either Agilent or Affymetrix. Pearson correlation for LCM luminal cells was 0.57 and Spearman coefficient was 0.43, which were the same values as that for MACS-sorted luminal cells using updated probe set definitions. Pearson correlation between LCM stromal cell data and MACS-sorted stromal cell data assayed by Affymetrix array was 0.68, Spearman coefficient was 0.78. Pearson correlation between LCM stromal cell data analyzed by Agilent array and Affymetrix MACS-sorted stromal cell data was 0.65, Spearman coefficient was 0.55. Pearson correlation between LCM stromal cell data analyzed by Agilent array and Affymetrix LCM stromal cell data was 0.33, Spearman coefficient was 0.59.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-246-S3.pdf>]

## Acknowledgements

We thank Tracy Sherertz and Martin Korb for their contribution to the database construction. We thank Paul Shannon for assistance in statistical analysis.

This work was supported by grant DK63630 and DK069690 from NIDDK. Additional funding came from grants CA85859, CA98699 and CA111244 from NCI, and PM50 GMO76547/Center for Systems Biology.

## References

- Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* 1999, **19(3)**:1720-1730.
- Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardias SL, Giordano TJ, Iannettoni MD, Orringer MB, Hanash SM, Beer DG: **Discordant protein and mRNA expression in lung adenocarcinomas.** *Mol Cell Proteomics* 2002, **1(4)**:304-313.
- Lichtinghagen R, Musholt PB, Lein M, Romer A, Rudolph B, Kristiansen G, Hauptmann S, Schnorr D, Loening SA, Jung K: **Different mRNA and protein expression of matrix metalloproteinases 2 and 9 and tissue inhibitor of metalloproteinases 1 in benign and malignant prostate tissue.** *Eur Urol* 2002, **42(4)**:398-406.
- Schlemmer BO, Sorensen BS, Overgaard J, Olsen KE, Gjerdrum LM, Nexø E: **Quantitative PCR--new diagnostic tool for quantifying specific mRNA and DNA molecules: HER2/neu DNA quantification with LightCycler real-time PCR in comparison with immunohistochemistry and fluorescence in situ hybridization.** *Scand J Clin Lab Invest* 2004, **64(5)**:511-522.
- Waghray A, Feroze F, Schober MS, Yao F, Wood C, Puravs E, Krause M, Hanash S, Chen YQ: **Identification of androgen-regulated genes in the prostate cancer cell line LNCaP by serial analysis of gene expression and proteomic analysis.** *Proteomics* 2001, **1(10)**:1327-1338.
- Lin B, White JT, Lu W, Xie T, Utleg AG, Yan X, Yi EC, Shannon P, Khrebtkova I, Lange PH, Goodlett DR, Zhou D, Vasicek TJ, Hood L: **Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease.** *Cancer Res* 2005, **65(8)**:3081-3091.
- Orntoft TF, Thykjaer T, Waldman FM, Wolf H, Celis JE: **Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas.** *Mol Cell Proteomics* 2002, **1(1)**:37-45.
- Schostak M, Krause H, Miller K, Schrader M, Weikert S, Christoph F, Kempkensteffen C, Kolleremann J: **Quantitative real-time RT-PCR of CD24 mRNA in the detection of prostate cancer.** *BMC Urol* 2006, **6**:7.
- Liu AY, Roudier MP, True LD: **Heterogeneity in primary and metastatic prostate cancer as defined by cell surface CD profile.** *Am J Pathol* 2004, **165(5)**:1543-1556.
- Kristiansen G, Pilarsky C, Wissmann C, Kaiser S, Bruemendorf T, Roepcke S, Dahl E, Hinzmann B, Specht T, Pervan J, Stephan C, Loening S, Dietel M, Rosenthal A: **Expression profiling of microdissected matched prostate cancer samples reveals CD166/MEMD and CD24 as new prognostic markers for patient survival.** *J Pathol* 2005, **205(3)**:359-376.
- Kristiansen G, Pilarsky C, Pervan J, Sturzebecher B, Stephan C, Jung K, Loening S, Rosenthal A, Dietel M: **CD24 expression is a significant predictor of PSA relapse and poor prognosis in low grade or organ confined prostate cancer.** *Prostate* 2004, **58(2)**:183-192.
- Liu AY, Peehl DM: **Characterization of cultured human prostatic epithelial cells by cluster designation antigen expression.** *Cell Tissue Res* 2001, **305(3)**:389-397.
- Oudes AJ, Campbell DS, Sorensen CM, Wlashek LS, True LD, Liu AY: **Transcriptomes of human prostate cells.** *BMC Genomics* 2006, **7**:92.
- Pascal LE, Deutsch EW, Campbell DS, Korb M, True LD, Liu AY: **The urologic epithelial stem cell database (UESC) - a web tool for cell type-specific gene expression and immunohistochemistry images of the prostate and bladder.** *BMC Urol* 2007, **7**:19.
- Deutsch EW, Ball CA, Berman JJ, Bova GS, Brazma A, Bumgarner RE, Campbell D, Causton HC, Christiansen JH, Daian F, Dauga D, Davidson D, Gimenez G, Goo YA, Grimmond S, Henrich T, Gerrmann BG, Johnson MH, Korb M, Mills JC, Oudes AJ, Parkinson HE, Pascal LE, Pollet N, Quackenbush J, Ramialison M, Ringwald M, Salgado D, Sansone SA, Sherlock G, Christian J, Stoeckert J, Swedlow J, Taylor RC, Wlashek L, Warford A, Wilkinson DG, Zhou Y, Zou LI, Liu AY, True LD: **Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments (MIS-FISHIE).** *Nature Biotechnology* in press.
- Dall P, Heider KH, Sinn HP, Skroch-Angel P, Adolf G, Kaufmann M, Herrlich P, Ponta H: **Comparison of immunohistochemistry and RT-PCR for detection of CD44v-expression, a new prognostic factor in human breast cancer.** *Int J Cancer* 1995, **60(4)**:471-477.
- Yu H, Wang F, Tu K, Xie L, Li YY, Li YX: **Transcript-level annotation of Affymetrix probesets improves the interpretation of gene expression data.** *BMC Bioinformatics* 2007, **8**:194.

18. Jackson DG, Buckley J, Bell JI: **Multiple variants of the human lymphocyte homing receptor CD44 generated by insertions at a single site in the extracellular domain.** *J Biol Chem* 1992, **267(7)**:4732-4739.
19. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F: **Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data.** *Nucleic Acids Res* 2005, **33(20)**:e175.
20. Liu AY, True LD: **Characterization of prostate cell types by CD cell surface molecules.** *Am J Pathol* 2002, **160(1)**:37-43.
21. Oudes AJ, Roach JC, Walashek LS, Eichner LJ, True LD, Vessella RL, Liu AY: **Application of Affymetrix array and Massively Parallel Signature Sequencing for identification of genes involved in prostate cancer progression.** *BMC Cancer* 2005, **5**:86.
22. Greenbaum D, Colangelo C, Williams K, Gerstein M: **Comparing protein abundance and mRNA expression levels on a genomic scale.** *Genome Biol* 2003, **4(9)**:117.
23. Nie L, Wu G, Zhang W: **Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: a quantitative analysis.** *Genetics* 2006, **174(4)**:2229-2243.
24. Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, Goodlett D, Berger JP, Gunter B, Linsley PS, Stoughton RB, Aebersold R, Collins SJ, Hanlon WA, Hood LE: **Integrated genomic and proteomic analyses of gene expression in Mammalian cells.** *Mol Cell Proteomics* 2004, **3(10)**:960-969.
25. Nie L, Wu G, Zhang W: **Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations.** *Biochem Biophys Res Commun* 2006, **339(2)**:603-610.
26. Pollard K, Lunny D, Holgate CS, Jackson P, Bird CC: **Fixation, processing, and immunochemical reagent effects on preservation of T-lymphocyte surface membrane antigens in paraffin-embedded tissue.** *J Histochem Cytochem* 1987, **35(11)**:1329-1338.
27. Hessner MJ, Xiang B, Jia S, Geoffrey R, Holmes S, Meyer L, Muheisen S, Wang X: **Three-color cDNA microarrays with prehybridization quality control yield gene expression data comparable to that of commercial platforms.** *Physiol Genomics* 2006, **25(1)**:166-178.
28. Mah N, Thelin A, Lu T, Nikolaus S, Kuhbacher T, Gurbuz Y, Eickhoff H, Kloppel G, Lehrach H, Mellgard B, Costello CM, Schreiber S: **A comparison of oligonucleotide and cDNA-based microarray systems.** *Physiol Genomics* 2004, **16(3)**:361-370.
29. Mecham BH, Klus GT, Strovel J, Augustus M, Byrne D, Bozso P, Wetmore DZ, Mariani TJ, Kohane IS, Szallasi Z: **Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements.** *Nucleic Acids Res* 2004, **32(9)**:e74.
30. Mecham BH, Wetmore DZ, Szallasi Z, Sadovsky Y, Kohane I, Mariani TJ: **Increased measurement accuracy for sequence-verified microarray probes.** *Physiol Genomics* 2004, **18(3)**:308-315.
31. Tan PK, Downey TJ, Spitznagel EL Jr., Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC: **Evaluation of gene expression measurements from commercial microarray platforms.** *Nucleic Acids Res* 2003, **31(19)**:5676-5684.
32. Pritchard CC, Hsu L, Delrow J, Nelson PS: **Project normal: defining normal variance in mouse gene expression.** *Proc Natl Acad Sci U S A* 2001, **98(23)**:13266-13271.
33. Bammler T, Beyer RP, Bhattacharya S, Boorman GA, Boyles A, Bradford BU, Bumgarner RE, Bushel PR, Chaturvedi K, Choi D, Cunningham ML, Deng S, Dressman HK, Fannin RD, Farin FM, Freedman JH, Fry RC, Harper A, Humble MC, Hurban P, Kavanagh TJ, Kaufmann WK, Kerr KF, Jing L, Lapidus JA, Lasarev MR, Li J, Li YJ, Lobenhofer EK, Lu X, Malek RL, Milton S, Nagalla SR, O'Malley J P, Palmer VS, Pattee P, Paules RS, Perou CM, Phillips K, Qin LX, Qiu Y, Quigley SD, Rodland M, Rusyn I, Samson LD, Schwartz DA, Shi Y, Shin JL, Sieber SO, Slifer S, Speer MC, Spencer PS, Sproles DI, Swenberg JA, Suk WA, Sullivan RC, Tian R, Tennant RW, Todd SA, Tucker CJ, Van Houten B, Weis BK, Xuan S, Zarbl H: **Standardizing global gene expression analysis between laboratories and across platforms.** *Nat Methods* 2005, **2(5)**:351-356.
34. Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martinez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye SQ, Yu W: **Multiple-laboratory comparison of microarray platforms.** *Nat Methods* 2005, **2(5)**:345-350.
35. Pascal LE, Oudes AJ, Petersen TW, Goo YA, Walashek LS, True LD, Liu AY: **Molecular and cellular characterization of ABCG2 in the prostate.** *BMC Urol* 2007, **7(1)**:6.
36. Liu AY, True LD, LaTray L, Nelson PS, Ellis WJ, Vessella RL, Lange PH, Hood L, van den Engh G: **Cell-cell interaction in prostate gene regulation and cytodifferentiation.** *Proc Natl Acad Sci U S A* 1997, **94(20)**:10705-10710.
37. Kassen A, Sutkowski DM, Ahn H, Sensibar JA, Kozlowski JM, Lee C: **Stromal cells of the human prostate: initial isolation and characterization.** *Prostate* 1996, **28(2)**:89-97.
38. Pascal LE, Deutsch EW, Campbell DS, Korb M, True LD, Liu AY: **The urologic epithelial stem cell database (UESC) - a web tool for cell type-specific gene expression and immunohistochemistry images of the prostate and bladder.** *BMC Urology* in press.
39. Goo YA, Goodlett DR, Pascal LE, Worthington KD, Vessella RL, True LD, Liu AY: **Stromal mesenchyme cell genes of the human prostate and bladder.** *BMC Urol* 2005, **5**:17.
40. True L, Coleman I, Hawley S, Huang CY, Gifford D, Coleman R, Beer TM, Gelmann E, Datta M, Mostaghel E, Knudsen B, Lange P, Vessella R, Lin D, Hood L, Nelson PS: **A molecular correlate to the Gleason grading system for prostate adenocarcinoma.** *Proc Natl Acad Sci U S A* 2006, **103(29)**:10991-10996.
41. Liu H, Zeeberg BR, Qu G, Koru AG, Ferrucci A, Kahn A, Ryan MC, Nuhanovic A, Munson PJ, Reinhold WC, Kane DW, Weinstein JN: **AffyProbeMiner: a web resource for computing or retrieving accurately redefined Affymetrix probe sets.** *Bioinformatics* 2007, **23(18)**:2385-2390.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
http://www.biomedcentral.com/info/publishing\_adv.asp

