Identification of S100A8-correlated genes for prediction of disease progression in non-muscle invasive bladder cancer

Seon-Kyu Kim (sk7944@cbnu.ac.kr)
Eun-Jung Kim (gstej@naver.com)
Sun-Hee Leem (shleem@dau.ac.kr)
Yun-Sok Ha (yunsokha@gmail.com)
Yong-June Kim (urokyj@cbnu.ac.kr)
Wun-Jae Kim (wjkim@chungbuk.ac.kr)

Identification of *S100A8*-correlated genes for prediction of disease progression in non-muscle invasive bladder cancer

Seon-Kyu Kim[1], Eun-Jung Kim[1,2], Sun-Hee Leem[3], Yun-Sok Ha[1], Yong-June Kim[1,2], and Wun-Jae Kim*[1,2]

[1]Department of Urology, College of Medicine, Chungbuk National University, Cheongju, Chungbuk, South Korea

[2]BK21 Chungbuk Biomedical Science Center, School of Medicine, Chungbuk National University, Chengju, Chungbuk, South Korea

[3]Department of Biological Science, Dong-A University, Busan, South Korea

*Corresponding author

Email addresses:

        SKK: sk7944@cbnu.ac.kr

        EJK: gstej@naver.com

        SHL: shleem@dau.ac.kr

        YSH: yunsokha@gmail.com

        YJK: urokyj@cbnu.ac.kr

        WJK: wjkim@chungbuk.ac.kr

Abstract

Background

S100 calcium binding protein A8 (*S100A8*) has been implicated as a prognostic indicator in several types of cancer. However, previous studies are limited in their ability to predict the clinical behavior of the cancer. Here, we sought to identify a molecular signature based on *S100A8* expression and to assess its usefulness as a prognostic indicator of disease progression in non-muscle invasive bladder cancer (NMIBC).

Methods

We used 103 primary NMIBC specimens for microarray gene expression profiling. The median follow-up period for all patients was 57.6 months (range: 3.2 to 137.0 months). Various statistical methods, including the leave-one-out cross validation method, were applied to identify a gene expression signature able to predict the likelihood of progression. The prognostic value of the gene expression signature was validated in an independent cohort (n=302).

Results

Kaplan-Meier estimates revealed significant differences in disease progression associated with the expression signature of *S100A8*-correlated genes (log-rank test, $P < 0.001$). Multivariate Cox regression analysis revealed that the expression signature of *S100A8*-correlated genes was a strong predictor of disease progression (hazard ratio = 15.225, 95% confidence interval = 1.746 to 133.52, $P = 0.014$). We validated our results in an independent cohort and confirmed that this signature produced consistent prediction patterns. Finally, gene network analyses of the signature revealed that *S100A8*, *IL1B*, and *S100A9* could be important mediators of the progression of NMIBC.

Conclusions

The prognostic molecular signature defined by *S100A8*-correlated genes represents a

promising diagnostic tool for the identification of NMIBC patients that have a high

risk of progression to muscle invasive bladder cancer.

Background

Non-muscle invasive bladder cancer (NMIBC) is the most common histological subtype of bladder cancer, accounting for approximately 80% of all cases. Approximately 20% of these patients experience disease progression to muscle invasive bladder cancer (MIBC) after treatment, a development that is associated with a very poor prognosis for survival. Conventional histopathological parameters, such as tumor stage or grade, are generally considered to be prognostic factors, and numerous biomarkers have been investigated as prognostic indicators of the likelihood that NMIBC will develop into MIBC [1-6].

Members of the *S100* family of calcium-binding proteins play essential roles in epithelial tissues and participate in a wide range of cellular processes, including transcription, proliferation, and differentiation [7-11]. At least 16 genes that encode members of the *S100* family, including the gene for *S100A8*, are clustered on human chromosome 1q21 [12, 13], in a region that frequently experiences chromosomal rearrangement during tumor development [14, 15]. *S100A8* is reportedly up-regulated in many cancers, including bladder cancer [16-23], and has been implicated in the regulation of tumor cell proliferation and metastasis [16, 24-26].

Although numerous diagnostic markers have been investigated as indicators of the risk of disease progression, none are able to sufficiently predict the behavior of NMIBC [1-6]. *S100A8* has been suggested to be a predictive biomarker of bladder cancer outcome in several studies [21-23], however, the regulation of *S100A8* gene expression and whether the genes associated with its expression provides additional insight into the mechanisms of disease progression or tumor invasion have not been studied. Therefore, we analyzed the expression pattern of *S100A8* and its correlated

genes to assess whether their molecular signature could identify patients with a higher

likelihood of disease progression.

Methods

Patients and tissue samples

Primary NMIBC tissue samples from 103 consecutive cases of patients with histologically diagnosed transitional cell carcinoma were obtained from Chungbuk National University Hospital. To reduce confounding factors for affecting the analyses, any patients diagnosed with a concomitant carcinoma in situ (CIS) lesion or only CIS lesion were excluded. All tumors were macro-dissected, typically within 15 minutes of surgical resection. Each bladder cancer specimen was confirmed as representative by analysis of adjacent tissue in fresh frozen sections from transurethral resection (TUR) specimens, and then frozen in liquid nitrogen and stored at –80 °C until use. The collection and analysis of all samples was approved by the Institutional Review Board of Chungbuk National University, and informed consent was obtained from each subject.

Tumors were staged and graded according to the 2002 TNM classification and the 2004 WHO grading system, respectively [27]. A second TUR was performed 2–4 week after the initial resection when it was incomplete or when a high-grade or T1 tumor was detected [27]. Patients with intermediate- or high-risk NMIBC received one cycle of intravesical BCG [27, 28]. All patients were followed and managed according to the standard recommendation for treatment of NMIBC [27-29]. In this study, we defined progression of the disease as an increase in stage from either Ta or T1 to T2 or higher after disease relapse [30].

RNA extraction, microarray experiments, and data processing

Total RNA was isolated by TRIzol reagent (Life Technologies, NY), according to the manufacturer's protocol. The quality and integrity of the RNA were confirmed by agarose gel electrophoresis and ethidium bromide staining, followed by

visual examination under ultraviolet light. Five-hundred nanograms of total RNA

were used for labeling hybridization according to the manufacturer's protocols

(Illumina HumanWG-6 BeadChip, version 2). Arrays were scanned with an Illumina

Bead Array Reader confocal scanner (BeadStation 500GXDW; Illumina, Inc., San

Diego, CA) according to the manufacturer's instructions. After scanning, the

microarray data were normalized using quantile normalization in the R language

environment (version 2.8.1, available at http://www.r-project.org/). Measured gene

expression values were log2 transformed and median centered across genes and

samples. The full microarray data set is available in the NCBI Gene Expression

Omnibus public database under the data series accession number GSE13507.

Statistical analysis

To classify patients into two groups, we used the median gene expression

value of *S100A8* as the cut-off. Pearson correlation coefficients were calculated to

evaluate the association between *S100A8* and its correlated genes. A hierarchical

clustering algorithm, using the uncentered correlation coefficient as the measure of

similarity and average linkage clustering, was applied as described in Eisen et al [31].

The Kaplan-Meier method was used to calculate the time to progression, and

differences between the times was assessed using log rank statistics. The prognostic

value of the *S100A8*-correlated gene signature was determined with multivariate Cox

proportional hazard regression models.

To validate the expression signature of *S100A8* and its correlated genes

identified in our original cohort, we applied it to independent microarray data from

302 patients with NMIBC, reported by Dyrskjot et al [32]. For validation of the

prognostic value of the molecular signature, we developed prediction models using

the compound covariate predictor [33], Bayesian compound covariate predictor [33],

linear discriminator analysis [34], nearest centroid classification [34], and support

vector machines [35]. The models incorporated genes that were differentially

expressed between the two classes using a two-sample t-test. Genes were considered

to have statistically significant differences in expression if the *P*-value was less than

0.001. We estimated the prediction error of each model using leave-one-out cross-

validation (LOOCV), as described by Simon et al [36]. For each LOOCV training set,

the entire model-building procedure was repeated, including the gene selection

process. Validation procedure was performed in BRB ArrayTools (version 3.7.1).

To explore the relationships between *S100A8*-correlated genes, we examined

functional associations among the genes and generated gene networks based on

whether they had more interconnected genes than would be expected to occur by

chance. The significance of each network was estimated using the scoring system

provided by the Ingenuity Pathway Analysis Tool (version 7.5). The scores were

determined by the number of differentially expressed genes within each of the

networks and the strength of the associations among the network members.

Results

Baseline characteristics

Table 1 details the baseline characteristics of the 103 primary NMIBC patients. The median age was 66.0 years (range: 24 to 88 years), and the median follow-up period after surgery was 57.6 months (range: 3.2 to 137.0 months). During the follow-up period, 11 of the 103 (10.7%) experienced disease progression.

Prognostic utility of *S100A8*-correlated genes

We analyzed 103 NMIBC samples and divided them into two groups according to the expression level of *S100A8*. The frequency of progression was significantly higher in the group with *S100A8* expression levels in the upper 50th percentile than in the group with *S100A8* expression levels in the lower 50th percentile (log-rank test, $P = 0.003$; Figure 1-A and 1-B).

We next sought to identify a gene expression signature that directly correlated with *S100A8* expression levels during disease progression and then use that signature to predict the likelihood of tumor progression. We identified 1,015 genes whose change in expression correlated with *S100A8* expression (Pearson correlation test, $P < 0.001$, $r < -0.3$ or $r > 0.3$). Based on hierarchical clustering analysis of the expression patterns of these genes, we divided the NMIBC samples into two groups: high *S100A8* cluster (HSC) and low *S100A8* cluster (LSC). The progression rate of HSC patients was significantly higher than that of the LSC patients (log-rank test, $P < 0.001$; Figure 2).

To evaluate the prognostic efficacy of the newly identified signature, we applied multivariate Cox regression analysis to the signature and known clinical and pathologic prognostic factors for NMIBC (Table 2). This analysis revealed that the molecular signature of *S100A8*-correlated genes (hazard ratio 15.225, 95% confidence

interval = 1.746 to 133.52, $P = 0.014$) was the only strong predicator of bladder cancer progression.

Validation of the signature in an independent cohort

We next sought to validate our findings by using gene expression data from an independent cohort of European patients with bladder cancer [32]. Patients with NMIBC from this cohort (n = 302) were divided into two groups according to the expression level of *S100A8*. Consistent with our results, progression of NMIBC was significantly higher in patients with *S100A8* expression levels in the upper 50th percentile than in those with *S100A8* expression levels in the lower 50th percentile (log-rank test, $P < 0.001$; Figure 1-C).

We also validated the association of the *S100A8*-correlated gene signature with disease progression in the European cohort. To overcome the peculiarities of any one particular prediction algorithm, we applied five different statistical methods to test the accuracy of our signature-based prediction of disease progression (Figure 3). We identified the genes with the greatest difference in expression level between the HSC and LSC subgroups in our original cohort (the training set). These genes were pooled to form a series of classifiers able to estimate the probability that a particular bladder cancer sample belonged to the HSC or LSC subgroup. The number of genes in the classifier set was optimized to minimize misclassification during LOOCV of the tumors in the training set. The performance of each prediction model is illustrated in Table 3. When applied to the European cohort (the test set), all five models produced consistent prediction patterns. Kaplan-Meier estimations in the test set revealed significant differences in the risk of progression between patients in subgroups HSC and LSC (Figure 3-B).

Biological insights into the signature for disease progression

To identify the predominant signalling networks active in the advancement from NMIBC to MIBC, gene network analysis of the 1,015 genes featured in the progression signature (Figure 2) was carried out using Ingenuity ^{TM} Pathways Analysis software. Of the 1,015 genes, 768 were mapped to gene networks defined by this tool. This analysis revealed a series of putative networks and associated functional categories. The 10 putative networks with highest scores are listed in Additional file 1 and their associated functions are illustrated in Figure 4.

As expected, genes involved in cellular growth and proliferation, cell cycle, cancer, and cell death were enriched, providing confidence in our results. We also found that genes involved in cell-mediated immune responses, immune cell trafficking, humoral immune responses, inflammatory responses, and immunological disease were significantly enriched (Figure 4).

Interestingly, among the top 10 putative networks, we observed functional connectivity between *S100A8* and *IL1B* in the third network, in which *S100A8* is a downstream effector of *IL1B* [37] (Figure 5). The expression level of *IL1B* was significantly higher in the HSC than in the LSC subgroup (two-sample t-test, $P < 0.001$; Figure 6-B), indicating its potential role in the progression of NMIBC to MIBC. We also found functional relationship between *S100A8* and *S100A9*, in which *S100A9* is a direct binding partner of *S100A8* (Figure 5). *S100A9* was more highly expressed in the HSC than in the LSC subgroup (two-sample t-test, $P < 0.001$; Figure 6-C).

Discussion

The data presented in this study demonstrates that the expression signature of *S100A8*-correlated genes is able to predict the likelihood of bladder cancer progression. The validity of this signature as a prognostic indicator was confirmed by analysis of 302 cancers from an independent European cohort. In addition, based on the results of gene network analysis, we have identified a putative mechanism that may be responsible for disease progression.

Although considerable effort has been devoted to the establishment of a prognostic model of NMIBC that can provide information concerning survival and treatment options at diagnosis [1-6], the ability to predict the course of disease progression for patients with bladder cancer remains a major clinical challenge. A number of the molecular markers that have been identified to date have been explored as potential predictors of disease progression. Some of these, such as *p53*, have been suggested to be independent markers, while others do not appear to be effective as prognostic indicators [4-6]. Thus, there is a crucial need for methods capable of identifying patients with NMIBC that is likely to develop into MIBC. In the present study, we developed a method to predict the progression of primary NMIBC based on a gene expression signature. We showed that our method has strong predictive value through multivariate regression analysis and a validation study in an independent cohort. These data underscore the effectiveness of this molecular signature as a prognostic indicator in NMIBC, and suggest that this signature could be clinically useful.

The accumulation of recent reports that document the deregulated expression of calcium-binding proteins of the *S100* family in a variety of human cancers implies that *S100* proteins are required for neoplastic or metastatic transformation of tumor

cells [26]. Previous observations that *S100A8* expression is enhanced in bladder

cancer suggest a role for this protein in the progression of bladder cancer [21-23]. The

data obtained from the current study demonstrated that upregulation of *S100A8*

expression was strongly associated with disease progression. This result substantiates

the involvement of *S100A8* during the progression of NMIBC.

Expression of *S100A8* is not the only indicator of *S100A8* activity, because it

is regulated by many different mechanisms. The identification of stable and reliable

human gene-to-gene relationships is an essential step towards unraveling the

interactions and functional correlations between human genes [38]. Therefore, we

performed gene network analysis to identify the association of *S100A8*-correlated

genes with progression of NMIBC. Strikingly, expression of both *S100A8* and its

correlated genes were strong predictors of the progression of bladder cancer (Figures

1 and 2). This finding was further supported by multivariate analysis, which showed

that the molecular signature of *S100A8*-correlated genes was a strong predictor of

cancer progression-free survival (Table 2). This result suggests that the *S100A8*-

correlated gene signature retains its prognostic relevance even after additional

pathological prognostic features have been taken into account.

The standard strategy for evaluating the accuracy of classification methods is

application of a training-validation approach, in which a training set is utilized to

identify the molecular signature and a validation set is used to estimate the degree of

reliability. For the validation assay, a large sample size is needed before expression

profiling can be utilized in a clinical setting [39]. In the current study, the accuracy of

a predictive *S100A8*-correlated gene signature was tested using five independent

algorithms with gene expression data from a large independent European cohort

(Figure 3). This result demonstrated not only a strong association between gene

expression patterns and progression, but also provided strong evidence of the reliability of the prediction.

Based on an analysis of the *S100A8*-correlated genes in the context of gene networks, we identified a putative role for *S100A8* in disease progression of bladder cancer. Because *S100A8* is the downstream target of *IL1B* [37], our results strongly support the potential involvement of *IL1B* in tumor progression. Recent studies indicate that *IL1B* is associated with tumor invasiveness and metastasis [40, 41]. Many of the satellite genes associated with *IL1B* (i.e., *CXCL2*, *PLA2G7*, *CCL11*, *S100A9*, *CD163*, *RNASE7*, and *OSM*) (Figure 5) participate in inflammatory and immune responses, which are the best known activities of *IL1B*. Previous studies have demonstrated that *IL1B* activates inflammation that promotes tumor invasiveness [40]. It was reported previously that *S100A9* is up-regulated in conjunction with *S100A8* in many cancers, including gastric cancer [20], prostate cancer [16, 17], and colorectal cancer [18, 19]. Both *S100A8* and *S100A9* have been implicated in the regulation of cell proliferation [16, 24] and metastatic processes [25]. Thus, *S100A9* may prove a valuable target for prevention of the migration of tumor cells to pre-metastatic sites [26].

Our results demonstrate that an expression signature consisting of *S100A8* and its correlated genes can be a reliable prognostic indicator of progression in NMIBC, independent of traditional pathologic prognostic parameters. The use of this signature as a predictive indicator could potentially enable the prognosis of heterogeneous NMIBC patients to be determined at diagnosis, which would allow for individualized treatment and evaluation.

Conclusions

We describe a prognostic molecular signature based on gene expression of *S100A8*-correlated genes that can identify patients with high-risk NMIBC that is likely to develop into MIBC. Identification of patients with high-risk NMIBC may improve the effectiveness of currently available treatments and provide opportunities for the development of new treatment modalities.

Abbreviations

NMIBC (non-muscle invasive bladder cancer); MIBC (muscle invasive bladder cancer); TUR (transurethral resection); LOOCV (leave-one-out cross validation); HSC (high *S100A8* cluster); LSC (low *S100A8* cluster)

Competing interests

None of the contributing authors have any conflicts of interest, including specific financial interests, relationships or affiliation relevant to the subject matter or materials discussed in the manuscript.

Authors' contributions

SKK conceived of the study, carried out the statistical analyses, and drafted the manuscript. EJK and SHL performed microarray preparation, sample selection, and RNA isolation. YSH investigated the clinical records of the considered patients and contributed to the interpretation of the results. YJK performed the statistical analyses, contributed to the interpretation of the results, and helped to draft the manuscript. WJK designed the study concept, interpreted the results and approved the final manuscript. All authors have read and approved the final manuscript.

Acknowledgements

References

1.   Pow-Sang JM, Seigne JD: **Contemporary management of superficial bladder cancer**. *Cancer Control* 2000, **7**(4):335-339.

2. Holmang S, Johansson SL: **Stage Ta-T1 bladder cancer: the relationship between findings at first followup cystoscopy and subsequent recurrence and progression**. *J Urol* 2002, **167**(4):1634-1637.

3. Herr HW: **Tumor progression and survival of patients with high grade, noninvasive papillary (TaG3) bladder tumors: 15-year outcome**. *J Urol* 2000, **163**(1):60-61; discussion 61-62.

4. Liou LS: **Urothelial cancer biomarkers for detection and surveillance**. *Urology* 2006, **67**(3 Suppl 1):25-33; discussion 33-24.

5. Cordon-Cardo C, Zhang ZF, Dalbagni G, Drobnjak M, Charytonowicz E, Hu SX, Xu HJ, Reuter VE, Benedict WF: **Cooperative effects of p53 and pRB alterations in primary superficial bladder tumors**. *Cancer Res* 1997, **57**(7):1217-1221.

6. Habuchi T, Marberger M, Droller MJ, Hemstreet GP, 3rd, Grossman HB, Schalken JA, Schmitz-Drager BJ, Murphy WM, Bono AV, Goebell P *et al*: **Prognostic markers for bladder cancer: International Consensus Panel on bladder tumor markers**. *Urology* 2005, **66**(6 Suppl 1):64-74.

7. Berridge MJ, Bootman MD, Roderick HL: **Calcium signalling: dynamics, homeostasis and remodelling**. *Nat Rev Mol Cell Biol* 2003, **4**(7):517-529.

8. Schafer BW, Heizmann CW: **The S100 family of EF-hand calcium-binding proteins: functions and pathology**. *Trends Biochem Sci* 1996, **21**(4):134-140.

9. Donato R: **S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles**. *Int J Biochem Cell Biol* 2001, **33**(7):637-668.

10. Donato R: **Intracellular and extracellular roles of S100 proteins**. *Microsc Res Tech* 2003, **60**(6):540-551.

11. Heizmann CW, Fritz G, Schafer BW: **S100 proteins: structure, functions and pathology**. *Front Biosci* 2002, **7**:d1356-1368.

12. Marenholz I, Heizmann CW, Fritz G: **S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature)**. *Biochem Biophys Res Commun* 2004, **322**(4):1111-1122.

13. Ravasi T, Hsu K, Goyette J, Schroder K, Yang Z, Rahimi F, Miranda LP, Alewood PF, Hume DA, Geczy C: **Probing the S100 protein family through genomic and functional analysis**. *Genomics* 2004, **84**(1):10-22.

14. Schafer BW, Wicki R, Engelkamp D, Mattei MG, Heizmann CW: **Isolation of a YAC clone covering a cluster of nine S100 genes on human chromosome 1q21: rationale for a new nomenclature of the S100 calcium-binding protein family**. *Genomics* 1995, **25**(3):638-643.

15. Schutte BC, Carpten JD, Forus A, Gregory SG, Horii A, White PS: **Report and abstracts of the sixth international workshop on human chromosome 1 mapping 2000. Iowa City, Iowa, USA. 30 September-3 October 2000**. *Cytogenet Cell Genet* 2001, **92**(1-2):23-41.

16. Hermani A, De Servi B, Medunjanin S, Tessier PA, Mayer D: **S100A8 and S100A9 activate MAP kinase and NF-kappaB signaling pathways and trigger translocation of RAGE in human prostate cancer cells**. *Exp Cell Res* 2006, **312**(2):184-197.

17. Hermani A, Hess J, De Servi B, Medunjanin S, Grobholz R, Trojan L, Angel P, Mayer D: **Calcium-binding proteins S100A8 and S100A9 as novel**

diagnostic markers in human prostate cancer. *Clin Cancer Res* 2005, **11**(14):5146-5152.

18. Stulik J, Koupilova K, Osterreicher J, Knizek J, Macela A, Bures J, Jandik P, Langr F, Dedic K, Jungblut PR: **Protein abundance alterations in matched sets of macroscopically normal colon mucosa and colorectal carcinoma**. *Electrophoresis* 1999, **20**(18):3638-3646.

19. Stulik J, Osterreicher J, Koupilova K, Knizek, Macela A, Bures J, Jandik P, Langr F, Dedic K, Jungblut PR: **The analysis of S100A9 and S100A8 expression in matched sets of macroscopically normal colon mucosa and colorectal carcinoma: the S100A9 and S100A8 positive cells underlie and invade tumor mass**. *Electrophoresis* 1999, **20**(4-5):1047-1054.

20. Yong HY, Moon A: **Roles of calcium-binding proteins, S100A8 and S100A9, in invasive phenotype of human gastric cancer cells**. *Arch Pharm Res* 2007, **30**(1):75-81.

21. Yao R, Lopez-Beltran A, Maclennan GT, Montironi R, Eble JN, Cheng L: **Expression of S100 protein family members in the pathogenesis of bladder tumors**. *Anticancer Res* 2007, **27**(5A):3051-3058.

22. Tolson JP, Flad T, Gnau V, Dihazi H, Hennenlotter J, Beck A, Mueller GA, Kuczyk M, Mueller CA: **Differential detection of S100A8 in transitional cell carcinoma of the bladder by pair wise tissue proteomic and immunohistochemical analysis**. *Proteomics* 2006, **6**(2):697-708.

23. Yao R, Davidson DD, Lopez-Beltran A, MacLennan GT, Montironi R, Cheng L: **The S100 proteins for screening and prognostic grading of bladder cancer**. *Histol Histopathol* 2007, **22**(9):1025-1032.

24. Cross SS, Hamdy FC, Deloulme JC, Rehman I: **Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are all overexpressed in common cancers**. *Histopathology* 2005, **46**(3):256-269.

25. Rafii S, Lyden D: **S100 chemokines mediate bookmarking of premetastatic niches**. *Nat Cell Biol* 2006, **8**(12):1321-1323.

26. Salama I, Malone PS, Mihaimeed F, Jones JL: **A review of the S100 proteins in cancer**. *Eur J Surg Oncol* 2008, **34**(4):357-364.

27. Babjuk M, Oosterlinck W, Sylvester R, Kaasinen E, Bohle A, Palou-Redorta J: **EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder**. *Eur Urol* 2008, **54**(2):303-314.

28. Hall MC, Chang SS, Dalbagni G, Pruthi RS, Seigne JD, Skinner EC, Wolf JS, Jr., Schellhammer PF: **Guideline for the management of nonmuscle invasive bladder cancer (stages Ta, T1, and Tis): 2007 update**. *J Urol* 2007, **178**(6):2314-2330.

29. Stenzl A, Cowan NC, De Santis M, Jakse G, Kuczyk MA, Merseburger AS, Ribal MJ, Sherif A, Witjes JA: **The Updated EAU Guidelines on Muscle-Invasive and Metastatic Bladder Cancer**. *Eur Urol* 2009, **55**(4):815-25.

30. Malmstrom PU, Sylvester RJ, Crawford DE, Friedrich M, Krege S, Rintala E, Solsona E, Di Stasi SM, Witjes JA: **An individual patient data meta-analysis of the long-term outcome of randomised studies comparing intravesical mitomycin C versus bacillus Calmette-Guerin for non-muscle-invasive bladder cancer**. *Eur Urol* 2009, **56**(2):247-256.

31. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns**. *Proc Natl Acad Sci U S A* 1998, **95**(25):14863-14868.

32. Dyrskjot L, Zieger K, Real FX, Malats N, Carrato A, Hurst C, Kotwal S, Knowles M, Malmstrom PU, de la Torre M *et al*: **Gene expression signatures predict outcome in non-muscle-invasive bladder carcinoma: a multicenter validation study**. *Clin Cancer Res* 2007, **13**(12):3545-3551.

33. Radmacher MD, McShane LM, Simon R: **A paradigm for class prediction using gene expression profiles**. *J Comput Biol* 2002, **9**(3):505-511.

34. Dudoit S, Fridlyand J, Speed TP: **Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data**. *J Am Stat Assoc* 2002, **97**(457):10.

35. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP *et al*: **Multiclass cancer diagnosis using tumor gene expression signatures**. *Proc Natl Acad Sci U S A* 2001, **98**(26):15149-15154.

36. Simon R, Radmacher MD, Dobbin K, McShane LM: **Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification**. *J Natl Cancer Inst* 2003, **95**(1):14-18.

37. Rahimi F, Hsu K, Endoh Y, Geczy CL: **FGF-2, IL-1beta and TGF-beta regulate fibroblast expression of S100A8**. *FEBS J* 2005, **272**(11):2811-2827.

38. Prieto C, Risueno A, Fontanillo C, De las Rivas J: **Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles**. *PLoS One* 2008, **3**(12):e3911.

39.   Thorgeirsson SS, Lee JS, Grisham JW: **Molecular prognostication of liver cancer: end of the beginning**. *J Hepatol* 2006, **44**(4):798-805.

40.   Apte RN, Dotan S, Elkabets M, White MR, Reich E, Carmi Y, Song X, Dvozkin T, Krelin Y, Voronov E: **The involvement of IL-1 in tumorigenesis, tumor invasiveness, metastasis and tumor-host interactions**. *Cancer Metastasis Rev* 2006, **25**(3):387-408.

41.   Yano S, Nokihara H, Yamamoto A, Goto H, Ogawa H, Kanematsu T, Miki T, Uehara H, Saijo Y, Nukiwa T *et al*: **Multifunctional interleukin-1beta promotes metastasis of human lung cancer cells in SCID mice via enhanced expression of adhesion-, invasion- and angiogenesis-related molecules**. *Cancer Sci* 2003, **94**(3):244-252.

Figure Legends

Figure 1 - Expression of *S100A8* and progression of non-muscle invasive bladder cancer

A: Expression of *S100A8* in 103 patients from the original cohort. B: Kaplan-Meier curves showing time to progression in the original cohort. C: Kaplan-Meier curves showing time to progression in the independent European cohort.

Figure 2 - Gene expression pattern of *S100A8*-correlated genes and progression of two clusters

A: Gene expression patterns of *S100A8* and its correlated genes. A total of 1,015 genes whose expression patterns are highly correlated with *S100A8* were selected for cluster analysis (Pearson correlation test, $P < 0.001$, $r < -0.3$ or $r > 0.3$). Patients were divided into two groups: high *S100A8* cluster (HSC) and low *S100A8* cluster (LSC). B: Kaplan-Meier curves showing time to progression. The progression rate of HSC patients was significantly higher than that of LSC patients (log-rank test, $P < 0.001$).

Figure 3 - Independent validation of the prognostic value of the signature

A: The validation strategy used for the construction of prediction models and the evaluation of predicted outcomes based on gene expression signature. B: Kaplan-Meier plots of progression of NMIBC patients from an independent European cohort predicted by compound covariate predictor (CCP), Bayesian compound covariate predictor (BCC), linear discriminator analysis (LDA), nearest centroid classification (NC), and support vector machines (SVM).

Figure 4 - Functional classification of *S100A8*-correlated genes

Classification enrichment was determined using Ingenuity Pathway Analysis software. The threshold of significance was –log ($P = 0.05$).

Figure 5 - Gene networks enriched with genes associated with *S100A8*

Gene networks of 1,015 genes that highly correlated with *S100A8*. Up- and down-regulated genes in the high *S100A8* cluster (HSC) group are indicated in red and green, respectively. The intensity of color is indicative of the degree of over- or under-expression. Genes without highlighted color are not part of the progression signature but are associated with the regulated genes. Each line and arrow represents functional and physical interactions between the genes and the direction of regulation reported in the literature.

Figure 6 - Comparison of expression levels between high *S100A8* cluster (HSC) and low *S100A8* cluster (LSC) patients

Two group box plot comparing expression levels of *S100A8* (A), *IL1B* (B), and *S100A9* (C) in HSC and LSC patients. *P*-value was obtained by two-sample t-test between HSC and LSC. The value of *r* indicates the correlation coefficient value of the gene compared with *S100A8*.

Tables

Table 1 - Baseline characteristics of primary non-muscle invasive bladder cancer patients

Variable	No. of patients (%)
Sex	
Male	87 (84.5)
Female	16 (15.5)
Grade	
Low	86 (83.5)
High	17 (16.5)
Stage	
Ta	23 (22.3)
T1	80 (77.7)
Progression	
No	92 (89.3)
Yes	11 (10.7)

Table 2 - Multivariate Cox regression analysis for prediction of disease progression

Variable	Progression	
	HR (95% CI)	*P*-value
Stage (Ta vs. T1)	0.258 (0.032 – 2.083)	0.204
Grade (low vs. high)	2.257 (0.445 – 11.449)	0.326
Number of tumors		
Single	Reference	-
2 to 7	1.684 (0.314 – 9.027)	0.543
>8	4.544 (0.624 – 32.141)	0.129
Size (> 3cm vs. ≤ 3cm)	1.933 (0.397 – 9.411)	0.414
Intravesical therapy (Yes vs. No)	2.141 (0.391 – 11.715)	0.38
S100A8-correlated gene signature (HSC vs. LSC)	15.225 (1.736 – 133.52)	0.014

Abbreviations: HR, hazards ratio; CI, confidence interval

Table 3 - Performance of prediction models

CCP

Class	Sensitivity	Specificity	PPV	NPV
HSC	0.92	0.943	0.939	0.926
LSC	0.943	0.92	0.926	0.939

BCC

Class	Sensitivity	Specificity	PPV	NPV
HSC	0.88	0.887	0.88	0.887
LSC	0.887	0.88	0.887	0.88

LDA

Class	Sensitivity	Specificity	PPV	NPV
HSC	0.9	0.943	0.938	0.909
LSC	0.943	0.9	0.909	0.938

NC

Class	Sensitivity	Specificity	PPV	NPV
HSC	0.9	0.962	0.957	0.911
LSC	0.962	0.9	0.911	0.957

SVM

Class	Sensitivity	Specificity	PPV	NPV
HSC	0.96	0.962	0.96	0.962
LSC	0.962	0.96	0.962	0.96

Sensitivity and specificity of compound covariate predictor (CCP), Bayesian compound covariate predictor (BCC), linear discriminator analysis (LDA), nearest centroid classification (NC), and support vector machines (SVM). Sensitivity is the probability for a class A sample to be correctly predicted as class A. Specificity is the probability for a non class A sample to be correctly predicted as non-A. Positive Predictive Value (PPV) is the probability that a sample predicted as class A actually belongs to class A. Negative Predictive Value (NPV) is the probability that a sample predicted as non class A actually does not belong to class A.

For some class high *S100A8* cluster (HSC), if $n11$ = number of class HSC samples predicted as HSC, $n12$ = number of class HSC samples predicted as low *S100A8*

cluster (LSC), n21 = number of LSC samples predicted as HSC, and n22 = number of LSC samples predicted as LSC, then the following parameters can characterize performance of classifiers: Sensitivity = n11/(n11+n12), Specificity = n22/(n21+n22), PPV = n11/(n11+n21), and NPV = n22/(n12+n22).

Additional files

Additional file 1 – The list of gene networks

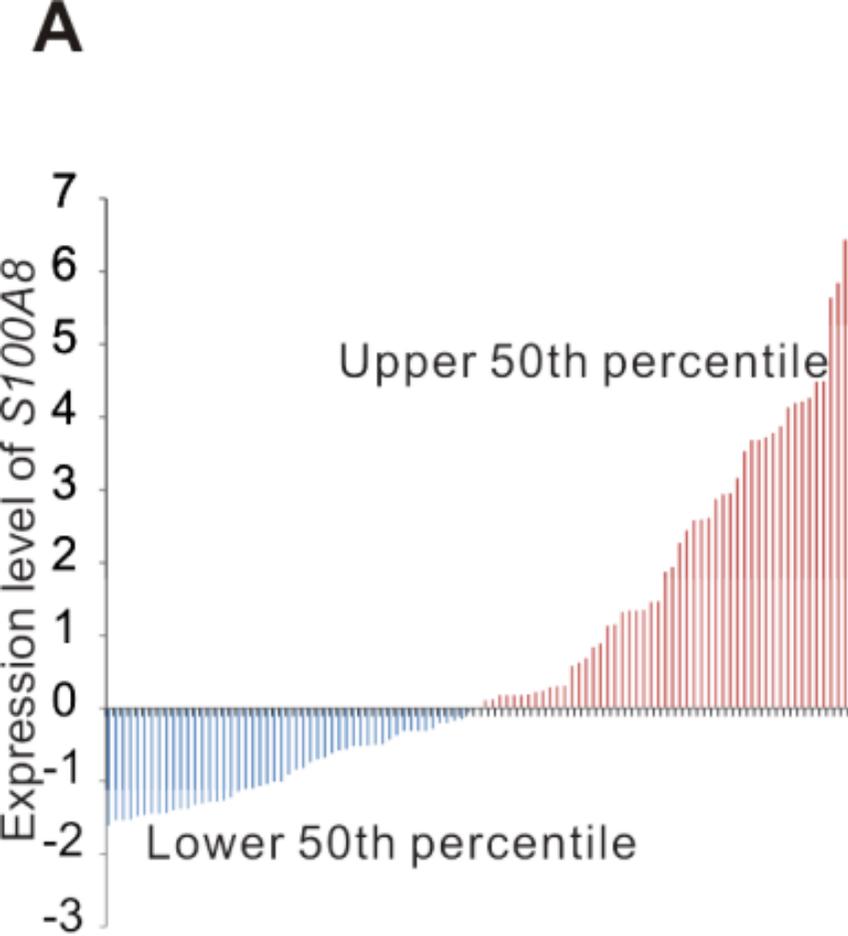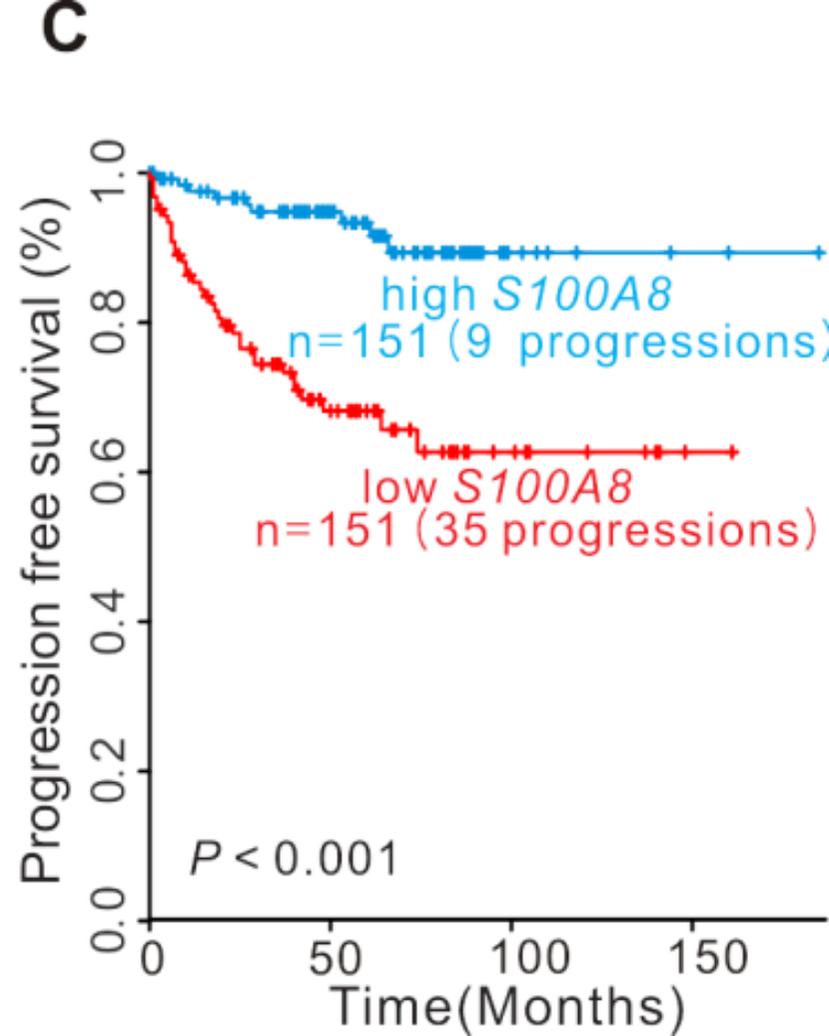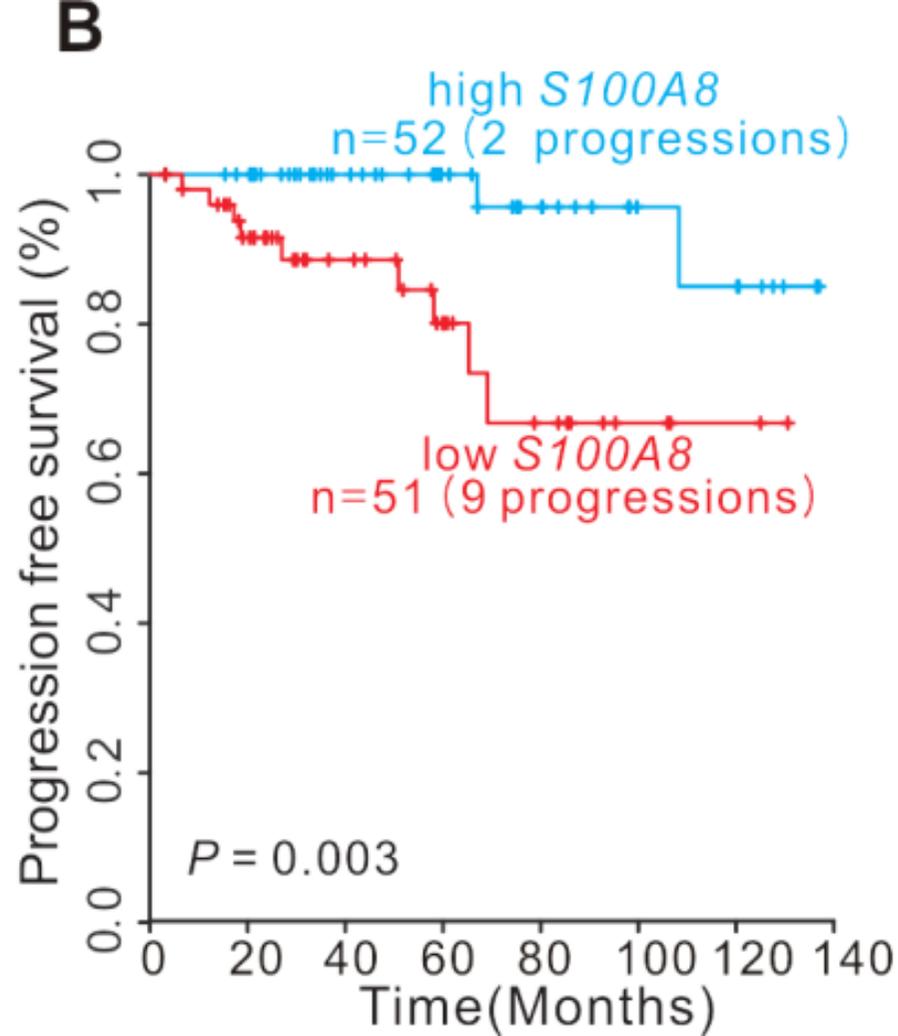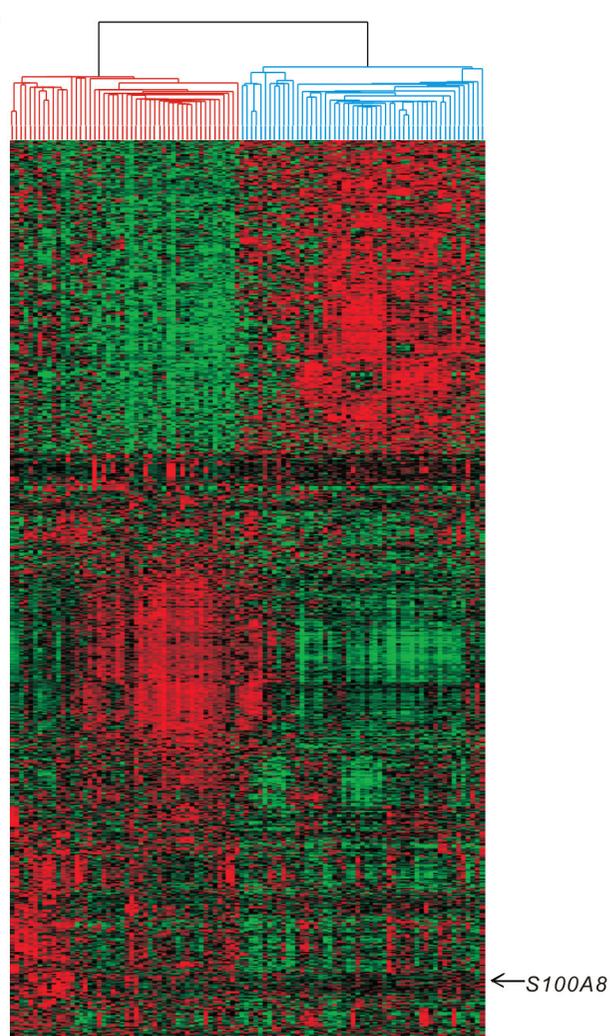Top10 list of gene networks constructed by Ingenuity^{TM} Pathway Analysis

A Expression level of *S100A8*

Upper 50th percentile

Lower 50th percentile

B Progression free survival (%)

high *S100A8*
n=52 (2 progressions)

low *S100A8*
n=51 (9 progressions)

P = 0.003

Time(Months)

C Progression free survival (%)

high *S100A8*
n=151 (9 progressions)

low *S100A8*
n=151 (35 progressions)

P < 0.001

Time(Months)

Figure 1

A

B

Progression free survival (%)

LSC
n=53 (1 progression)

HSC
n=50 (10 progressions)

P < 0.001

Time (months)

← *S100A8*

■ HSC, high *S100A8* cluster
■ LSC, low *S100A8* cluster

Figure 2

Figure 3

Function Enrichment

Figure 4

Figure 5
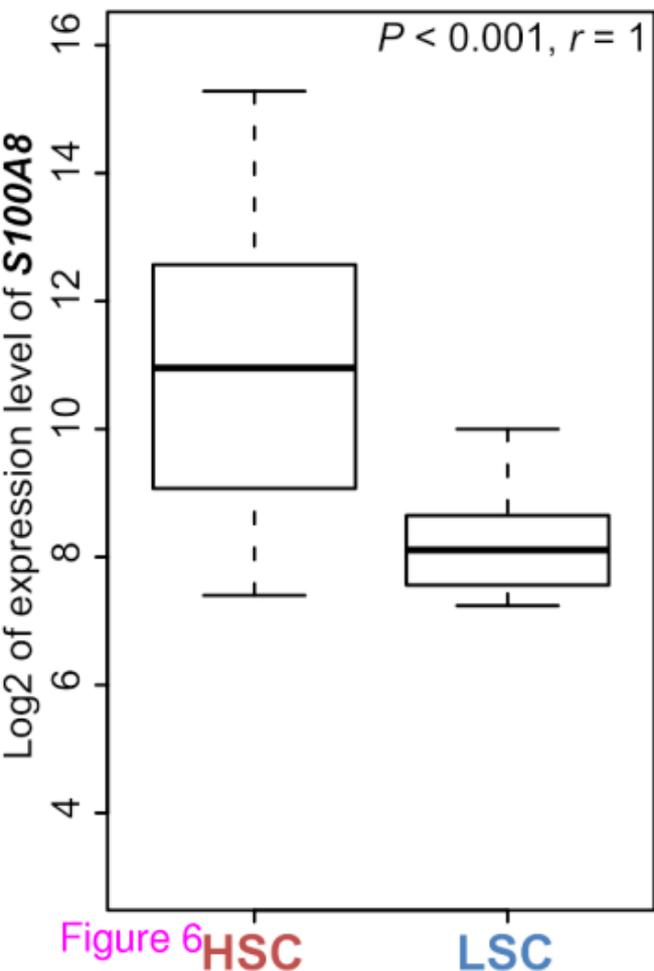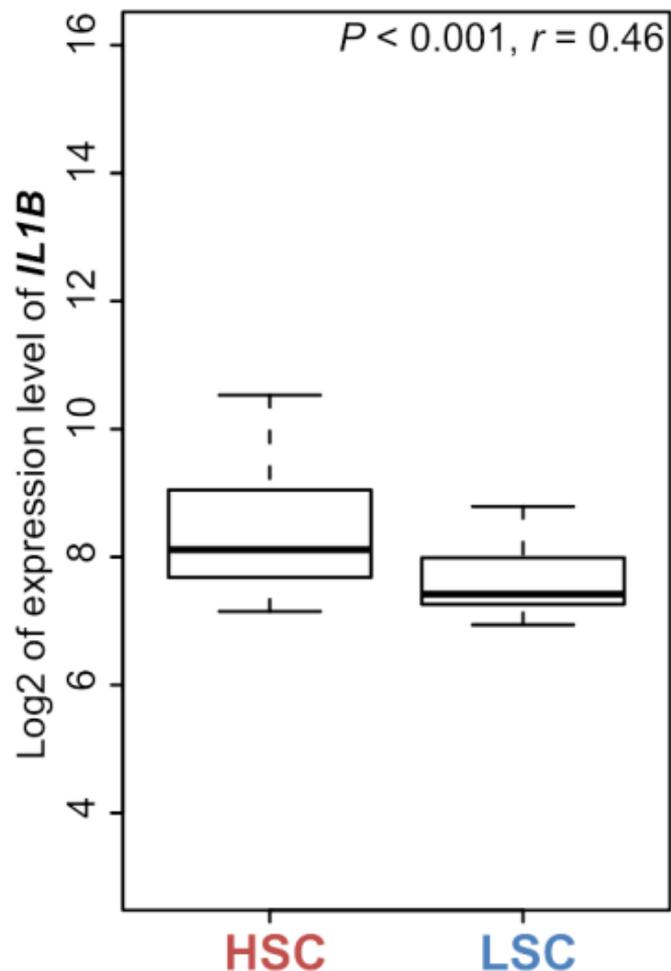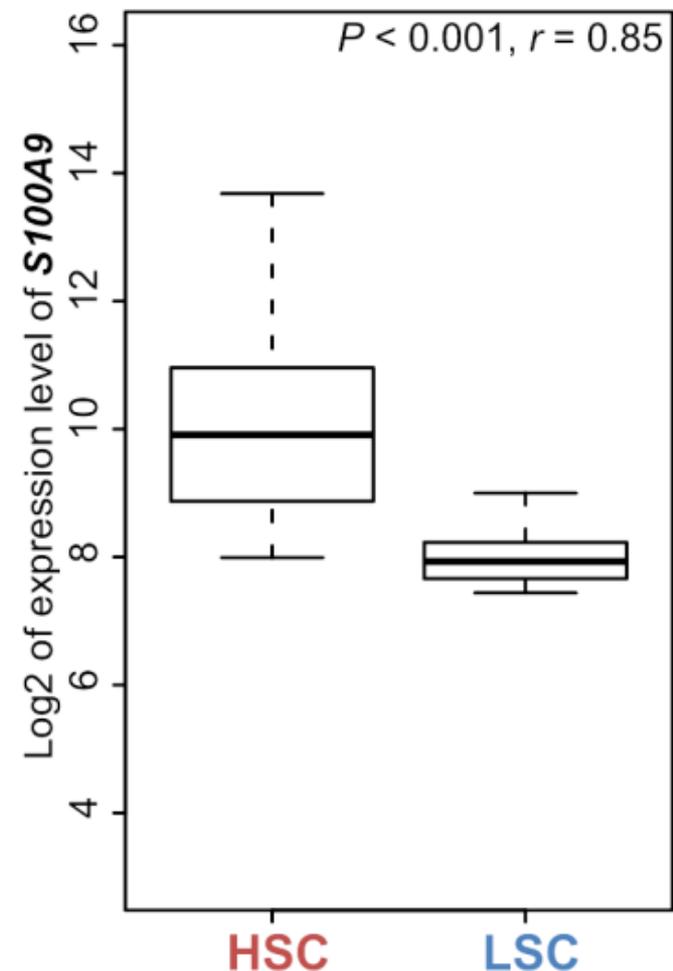
Figure 6

Additional files provided with this submission:

Additional file 1: Additional File 1.xls, 30K
http://www.biomedcentral.com/imedia/1790154151333281/supp1.xls