

**Table 1.** The 13 patient cohorts (N = 1454) included in *BC-BET*. The numbers in the table correspond to the number of patients with each clinical characteristic or available endpoint that are included in the database and analyzed. A '-' denotes insufficient sample size for analysis.

Cohort (availability)*	Platform	# of samples						Total (N)
		Normal, Tumor	LG, HG	NMI, MI	DSS	OS	RFS	
AUH-1 [1] (GSE3167)	Affymetrix Human Genome U133A	9,41	8, 32	28, 13	-	-	-	50
AUH-2 [2] (GSE5479)	MDL Human 3k	-	98, 271	351, 51	-	-	-	404
Blaveri [3] (S)	UCSF Human Array 2.0	-	10, 68	27, 53	-	74	-	74
CNUH [4] (GSE13507)	Illumina human-6 v2.0	10, 165	105, 60	104, 61	165	165	-	175
DFCI [5] (GSE31684)	Affymetrix Human Genome U133 Plus 2.0	-	6, 84	15, 78	-	-	90	93
Lindgren [6](GSE19915)	Swegene	12,144	72, 72	97, 45	-	142	-	156
Lindgren-2 [7] (GSE32548)	Illumina HumanHT-12 V3.0	-	56, 75	92, 38	-	89	-	131
MDA-1 [8](GSE48276)	Illumina HumanHT-12 WG-DASL V4.0 R2	-	-	-	-	22 <sup>†</sup>	-	22
MDA-2 [8](GSE48075)	Illumina HumanHT-12 V3.0	-	-	67, 73	-	73 <sup>†</sup>	-	140
MSKCC [9] (S)	Affymetrix Human Genome U133A	38,91	18, 73	25, 66	87	-	-	129
UVA[10] (GSE37317)	Affymetrix Human Genome U133A	-	-	8, 10	-	-	-	18
Stransky-1[11] (E-TABM-147)	Affymetrix Human Genome U95A	5,26	11, 15	9, 17	-	-	-	31
Stransky-2[11] (E-TABM-147)	Affymetrix Human Genome U95Av2	-	13, 16	16, 15	-	-	-	31
Total		74, 467	397, 769	839, 523	252	565	90	1454

\*Gene expression data for all cohorts are publicly available from the Gene Expression Omnibus (GEO) [12] with the given Accession # (GSE ID), from Array Express [13] (Accession # E-TABM-147) or as Supplementary material to publication (S). <sup>†</sup>patients have MI, HG tumors (MDA-1) or MI tumors with unspecified grade (MDA-2). Abbreviations: LG, low grade; HG, high grade; NMI, non-muscle invasive; MI, muscle-invasive; DSS, disease-specific survival; OS, overall survival; RFS, recurrence-free survival.

**Supporting Table 1.** Summary of patient cohorts, sample exclusion and processing for *BC-BET*. The **Complete cohort** columns correspond to all samples available at the given accession number or publication, while the **BC-BET** columns corresponds to the samples included in the database.

Cohort (availability)*	<b>Complete cohort</b>		<b>BC-BET</b>	
	N	Description	N	Sample Exclusion, Processing, and Additional Notes
AUH-1 [1] (GSE3167)	60	Forty-one tumor biopsies, 9 biopsies of normal bladder mucosa from patients without bladder cancer, and 5 paired samples consisting of a normal biopsy and a biopsy with CIS were profiled.	50	The paired biopsy samples are excluded.

AUH-2 [2] (GSE5479)	808	Tumors from 404 patients were obtained from bladder cancer patients from hospitals in Denmark, Sweden, England, Spain, and France and were profiled in duplicate.	404	Replicate profiles were averaged to produce a single gene expression profile for each patient. Eighteen samples are also profiled in AUH-1 and are removed if AUH-1 is analyzed.
Blaveri [3] (S)	80	Eighty tumor samples were obtained from the tissue bank of the University of California at San Francisco Comprehensive Cancer Center. Six of these samples had squamous histology. All patients receive either transurethral resection of the bladder (TURBT) or cystectomy. Specifically, of those from transitional cell carcinomas, 25/27 patients with NMI tumors underwent TURBT while 44/47 patients with MI tumors receive cystectomies.	74	The 6 samples with squamous histology are excluded. For prognostic biomarker evaluation, only the NMI samples obtained from TURBT and MI samples obtained from cystectomy are analyzed.
CNUH [4] (GSE13507)	256	Samples were collected at Chungbuk National University Hospital (CNUH), and consisted of 165 primary bladder cancer samples, 23 recurrent tumor samples, 58 normal samples surrounding tumors, and 10 normal bladder samples from patients with benign disease. Profiles from NMI tumors were obtained by transurethral resection of the bladder while those from MI tumors were obtained by cystectomy. Fifty-six patients with intermediate- or high-risk NMI tumors received intravesical Bacillus Calmette-Guerin therapy while 26 patients with MI tumors and 1 patient with an MI tumor received cisplatin-based adjuvant chemotherapy.	175	Samples from recurrent tumors and samples of bladder mucosae surrounding cancer are excluded. The user has the option of excluding patients treated with intravesical therapy or chemotherapy in the prognostic biomarker evaluation.
DFCI [5] (GSE31684)	93	Tumors were profiled from 93 patients undergoing radical cystectomy at Memorial Sloan Kettering Cancer Center between 1993 and 2004. Three patients received neoadjuvant chemotherapy, 16 received adjuvant chemotherapy, and 19 patients received salvage chemotherapy for recurrent disease.	93	Patients receiving neoadjuvant chemotherapy are excluded from the prognostic biomarker evaluation. Optionally, the user can exclude chemotherapy-treated patients from the prognostic biomarker evaluation.
Lindgren [6] (GSE19915)	156	Tumor profiles were obtained from 144 patients undergoing transurethral resection of the bladder from the University Hospital of Lund, Sweden, as well as normal profiles from 12 patients with non-bladder cancer related disorders. Forty-six patients are later cystectomized, the majority (N = 32) who had HG, MI tumors.	156	In the prognostic biomarker evaluation in patients with HG, MI tumors, 11 patients not cystectomized are excluded. Ninety patients are profiled in Lindgren-2 and are excluded if Lindgren-2 is analyzed.
Lindgren-2 [7] (GSE32548)	131	Tumor profiles were obtained from 131 patients undergoing transurethral resection of the bladder from University Hospital of Lund, Sweden. Thirty-two patients are later cystectomized, the majority (N = 21) who had HG, MI tumors.	131	In the prognostic biomarker evaluation in patients with HG, MI tumors, 14 patients not cystectomized are excluded.
MDA-1 [8] (GSE48276)	128	Profiles were obtained from 128 bladder tumors from patients at MD Anderson (MDA) or Fox-Chase Cancer Center and Thomas Jefferson University (Philadelphia samples). Of the MDA tumor samples, only 24 had urothelial histology at cystectomy and were not treated by neoadjuvant chemotherapy	22	The Philadelphia samples were all MI and did not have survival information, and were therefore not analyzed. Of the MDA samples, tumors not having urothelial histology at cystectomy and those treated with neoadjuvant chemotherapy were excluded. Two additional NMI tumors were also excluded.
MDA-2 [8] (GSE48075)	142	Gene expression profiles from 142 primary bladder tumors were obtained.	142	No samples were excluded.
MSKCC [9] (S)	157	Gene expression profiles from 105 patients with bladder cancer were obtained along with normal samples, with some samples in duplicate or triplicate.	91	Previously processed data is included [14], which averaged replicate profiles to produce a single gene expression profile for each patient and removed several low quality samples.
UVA[10]	19	Gene expression profiles of 19 bladder cancer	18	One sample with squamous cell histology

(GSE37317)		patients were obtained at the University of Virginia.		was excluded.
Stransky-1 and Stransky-2[11] (E-TABM-147)	103	Gene expression profiles from 52 bladder cancer patients from Henri Mondor Hospital are available, with many samples profiled in duplicate. Ten normal samples were also obtained.	62	Replicate profiles were averaged to produce a single gene expression profile for each patient

## References

1. Dyrskjot L, Kruhoffer M, Thykjaer T, Marcussen N, Jensen JL, Moller K, Orntoft TF: **Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification.** *Cancer research* 2004, **64**(11):4040-4048.
2. 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.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2007, **13**(12):3545-3551.
3. Blaveri E, Simko JP, Korkola JE, Brewer JL, Baehner F, Mehta K, Devries S, Koppie T, Pejavar S, Carroll P *et al*: **Bladder cancer outcome and subtype classification by gene expression.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2005, **11**(11):4044-4055.
4. Kim WJ, Kim EJ, Kim SK, Kim YJ, Ha YS, Jeong P, Kim MJ, Yun SJ, Lee KM, Moon SK *et al*: **Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer.** *Molecular cancer* 2010, **9**:3.
5. Riester M, Taylor JM, Feifer A, Koppie T, Rosenberg JE, Downey RJ, Bochner BH, Michor F: **Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012, **18**(5):1323-1333.
6. Lindgren D, Frigyesi A, Gudjonsson S, Sjodahl G, Hallden C, Chebil G, Veerla S, Ryden T, Mansson W, Liedberg F *et al*: **Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome.** *Cancer research* 2010, **70**(9):3463-3472.
7. Lindgren D, Sjodahl G, Lauss M, Staaf J, Chebil G, Lovgren K, Gudjonsson S, Liedberg F, Patschan O, Mansson W *et al*: **Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma.** *PloS one* 2012, **7**(6):e38863.
8. Choi W, Porten S, Kim S, Willis D, Plimack ER, Hoffman-Censits J, Roth B, Cheng T, Tran M, Lee IL *et al*: **Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy.** *Cancer cell* 2014, **25**(2):152-165.
9. Sanchez-Carbayo M, Socci ND, Lozano J, Saint F, Cordon-Cardo C: **Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays.** *J Clin Oncol* 2006, **24**(5):778-789.
10. Smith SC, Baras AS, Owens CR, Dancik G, Theodorescu D: **Transcriptional signatures of Ral GTPase are associated with aggressive clinicopathologic characteristics in human cancer.** *Cancer research* 2012, **72**(14):3480-3491.

11. Stransky N, Vallot C, Reyat F, Bernard-Pierrot I, de Medina SG, Segraves R, de Rycke Y, Elvin P, Cassidy A, Spraggon C *et al*: **Regional copy number-independent deregulation of transcription in cancer.** *Nature genetics* 2006, **38**(12):1386-1396.
12. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM *et al*: **NCBI GEO: archive for functional genomics data sets--10 years on.** *Nucleic acids research* 2011, **39**(Database issue):D1005-1010.
13. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M *et al*: **ArrayExpress update--trends in database growth and links to data analysis tools.** *Nucleic acids research* 2013, **41**(Database issue):D987-990.
14. Smith SC, Baras AS, Dancik G, Ru Y, Ding KF, Moskaluk CA, Fradet Y, Lehmann J, Stockle M, Hartmann A *et al*: **A 20-gene model for molecular nodal staging of bladder cancer: development and prospective assessment.** *The Lancet Oncology* 2011, **12**(2):137-143.