This is a repository copy of *Genome-wide meta-analysis identifies novel genes associated with recurrence and progression in non–muscle-invasive bladder cancer*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/185321/

Version: Published Version

Article:
Galesloot, T.E., Grotenhuis, A.J., Kolev, D. et al. (16 more authors) (2022) Genome-wide meta-analysis identifies novel genes associated with recurrence and progression in non–muscle-invasive bladder cancer. European Urology Oncology, 5 (1). pp. 70-83. ISSN 2588-9311

https://doi.org/10.1016/j.euo.2021.07.001

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

available at www.sciencedirect.com
journal homepage: euoncology.europeanurology.com

European Association of Urology

Priority Article

Genome-wide Meta-analysis Identifies Novel Genes Associated with Recurrence and Progression in Non–muscle-invasive Bladder Cancer

Tessel E. Galesloot [a,], Anne J. Grotenhuis [a], Dimitar Kolev [a], Katja K. Aben [a,b], Richard T. Bryan [c,d], James W.F. Catto [e], Kar K. Cheng [f], Samantha Conroy [e], Lars Dyrskjøt [g,h], Neil E. Fleshner [i], Nicholas D. James [c], Philippe Lamy [g], Sia Viborg Lindskrog [g,h], Núria Malats [j,k], Lourdes Mengual [l], Gerald Verhaegh [m], Maurice P. Zeegers [c,n,o], Lambertus A.L.M. Kiemeney [a,m,†], Sita H. Vermeulen [a,†]*

[a] *Department for Health Evidence, Radboud university medical center, Nijmegen, The Netherlands;* [b] *Netherlands Comprehensive Cancer Organisation, Utrecht, The Netherlands;* [c] *Institute of Cancer & Genomic Sciences, University of Birmingham, Birmingham, UK;* [d] *Bladder Cancer Research Centre, University of Birmingham, Birmingham, UK;* [e] *Academic Urology Unit, University of Sheffield, Sheffield, UK;* [f] *Institute of Applied Health Research, University of Birmingham, Birmingham, UK;* [g] *Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark;* [h] *Department of Clinical Medicine, Aarhus University, Aarhus, Denmark;* [i] *Department of Urology, Princess Margaret Cancer Centre, Toronto, Canada;* [j] *Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), Madrid, Spain;* [k] *CIBERONC, Madrid, Spain;* [l] *Department and Laboratory of Urology, Hospital Clínic, IDIBAPS, Universitat de Barcelona, Barcelona, Spain;* [m] *Department of Urology, Radboud university medical center, Nijmegen, The Netherlands;* [n] *Department of Complex Genetics and Epidemiology, School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, The Netherlands;* [o] *CAPHRI School for Public Health and Primary Care, University of Maastricht, Maastricht, The Netherlands*

Article info

Article history:
Received 9 June 2021
Received in revised form
15 June 2021
Accepted July 1, 2021

Associate Editor:
Ashish Kamat

Keywords:
Non–muscle-invasive bladder cancer
Prognosis
Recurrence
Progression
Genome-wide association study
Meta-analysis

Abstract

Background: Non–muscle-invasive bladder cancer (NMIBC) is characterized by frequent recurrences and a risk of progression in stage and grade. Increased knowledge of underlying biological mechanisms is needed.

Objective: To identify single nucleotide polymorphisms (SNPs) associated with recurrence-free (RFS) and progression-free (PFS) survival in NMIBC.

Design, setting, and participants: We analyzed outcome data from 3400 newly diagnosed NMIBC patients from the Netherlands, the UK, Canada, and Spain. We generated genome-wide germline SNP data using Illumina OmniExpress and Infinium Global Screening Array in combination with genotype imputation.

Outcome measurements and statistical analysis: Cohort-specific genome-wide association studies (GWASs) for RFS and PFS were performed using a Cox proportional hazard model. Results were combined in a fixed-effect inverse-variance weighted meta-analysis. Candidate genes for the identified SNP associations were prioritized using functional annotation, gene-based analysis, expression quantitative trait locus analysis, and transcription factor binding site databases. Tumor expression levels of prioritized genes were tested for association with RFS and PFS in an independent NMIBC cohort.

[†] These authors contributed equally to this manuscript.

* Corresponding author. Department for Health Evidence, Radboud university medical center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. Tel.: +31 (0) 24 3614266.
E-mail address: Tessel.Galesloot@radboudumc.nl (T.E. Galesloot).

https://doi.org/10.1016/j.euo.2021.07.001
2588-9311/© 2021 The Authors. Published by Elsevier B.V. on behalf of European Association of Urology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Results and limitations: This meta-analysis revealed a genome-wide significant locus for RFS on chromosome 14 (lead SNP rs12885353, hazard ratio [HR] C vs T allele 1.55, 95% confidence interval [CI] 1.33–1.82, $p = 4.0 \times 10^{-8}$), containing genes *G2E3* and *SCFD1*. Higher expression of *SCFD1* was associated with increased RFS (HR 0.70, 95% CI 0.59–0.84, $p_{FDR} = 0.003$). Twelve other loci were suggestively associated with RFS ($p < 10^{-5}$), pointing toward 18 additional candidate genes. For PFS, ten loci showed suggestive evidence of association, indicating 36 candidate genes. Expression levels of ten of these genes were statistically significantly associated with PFS, of which four (*IFT140*, *UBE2I*, *FAHD1*, and *NME3*) showed directional consistency with our meta-analysis results and published literature.

Conclusions: In this first prognostic GWAS in NMIBC, we identified several novel candidate loci and five genes that showed convincing associations with recurrence or progression.

Patient summary: In this study, we searched for inherited DNA changes that affect the outcome of non–muscle-invasive bladder cancer (NMIBC). We identified several genes that are associated with disease recurrence and progression. The roles and mechanisms of these genes in NMIBC prognosis should be investigated in future studies.

© 2021 The Authors. Published by Elsevier B.V. on behalf of European Association of Urology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Approximately 75% of urinary bladder cancer (UBC) patients present with non–muscle-invasive bladder cancer (NMIBC) [1]. NMIBC patients generally have a good prognosis with 5-yr disease-specific survival of 90–95%, but the disease is characterized by frequent local recurrences and a risk of progression to the more lethal muscle-invasive bladder cancer (MIBC) [2,3]. Consequently, frequent surveillance episodes and repeated treatments are necessary, resulting in considerable patient and health care burden [4].

Significant heterogeneity in outcomes, that is, recurrence and progression rates, exists among NMIBC patients [2,3]; hence, improved knowledge of underlying biological mechanisms and identification of novel prognostic biomarkers are needed to improve clinical management. Notwithstanding, analyses of genetic markers of cancer outcomes demonstrate increasing evidence that, in addition to somatic events, germline genetic variation plays a role in cancer outcomes and response to cancer treatment. For example, in 2010, Chen et al [5] reported a replicated association between germline genetic variations in the sonic hedgehog pathway and clinical outcomes in NMIBC. A recent review by Lipunova et al [6] provides an overview of the currently published germline genetic associations for NMIBC and MIBC outcomes, also for specific subgroups such as bacillus Calmette-Guérin (BCG)-treated patients. A total of 81 associations for recurrence and 24 for progression were reported, based on a final set of 112 articles.

Unfortunately, the validity of these associations is unknown, as most findings were not replicated in independent series [6]. Indeed, in a previous replication study from our group, we were able to replicate only six out of 114 previously reported associations for UBC prognosis and treatment response [7].

As previous studies were candidate-gene studies and thus based on existing knowledge and hypotheses of cancer biology, there is a clear need for an *agnostic* genome-wide approach to allow for identification of new genetic loci for NMIBC prognosis as well as to evaluate the replicability of previous candidate-gene findings. Such an approach has already been successful for prognostic outcomes in pancreatic and prostate cancer [8,9]. Here, we report the first meta-analysis of genome-wide association studies (GWASs) for NMIBC recurrence and progression.

2. Patients and methods

2.1. Study populations

We included six cohorts: the Nijmegen Bladder Cancer Study (NBCS, Nijmegen, The Netherlands; $N = 1451$), the Bladder Cancer Prognosis Programme (BCPP, Birmingham, UK; $N = 684$), two cohorts from the Genito-Urinary BioBank (GUB-1 and GUB-2, Toronto, Canada; $N = 353$ and 432, respectively), and biobanked case series from the University of Sheffield (Sheffield, UK; $N = 244$) and the Hospital Clínic of Barcelona (Barcelona, Spain; $N = 238$). Details about the cohorts (outcome definitions, genotyping, quality control, and imputation) are provided in the Supplementary material. Each study was approved by local research ethics committees. All participants provided informed consent.

2.2. GWAS and meta-analysis

Single nucleotide polymorphisms (SNPs) were tested for association with recurrence-free (RFS) and progression-free (PFS) survival in cohort-specific GWASs. We used multivariable Cox proportional hazard regression models in gwasurvivr software version 1.0.0 [10], according to an additive genotype model including ten multidimensional scaling components to prevent population stratification bias. Results were combined in a fixed-effect inverse-variance weighted meta-analysis using METAL software (total N max 3400) applying genomic control to the GWAS results and a minor allele frequency (MAF) filter of 5%,

resulting in a total of 7591411 and 7582931 association tests for RFS and PFS, respectively. Meta-analysis results were filtered for inclusion of at least three cohorts in the meta-analysis (a SNP can lack in a cohort due to MAF <5% or no genotyping/imputation) and directionally consistent beta coefficients among cohorts. SNPs with $p < 5 \times 10^{-8}$ using a one degree of freedom Wald test were considered genome-wide statistically significant. A secondary threshold for suggestive evidence of association was set at $p < 1 \times 10^{-5}$. Heterogeneity of effect estimates across study cohorts was assessed using the I^2 statistic. Top associated loci (loci with at least one SNP with $p < 1 \times 10^{-5}$) were summarized by a lead index SNP defined as the SNP with the smallest p value in the region. Lead SNPs were also investigated for their effect on RFS and PFS after stratification of patients into low- and high-risk groups (Supplementary material).

2.3. Gene prioritization strategies

To identify candidate genes that might mediate the association of the lead SNPs with RFS and PFS, we performed expression quantitative trait locus (eQTL) analyses using (1) functional mapping and annotation (FUMA; v1.3.6) [11] for expression in whole blood, (2) PancanQTL [12] for expression in MIBCs from The Cancer Genome Atlas, and (3) data from the UROMOL study [13] for expression in NMIBCs (Supplementary material). Additionally, lead SNPs were checked for chromatin interaction in FUMA and for their effect on transcription factor binding sites (TFBSs) using SNP2TFB2 [14].

2.4. Gene-based and gene-set analyses

Gene-based and gene-set analyses were performed using meta-GWAS summary statistics as input and MAGMA software (v1.07) as implemented in FUMA [15], with a Bonferroni-adjusted threshold of 2.64×10^{-6} for statistical significance (adjusted for 18 957 genes; Supplementary material).

2.5. Correlations of tumor gene expression of prioritized genes with recurrence and progression

Prioritized genes were identified from annotation of lead SNPs, gene-based-analyses, and TFBS and eQTL analyses, resulting in 20 and 36 genes

Table 1 – Baseline patient and tumor characteristics of NMIBC patients per cohort

		NBCS (*N* = 1451)	BCPP (*N* = 684)	GUB-1 (*N* = 353)	GUB-2 (*N* = 432)	Sheffield (*N* = 244)	Barcelona (*N* = 238)
Female sex, *N* (%)		261 (17)	150 (22)	56 (16)	93 (22)	53 (22)	33 (14)
Age (yr), median (range)		64 (25–91)	71 (34–92)	67 (22–98)	73 (22–97)	71 (32–92)	69 (29–100)
Smoking status, *N* (%)	Never	245 (17)	147 (22)	86 (24)	111 (26)	25 (10)	47 (20)
	Former	677 (47)	356 (52)	198 (56)	187 (43)	94 (39)	105 (44)
	Current	424 (29)	130 (19)	41 (12)	78 (18)	34 (14)	58 (24)
	Unknown	105 (7)	51 (7)	28 (8)	56 (13)	91 (37)	28 (12)
Tumor stage, *N* (%)	Ta	1010 (70)	460 (67)	200 (57)	263 (61)	163 (67)	113 (48)
	CIS	55 (4)	12 (2)	33 (9)	32 (7)	8 (3)	8 (3)
	T1	365 (25)	211 (31)	102 (29)	137 (32)	73 (30)	115 (48)
	Unknown	21 (1)	1 (0)	18 (5)	0 (0)	0 (0)	2 (1)
Concomitant CIS, *N* (%)	No	1321 (91)	370 (54)	246 (70)	334 (77)	118 (48)	213 (89)
	Yes	111 (8)	88 (13)	106 (30)	90 (21)	59 (24)	25 (11)
	Unknown	19 (1)	226 (33)	1 (0)	8 (2)	67 (27)	0
Tumor grade, *N* (%)	G1 or low grade/PUNLMP	913 (63)	198 (29)	148 (42)	231 (53)	62 (25)	97 (41)
	G2	NA	254 (37)	17 (5)	NA	99 (41)	30 (13)
	G3 or high grade	521 (36)	220 (32)	172 (49)	198 (46)	83 (34)	100 (42)
	Unknown	15 (1)	12 (2)	16 (5)	3 (1)	0	11 (5)
Tumor size (cm), *N* (%)	<3	192 (13)	406 (59)	295 (84)	292 (67)	57 (23)	165 (69)
	≥3	112 (8)	256 (37)	20 (6)	103 (24)	96 (39)	48 (20)
	Unknown	1147 (79)	22 (3)	38 (11)	37 (9)	91 (37)	25 (11)
Tumor focality, *N* (%)	Solitary	784 (54)	396 (58)	154 (44)	249 (58)	63 (26)	131 (55)
	Multifocal	584 (40)	267 (39)	169 (48)	166 (38)	87 (36)	81 (34)
	Unknown	83 (6)	21 (3)	30 (8)	17 (4)	94 (39)	26 (11)
Follow-up (yr), median (min–max) [a]		7.2 (0.1–35.3)	3.7 (0–7.8)	3.3 (0–46.0)	8.9 (0–44.9)	2.0 (0.2–17.0)	3.9 (0–25.5)
Recurrence within 5 yr, *N* (%)		591 (41)	253 (37)	178 (51)	221 (52)	73 (34)	150 (63)
Kaplan-Meier 5-yr risk of recurrence (%)		47.3	43.6	58.2	53.7	53.9	78.6
Progression within 5 yr, *N* (%)		171 (12)	74 (11)	72 (20)	73 (17)	43 (18)	37 (16)
Kaplan-Meier 5-yr risk of progression (%)		13.9	14.5	25.3	17.5	31.5	20.4

BCPP = Bladder Cancer Prognosis Programme (Birmingham, UK); CIS = carcinoma in situ; G = grade; GUB = Genito-Urinary BioBank (Toronto, Canada); *N* = number; NA = not applicable; NBCS = Nijmegen Bladder Cancer Study (Nijmegen, The Netherlands); NMIBC = non–muscle-invasive bladder cancer; PUNLMP = papillary urothelial neoplasm of low malignant potential.
For initial NMIBC treatment per cohort, see Supplementary Table 1.
Tumor histology is 100% urothelial cell carcinoma for NBCS, GUB-1, GUB-2, Sheffield, and Barcelona, and 98.3% for BCPP (1.3% unknown, 0.4% other).
[a] Follow-up was defined as the time between primary NMIBC diagnosis (initial transurethral resection of the tumor) and date of the last urological check-up visit. If the latter was not available, date last alive or date of death was used.

Table 2 – Top associated loci ($p < 1 \times 10^{-5}$) for RFS summarized by a lead index SNP defined as the SNP with the smallest association *p* value in the region

SNP	CHR	BP	A1	A2	Direction [a]	HR	95% CI	*p* value	MAF	*N*	Info [b]	Het I^2 (*p* value) [c]	(Nearest) Gene(s) [d]
rs12885353 [e]	14	31078574	C	T	?++?++	1.86	1.48–2.34	9.5×10^{-8}	0.05–0.07 (C)	1231	>0.90	70.2 (0.02)	*G2E3*
rs192039210	7	2207116	A	G	?+++??	1.94	1.50–2.50	3.6×10^{-7}	0.05–0.08 (A)	1461	0.63	0 (0.55)	*MAD1L1*
rs17834128	12	14072420	T	C	++++++	1.28	1.16–1.41	3.8×10^{-7}	0.16–0.18 (T)	3366	>0.98	0 (0.52)	*GRIN2B*
rs7329778	13	36281898	G	C	++++++	1.30	1.17–1.44	1.4×10^{-6}	0.18–0.21 (G)	3366	0.65–0.72	0 (0.45)	*#LINC00445* [f]
rs12967544	18	50910125	C	T	++++++	1.29	1.16–1.44	1.7×10^{-6}	0.12–0.17 (C)	3366	>0.99	0 (0.91)	*DCC*
rs9992900	4	14684423	A	G	++++++	1.30	1.17–1.46	3.0×10^{-6}	0.18–0.37 (G)	3366	0.73–0.96	0 (0.75)	*LINC00504* [g]
rs4420730	2	2815611	T	C	?+++??	1.71	1.36–2.14	4.0×10^{-6}	0.03–0.06 (T)	1461	>0.92	63.6 (0.06)	*#LINC01250* [h]
rs3808347	7	939976	T	G	++++++	1.43	1.23–1.67	4.0×10^{-6}	0.06–0.07 (T)	3366	0.80–0.95	0 (0.89)	*ADAP1*
rs9935790	16	25066416	T	C	++++++	1.29	1.16–1.44	4.3×10^{-6}	0.12–0.18 (T)	3366	>0.85	0 (0.46)	*#LINC02175* [i]
rs2839488	21	43786186	C	G	++++++	1.20	1.11–1.29	5.0×10^{-6}	0.34–0.44 (C) [j]	3366	>0.95	10.4 (0.35)	*TFF1*
rs4351611	1	4157842	T	C	++++++	1.20	1.11–1.30	7.2×10^{-6}	0.37–0.41 (C)	3366	>0.94	23.3 (0.26)	*#EEF1DP6* [k]
rs7091482	10	7316841	A	G	+??+?+	1.55	1.28–1.88	9.3×10^{-6}	0.04–0.07 (A)	2351	0.82–0.97	0 (0.67)	*SFMBT2*
rs73038204	3	169604978	T	C	++++++	1.34	1.18–1.52	9.5×10^{-6}	0.06–0.12 (T)	3366	0.87	0 (0.70)	*#KRT18P43* [l]

A1 = effect allele; A2 = reference allele; BCPP = Bladder Cancer Prognosis Programme (Birmingham, UK); BP = basepair position; CHR = chromosome; CI = confidence interval; GUB = Genito-Urinary BioBank (Toronto, Canada); Het = heterogeneity; HR = hazard ratio; LINC = long intergenic or intronic non–protein coding RNA; MAF = minor allele frequency; *N* = number; NBCS = Nijmegen Bladder Cancer Study (Nijmegen, The Netherlands); RFS = recurrence-free survival; SNP = single nucleotide polymorphism.

SNPs with a cohort-specific MAF >5% were included in the meta-analysis. This table contains all top-ranked SNPs that are based on meta-analysis of at least three cohorts and directional consistency in effect estimates between cohorts. Only the strongest signal per locus is shown (for the complete list of SNPs associated with RFS with $p < 1 \times 10^{-5}$, see Supplementary Table 2).

Basepair positions and annotations are according to GRCh37/hg19 NCBI Reference Sequence collection.

[a] Order of the cohorts: NBCS, GUB-1, GUB-2, BCPP, Barcelona, and Sheffield.

[b] Impute info score, which entails the quality of imputation, with 1 = directly measured SNP and >0.8 for high-quality imputation.

[c] I^2 statistic measures heterogeneity on a scale of 0–100%.

[d] Genes indicated with a # are nearest genes. All SNPs in this table that reside within a gene are intronic variants.

[e] After a meta-analysis of all six cohorts (no MAF threshold), rs12885353 reached genome-wide significance (HR C vs T 1.55, 95% CI 1.33–1.82, $p = 4.0 \times 10^{-8}$) as well as rs34339578 in the same locus (HR T vs G 1.60, 95% CI 1.35–1.89, $p = 4.8 \times 10^{-8}$; intronic variant of gene *SCFD1*), although moderate to substantial effect heterogeneity was present (Table 3, and Supplementary Fig. 3 and 4).

[f] Nearest protein coding gene is *NBEA*.

[g] Nearest protein coding gene is *CPEB2*.

[h] Nearest protein coding gene is *EIPR1*.

[i] Nearest protein coding gene is *ARHGAP17*.

[j] Observed MAF differences correspond to 1000 genome frequencies, that is, 37% in CEU population, 47% in GBR population, and 36% in IBS population.

[k] *EEF1DP6* is a pseudogene. Nearest protein coding gene is *C1orf174*.

[l] *KRT18P43* is a pseudogene. Nearest protein coding gene is *LRRC31*.

for RFS and PFS, respectively. Tumor gene expression was tested for association with RFS and PFS in the NMIBC patient cohort from UROMOL (*N* = 511 with 348 events for RFS and *N* = 530 with 65 events for PFS) [13].

2.6. Replication of previously reported loci

We checked the associations of SNPs that were previously published for association with RFS or PFS in the whole NMIBC group, as summarized by Lipunova et al [6] plus rs4976845 (published for association with progression after the review [16]).

3. Results

3.1. Characteristics of NMIBC patients in the study cohorts

Among the cohorts, the median age at NMIBC diagnosis ranged from 64 to 71 yr and the percentage of females ranged between 14% and 22% (Table 1 and Supplementary Table 1). Kaplan-Meier 5-yr risk of recurrence and progression in stage and/or grade ranged between 44% and 79% and between 14% and 32%, respectively.

3.2. Genome-wide scan for RFS

3.2.1. Single SNP associations

Results of the meta-analysis for RFS (total *N* = 3366) are summarized in Table 2, Figure 1, Supplementary Figures 1 and 2, and Supplementary Table 2. The strongest association was found for rs12885353 on chromosome 14 (intron variant of gene *G2E3*) based on a meta-analysis of four out of six cohorts (two cohorts failed the MAF >5% threshold; Tables 2 and 3, and Supplementary Fig. 3 and 4). After removing the MAF threshold to include all cohorts, this SNP reached genome-wide significance (hazard ratio [HR] C vs T 1.55, 95% confidence interval [CI] 1.33–1.82, $p = 4.0 \times 10^{-8}$) as did rs34339578 in the same locus (HR T vs G 1.60, 95% CI 1.35–1.89, $p = 4.8 \times 10^{-8}$; intronic variant of gene *SCFD1*), although moderate to substantial effect heterogeneity was present (Table 3 and Supplementary Fig. 4).

Twelve other loci were suggestively associated with RFS ($p < 1 \times 10^{-5}$; Table 2), of which three were based on high imputation quality and low effect heterogeneity (*GRIN2B*,

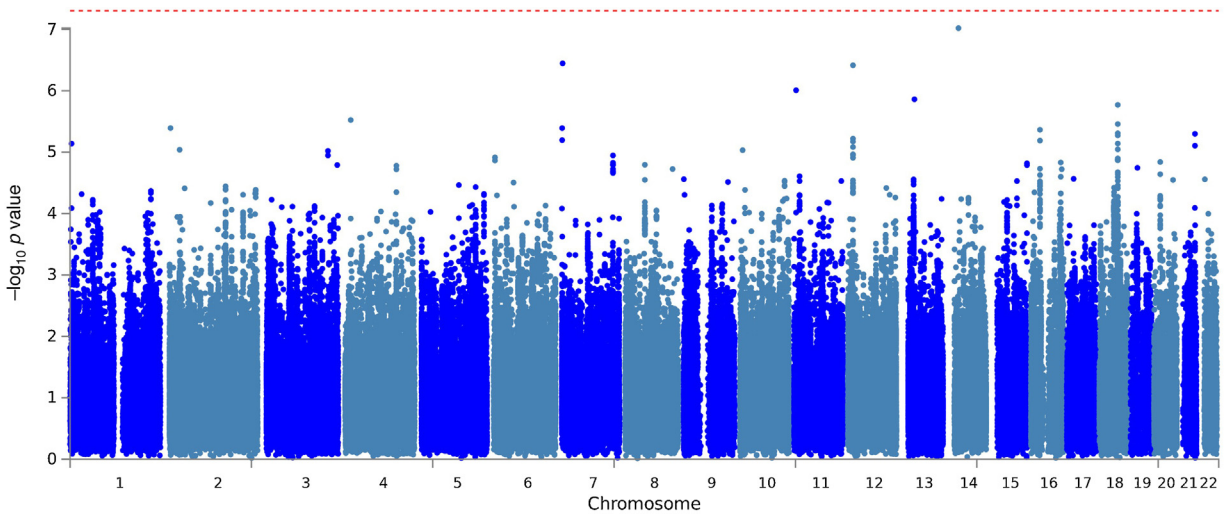

Fig. 1 – Manhattan plot for the meta-analysis association results showing the chromosomal position of genotyped or imputed SNPs plotted against the $-\log_{10}$ *p* value of their association with RFS. SNP associations based on meta-analysis of at least three cohorts, directional consistency in effect estimates between cohorts, and a cohort-specific MAF of >5% are included. The horizontal red dotted line indicates the threshold for genome-wide statistical significance ($p = 5 \times 10^{-8}$). MAF = minor allele frequency; RFS = recurrence-free survival; SNP = single nucleotide polymorphism.

Table 3 – Genome-wide significant associations for SNPs on chromosome 14 with RFS after reduction of the MAF threshold to 1%

SNP	CHR	BP	A1	A2	Direction [a]	HR	95% CI	*p* value	MAF	*N*	Info [b]	Het I^2 (*p* value) [c]	Gene
rs12885353	14	31078574	C	T	++++++	1.55	1.33–1.82	4.0×10^{-8}	0.047–0.07 (C)	3366	>0.89	65.8 (0.01)	*G2E3*
rs34339578	14	31139335	T	G	++++++	1.60	1.35–1.89	4.8×10^{-8}	0.040–0.07 (T)	3366	>0.88	55.0 (0.05)	*SCFD1*

A1 = effect allele; A2 = reference allele; BCPP = Bladder Cancer Prognosis Programme (Birmingham, UK); BP = basepair position; CHR = chromosome; CI = confidence interval; GUB = Genito-Urinary BioBank (Toronto, Canada); Het = heterogeneity; HR = hazard ratio; MAF = minor allele frequency; *N* = number; NBCS = Nijmegen Bladder Cancer Study (Nijmegen, The Netherlands); RFS = recurrence-free survival; SNP = single nucleotide polymorphism.
Basepair positions and annotations are according to GRCh37/hg19 NCBI Reference Sequence collection.
[a] Order of the cohorts: NBCS, GUB-1, GUB-2, BCPP, Barcelona, and Sheffield.
[b] Impute info score, which entails the quality of imputation, with 1 = directly measured SNP and >0.8 for high-quality imputation.
[c] I^2 statistic measures heterogeneity on a scale of 0–100%.

DCC, and *TFF1*) and nine were not (Supplementary Fig. 5–20, and Supplementary Tables 3 and 4).

Lead SNP association results stratified by a high/low-risk profile at the time of diagnosis showed similar effect estimates, although point estimates of effect sizes were somewhat higher (that is, indicated a stronger effect) in the high-risk group for five out of 13 lead SNPs (Supplementary Table 5).

3.2.2. Gene prioritization

None of the lead SNPs was an eQTL in NMIBC or MIBC tissue, but eQTL analyses in whole blood revealed SNPs in three loci (chromosome 14, 16, and 21) as eQTLs of six genes: *SCFD1*, *SLC5A11*, *TMPRSS3*, *ARHGAP17*, *SLC37A1*, and *TNRC6A* (Supplementary Table 6).

Two lead SNPs were found to affect TFBSs: rs7329778 affects the binding site of FOXH1 and ZNF354C, and rs3808347 affects the binding site of ZNF263 and SP1.

No new candidate genes were discovered based on chromatin interaction mapping.

3.2.3. Gene-based and gene-set analyses

The gene-based analysis did not reveal any statistically significant associations (Supplementary Table 7, and Supplementary Fig. 21 and 22). The strongest association was observed for *HIVEP2* on chromosome 6 ($p = 6.7 \times 10^{-5}$). Gene-based association results for prioritized candidate genes revealed that all but six genes and the transcription factors reached nominal significance ($p < 0.05$), with the strongest association for *DCC* ($p = 1.55 \times 10^{-4}$; Supplementary Table 8).

The gene-set analysis revealed no gene sets to be suggestively associated with RFS (Supplementary Table 9).

3.2.4. Association of tumor gene expression of prioritized genes with RFS

Tumor expression of three out of the 20 prioritized genes (*SCFD1*, *ARHGAP17*, and *TNRC6A*) analyzed was associated with RFS according to $p_{FDR} < 0.05$ (Supplementary Table 10). The strongest association was identified for *SCFD1*, which demonstrated a reduced recurrence risk for higher expression levels (HR 0.70, 95% CI 0.59–0.84, $p_{FDR} = 0.003$), corroborating the directions of the effect of the meta-GWAS and eQTL analyses. Increased tumor expression levels of *ARHGAP17* and *TNRC6A* were associated with an increased recurrence risk, but directions of effect of lead SNPs and eQTL analyses did not match.

Table 4 – Top associated loci ($p < 1 \times 10^{-5}$) for PFS summarized by a lead index SNP defined as the SNP with the smallest association *p* value in the region

SNP	CHR	BP	A1	A2	Direction [a]	HR	95% CI	*p* value	MAF	*N*	Info [b]	Het I^2 (*p* value) [c]	(Nearest) Gene(s) [d]
rs76607989 [e]	8	39329129	C	T	?+??++	3.06	2.05–4.56	4.2×10^{-8}	0.05–0.07 (C)	834	>0.70	0 (0.75)	*ADAM3A*
rs16847917	2	213080786	T	C	??+?++	2.76	1.88–4.04	1.9×10^{-7}	0.05–0.06 (T)	913	>0.98	53.6 (0.12)	*ERBB4*
rs113601380	12	19843410	T	TTGAG	++?+?+	1.96	1.51–2.52	2.7×10^{-7}	0.06–0.07 (T)	2731	>0.95	14.2 (0.32)	*#AEBP2*
rs754149	10	2531999	G	C	------	0.71	0.62–0.82	2.0×10^{-6}	0.41–0.44 (G)	3400	>0.97	0 (0.88)	*LINC02645* [f]
rs2065281	21	31871307	A	G	++++++	1.39	1.21–1.59	2.1×10^{-6}	0.27–0.32 (A)	3400	>0.97[&]	0 (0.72)	*KRTAP19-4*
rs28677138	7	155679499	T	C	++++++	1.61	1.32–1.97	2.6×10^{-6}	0.09–0.13 (T)	3400	>0.93	0 (0.48)	*#SHH*
rs140189706	16	1594376 [g]	T	C	------	0.57	0.45–0.72	3.3×10^{-6}	0.06–0.07 (C)	3400	>0.92	15.2 (0.32)	*TMEM204 & IFT140*
rs931105	11	80021523	A	G	++++++	1.39	1.21–1.61	6.6×10^{-6}	0.32–0.39 (G)	3400	>0.99[&]	0 (0.44)	*#RNU6-544P* [h]
rs67816797	16	1757001[g]	C	T	++++++	1.63	1.31–2.02	8.7×10^{-6}	0.07–0.08 (C)	3400	>0.97	0 (0.55)	*MAPK8IP3*
rs11151504	18	67028817	T	C	++++++	1.63	1.31–2.03	9.5×10^{-6}	0.07–0.09 (T)	3400	>0.86	49.4 (0.08)	*#DOK6*

A1 = effect allele; A2 = reference allele; BCPP = Bladder Cancer Prognosis Programme (Birmingham, UK); BP = basepair position; CHR = chromosome; CI = confidence interval; GUB = Genito-Urinary BioBank (Toronto, Canada); Het = heterogeneity; HR = hazard ratio; LINC = long intergenic or intronic non–protein coding RNA; MAF = minor allele frequency; *N* = number; NBCS = Nijmegen Bladder Cancer Study (Nijmegen, The Netherlands); PFS = progression-free survival; SNP = single nucleotide polymorphism.

SNPs with a cohort-specific MAF of >5% were included in the meta-analysis. This table contains all top-ranked SNPs that are based on meta-analysis of at least three cohorts and directional consistency in effect estimates between cohorts. Only the strongest signal per locus is shown (for the complete list of SNPs associated with PFS with $p < 1 \times 10^{-5}$, see Supplementary Table 12).

Basepair positions and annotations are according to GRCh37/hg19 NCBI Reference Sequence collection.

[a] Order of the cohorts: NBCS, GUB-1, GUB-2, BCPP, Barcelona, and Sheffield.

[b] Impute info score, which entails the quality of imputation, with 1 = directly measured SNP and >0.8 for high-quality imputation. For values indicated with an &, these SNPs were directly genotyped in NBCS, GUB-1, GUB-2, and BCPP.

[c] I^2 statistic measures heterogeneity on a scale of 0–100%.

[d] Genes indicated with a # are the nearest genes. All SNPs in this table that reside within a gene are intronic variants except for rs2065281 (upstream variant).

[e] This association became less strong after meta-analysis of all six cohorts due to an inconsistent direction of effect for this SNP (and correlated SNPs in the region) in NBCS as compared with the other cohorts (Supplementary Tables 13 and 14, and Supplementary Fig. 27 and 28).

[f] Nearest protein coding gene is *PFKP*.

[g] These two SNPs are part of the same association signal that spans a large region containing many genes, see Supplementary Figure 40.

[h] RNU6-544P is a pseudogene. Nearest protein coding gene is *TENM4*.

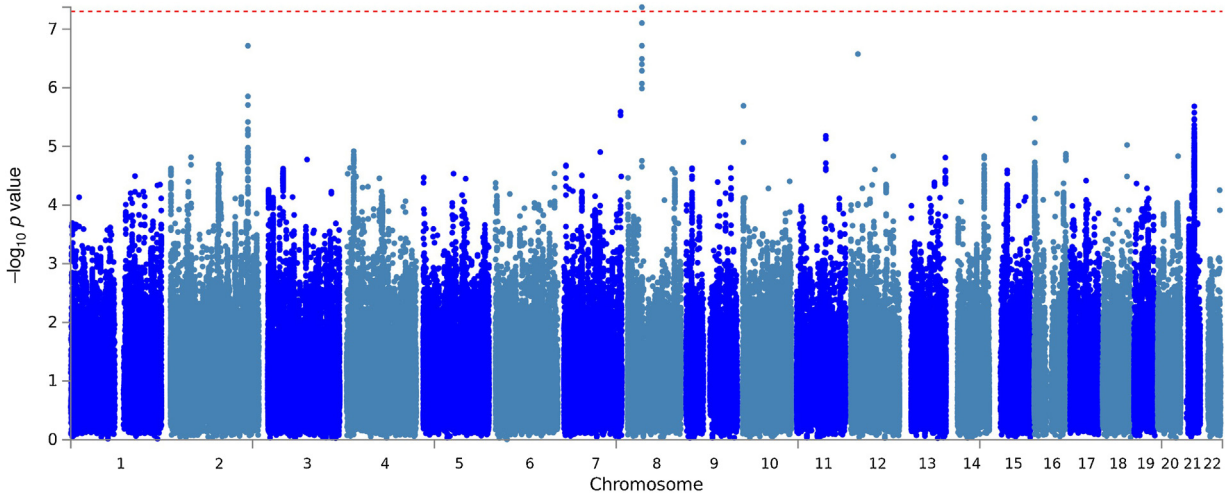

Fig. 2 – Manhattan plot for the meta-analysis association results showing the chromosomal position of genotyped or imputed SNPs plotted against the $-\log_{10}$ *p* value of their association with PFS. SNP associations based on meta-analysis of at least three cohorts, directional consistency in effect estimates between cohorts and a cohort-specific MAF of >5% are included. The horizontal red dotted line indicates the threshold for genome-wide statistical significance ($p = 5 \times 10^{-8}$). MAF = minor allele frequency; PFS = progression-free survival; SNP = single nucleotide polymorphism.

3.2.5. Replication of previously reported loci

Only rs3795617 (residing in *RGS13*) out of the ten previously reported SNPs for RFS in NMIBC reached nominal significance (p = 0.03; Supplementary Table 11).

3.3. Genome-wide scan for PFS

3.3.1. Single SNP associations

Results of the meta-analysis for PFS (total N = 3400) are summarized in Table 4, Figure 2, Supplementary Figures 23 and 24, and Supplementary Table 12. SNP rs76607989 on chromosome 8 (intron variant of *ADAM3A*) reached genome-wide significance based on a meta-analysis of three cohorts with MAF >5% (Supplementary Fig. 25 and 26), but this signal disappeared after inclusion of all cohorts, revealing an opposite direction of effect in the NBCS cohort compared with the others (HR C vs T 1.97, 95% CI 1.43–2.72, $p = 3.0 \times 10^{-5}$, heterogeneity I^2 75%; Supplementary Fig. 27, and Supplementary Tables 13 and 14).

Nine other loci reached the threshold for suggestive evidence of association ($p < 1 \times 10^{-5}$), of which eight were based on high imputation quality and no effect heterogeneity (Table 4, Supplementary Table 15, and Supplementary Fig. 28–39) and one was not (rs11151504; Supplementary Fig. 40).

Risk stratification into low- and high-risk profiles at the time of diagnosis showed similar effect estimates for the lead SNPs, although point estimates of effect sizes were somewhat higher (ie, indicated a stronger effect) in the low-risk group for five out of ten lead SNPs (Supplementary Table 16).

3.3.2. Gene prioritization

SNP rs67816797 was identified as an eQTL for *UBE2I* in MIBC, but the eQTL analysis of NMIBC data did not identify any additional candidate genes. The eQTL analyses in whole blood revealed SNPs in loci on chromosomes 8 and 16 to influence expression of 13 genes, including *IFT140, FAHD1*, and *NME3* (Supplementary Table 17).

Two SNPs were found to affect TFBSs: rs76607989 was found to affect the binding site of HSF1, ZNF263, and SP1, and rs11151504 was found to affect the binding site of MZF1_1-4.

No new candidate genes were discovered based on chromatin interaction mapping.

3.3.3. Gene-based and gene-set analyses

No statistically significant associations were found (Supplementary Table 18, and Supplementary Fig. 41 and 42). Gene *KRTAP19-6* on chromosome 21 showed the strongest association ($p = 3.93 \times 10^{-5}$), but this association was based on one SNP only (rs1023364). Gene-based association results for 36 prioritized candidate genes revealed 20 genes that reached nominal significance (Supplementary Table 19).

The gene-set analysis revealed one GO gene set suggestively associated with PFS: the ATP-binding cassette (ABC) transporter complex ($p = 5.9 \times 10^{-6}$; Supplementary Table 20).

3.3.4. Association of tumor gene expression of prioritized genes with PFS

Tumor gene expression of nine of the prioritized genes for PFS was associated with PFS according to $p_{FDR} < 0.05$ (Supplementary Table 21). The strongest association was found for *IFT140* (HR 0.72, 95% CI 0.62–0.83, $p_{FDR} = 5.60 \times 10^{-5}$), and directions of effect were consistent with meta-GWAS and eQTL results. The same holds for *UBE2I*, *FAHD1*, and *NME3*, but not for *KRT17*, *ZNF263*, *MRPS34*, *SHH*, and *ADAM9* (directional inconsistencies between results of meta-GWAS, eQTL analysis, gene expression-PFS correlations, and/or previously published effects in literature).

3.3.5. Replication of previously reported loci

Two SNPs (rs12628 in *HRAS* and rs11585883 residing between *RGS4* and *RGS5*) of the eight previously reported SNPs for PFS in NMIBC reached nominal significance (p = 0.046 and 0.034, respectively; Supplementary Table 22).

4. Discussion

We performed the first meta-analysis of GWAS to date for recurrence and progression in NMIBC using six cohorts from Europe and Canada (total N = 3400). Out of >7 500 000 SNPs tested, we have identified several common genetic variants that are reproducibly associated with RFS and PFS. These associations were not influenced by risk stratification into low- and high-risk profiles based on stage and grade at diagnosis. In addition, none of the top associated loci were found to be involved in both RFS and PFS, and none of the SNPs that were previously reported in literature for RFS and PFS reached the suggestive threshold of significance in our meta-GWAS. Through several gene prioritization strategies, we have selected a total of 54 candidate genes, including six transcription factors for RFS and PFS. NMIBC tumor expression levels of one and four of these genes were directionally consistently associated with RFS and PFS, respectively, thereby providing strong support functional relevance (Tables 5 and 6).

The strongest evidence for RFS was found for gene *SCFD1* (Sec1 family domain-containing protein 1) on chromosome 14, also known as *SLY1* (Table 5). *SCFD1* is known for its role in intra-Golgi transport, endoplasmic reticulum (ER) to Golgi transport, and Golgi-to-ER retrograde transport [17,18]. In addition, it was shown that the SLY1 protein is a potential contributor to protein trafficking response to cellular stress in neuroblastoma cells and that suppression of SLY1 is associated with accelerated apoptosis [19]. Finally, *SCFD1* belongs to the same family of signal adapter proteins as gene *SASH1*, a tumor suppressor gene in, among others, breast cancer, ovarian cancer, and colon cancer [20–22].

Four genes [*IFT140*, *UBE2I*, *FAHD1*, and *NME3*] showed strong evidence for association with PFS (Table 6). Of these, *UBE2I* (ubiquitin-conjugating enzyme E2 I, also known as *UBC9*) was very recently shown to play a dual role in bladder

Table 5 – Candidate genes for recurrence-free survival and evidence based on lead SNP annotation and association, eQTL analysis, transcription factor binding site analysis, gene-based analysis, previous associations as recorded in the GWAS catalog, and the association of tumor gene expression with recurrence-free survival in the NMIBC patient cohort from UROMOL.

Gene (location)	Lead SNP + annotation	Lead SNP association	LocusZoom plot	eQTL analysis	Lead SNP influences a TFBS	Gene-based association signal	Previously published gene associations in GWAS catalog	Association with RFS in UROMOL (N = 511 with 348 events)
G2E3 (14q12)	rs12885353; intron variant	Association $p < 5 \times 10^{-8}$; substantial effect heterogeneity	Only a few SNPs in the region show association	–	–	Nominally significant	Body height, mathematical ability, educational attainment, lysophosphatidylcholine 20:3 measurement	–
SCFD1 (14q12)	rs34339578; intron variant	Association $p < 5 \times 10^{-8}$; substantial effect heterogeneity	Only a few SNPs in the region show association	rs34339578, rs12885353 + 1 LD proxy [a] (whole blood)	–	Nominally significant	Amyotrophic lateral sclerosis, blood protein measurement	Strongest association (p_{FDR} = 0.003); direction of effect consistent with lead SNP effect and eQTL results
GRIN2B (12p13.1)	rs17834128; intron variant	High imputation quality (info >0.98); no effect heterogeneity	Strong regional evidence	–	–	Nominally significant	29 traits including cognitive performance, acute myeloid leukemia, and immune response to smallpox	–
DCC (18q21.2)	rs12967544; intron variant	High imputation quality (info >0.99); no effect heterogeneity	Moderate regional evidence	–	–	Top 10 ($p = 1.55 \times 10^{-4}$)	56 traits including smoking initiation, body height, and depression	–
TFF1 (21q22.3)	rs2839488; intron variant	High imputation quality (info >0.95); low effect heterogeneity	Moderate regional evidence	–	–	Top 10 ($p = 1.28 \times 10^{-4}$)	Blood protein levels, pancreatic carcinoma	–
TFF2 (21q22.3)	rs2839488; downstream flanking gene (at ~15 kb)			–	–	Nominally significant	Spatial memory, pancreatic cancer	–
TFF3 (21q22.3)	rs2839488; second downstream flanking gene (at ~50 kb)			–	–	Nominally significant	Parental longevity, spatial memory	–
TMPRSS3 (21q22.3)	rs2839488; upstream flanking gene (at ~6 kb)			rs2839488 + 1 LD proxy (whole blood)	–	Nominally significant	Blood protein levels, diverticular disease, face memory	–
SLC37A1 (21q22.3)	rs2839488; downstream gene (at ~134 kb)		Gene resides outside of association region (separated by a recombination hot spot)	rs2839488 + 1 LD proxy (whole blood)	–	–	Posterior-cingulate cortex volume, educational attainment	–
HIVEP2 (6q24.2)	NA	NA	NA	–	–	Strongest association ($p = 6.7 \times 10^{-5}$)	19 traits, including body height, asthma, cardiovascular disease	–
MAD1L1 (7p22.3)	rs192039210; intron variant	Low imputation quality (info 0.63); substantial effect heterogeneity; meta-analysis based on four cohorts (SNP not present in cohorts of Barcelona and Sheffield)	Very weak regional evidence	–	–	–	55 traits, including smoking status, prostate carcinoma, and testicular carcinoma	–

Table 5 (*Continued*)

Gene (location)	Lead SNP + annotation	Lead SNP association	LocusZoom plot	eQTL analysis	Lead SNP influences a TFBS	Gene-based association signal	Previously published gene associations in GWAS catalog	Association with RFS in UROMOL (N = 511 with 348 events)
SFMBT2 (10p14)	rs7091482; intron variant	Moderate imputation quality (info 0.82–0.97); no effect heterogeneity; deviating direction of effect in Barcelona cohort	Moderate regional evidence	–	–	–	7 traits, including estrogen receptor–positive breast cancer, smoking behavior	–
ADAP1 (7p22.3)	rs3808347; intron variant	Moderate imputation quality (info 0.80–0.95); no effect heterogeneity	Moderate-weak regional evidence	–	–	Nominally significant	Facial morphology	–
ARHGAP17 (16p12.1)	rs9935790; downstream gene (at ~40 kb)	Moderate imputation quality (info 0.85); no effect heterogeneity	Contradictory regional evidence (SNPs strongly correlated with the lead SNP show no association)	32 LD proxies of rs9935790 (whole blood)	–	–	4 traits: blood metabolite levels, urinary metabolites, appendicular lean mass, and chronic kidney disease	Significant association (p_{FDR} = 0.013); directions of effect of lead SNP and eQTL analyses do not match
SLC5A11 (16p12.1)	rs9935790; upstream gene (at ~143 kb)			rs9935790 and 47 LD proxies (whole blood)	–	–	5 traits, including leukocyte count and chronic kidney disease	–
TNRC6A (16p12.1)	rs9935790; upstream gene (at ~229 kb)			17 LD proxies of rs9935790 (whole blood)	–	–	31 traits, including alcohol consumption, smoking status, metabolic syndrome, and acute myeloid leukemia	Significant association (p_{FDR} = 0.033); directions of effect of lead SNP and eQTL analyses do not match
FOXH1 [b] (8q24.3)	NA	NA	NA	–	rs7329778	–	7 traits, including smoking status measurement and intelligence	–
ZNF354C [b] (5q35.3)	NA	NA	NA	–	rs7329778	–	–	–
ZNF263 [b] (16p13.3)	NA	NA	NA	–	rs3808347	–	Lung function (FVC)	–
SP1 [b] (12q13.13)	NA	NA	NA	–	rs3808347	–	19 traits, including erythrocyte count, HDL cholesterol, and neutrophil count	–

eQTL = expression quantitative trait locus; FDR = false discovery rate; FVC = forced vital capacity; GWAS = genome-wide association study; HDL = high-density lipoprotein; LD = linkage disequilibrium; NA = not applicable; RFS = recurrence-free survival; SNP = single nucleotide polymorphism; TFBS = transcription factor binding site.
Lead SNPs were defined as the SNPs with the smallest association p value in top associated loci (regions with at least one SNP with $p < 1 \times 10^{-5}$).
Chromatin interaction mapping is not included in the table, as it revealed no candidate genes.
[a] LD proxy is defined as an SNP with $r^2 > 0.8$ with a lead SNP.
[b] Transcription factors that were identified based on lead SNPs influencing their binding site.

Table 6 – Candidate genes for progression-free survival and evidence based on lead SNP annotation and association, eQTL analysis, transcription factor binding site analysis, gene-based analysis, previous associations as recorded in the GWAS catalog, and the association of tumor gene expression with progression-free survival in the NMIBC patient cohort from UROMOL.

Gene (location)	Lead SNP + annotation	Lead SNP association	LocusZoom plot	eQTL analysis	Lead SNP influences a TFBS	Gene-based association signal	Previously published gene associations in GWAS catalog	Association with PFS in UROMOL (N = 530 with 65 events)
ADAM3A (8p11.22)	rs76607989; intron variant	Reasonable imputation quality (info >0.70); high effect heterogeneity; deviating direction of effect in NBCS	Strong regional evidence	–	–	–	–	–
ADAM9 (8p11.22)	rs76607989; upstream gene (at ~366 kb)			2 LD proxies [a] of rs76607989 (whole blood)	–	–	–	Significant association (p_{FDR} = 0.015); inconsistent directions between lead SNP effect, eQTL analysis results, correlation of gene expression with PFS, and effects in literature
ERBB4 (2q34)	rs16847917; intron variant + rs10207206; intron variant	High imputation quality (info >0.98); high effect heterogeneity for rs16847917 due to deviating direction of effect in GUB-1; no effect heterogeneity for rs10207206	Moderate regional evidence	rs16847917 (lung adenocarcinoma)	–	Nominally significant	48 traits, including smoking initiation and breast cancer	–
AEBP2 (12p12.3)	rs113601380; intron variant	High imputation quality (>0.95); low effect heterogeneity	Contradictory regional evidence (SNPs strongly correlated with the lead SNP show no association)	–	–	–	10 traits, including lung function and lifetime average cigarettes per day in chronic obstructive pulmonary disease	–
SHH (7q36.3)	rs28677138; downstream flanking gene (at ~74 kb), but separated by a recombination hot spot	High imputation quality (>0.93); no effect heterogeneity	Moderate regional association; SHH is separated from the association region by a recombination hot spot	–	–	Nominally significant	10 traits, including creatinine and lung function	Significant association (p_{FDR} = 2.6 × 10^{-4}); inconsistent direction of effect between correlation of gene expression with PFS and effects in literature
TMEM204 (16p13.3)	rs140189706; intron variant	High imputation quality (0.93); low effect heterogeneity	Strong regional evidence; however, some SNPs that are strongly correlated with the lead SNP show no association	rs67816797, rs140189706 (breast cancer and whole blood); 140 LD proxies of rs140189706 and rs67816797 (whole blood)	–	Nominally significant	Coronary artery disease	–
IFT140 (16p13.3)	rs140189706; intron variant			133 LD proxies of rs140189706 and rs67816797 (whole blood)	–	Nominally significant	Coronary artery disease	Strongest association (p_{FDR} = 5.4 × 10^{-5}); direction of effect consistent with lead SNP effect and eQTL results
TELO2 (16p13.3)	rs140189706; upstream gene (at ~34 kb)			136 LD proxies of rs140189706 and rs67816797 (whole blood)	–	–	Rheumatoid arthritis	–
UBE2I (16p13.3)	rs140189706; upstream gene (at ~217 kb)			rs67816797 (bladder urothelial carcinoma)	–	–	Systolic blood pressure, white blood cell count, refractive error	Significant association (p_{FDR} = 0.039); direction of effect consistent with lead SNP effect and eQTL results
BAIAP3 (16p13.3)	rs140189706; upstream gene (at ~194 kb)			95 LD proxies of rs140189706 and rs67816797 (whole blood)	–	–	–	–
CLCN7 (16p13.3)	rs140189706; downstream gene (at ~69 kb)			131 LD proxies of rs140189706 and rs67816797 (whole blood)	–	–	–	–

Table 6 (*Continued*)

Gene (location)	Lead SNP + annotation	Lead SNP association	LocusZoom plot	eQTL analysis	Lead SNP influences a TFBS	Gene-based association signal	Previously published gene associations in GWAS catalog	Association with PFS in UROMOL (*N* = 530 with 65 events)
MAPK8IP3 (16p13.3)	rs67816797; intron variant	High imputation quality (0.97); no effect heterogeneity	Strong regional evidence; however, some SNPs that are strongly correlated with the lead SNP show no association	rs67816797, rs140189706 (head and neck squamous cell carcinoma); 61 LD proxies of rs140189706 and rs67816797 (whole blood)	–	Nominally significant	Intraocular pressure, C-reactive protein	–
CRAMP1L (16p13.3)	rs67816797; upstream gene (at ~29 kb)			139 LD proxies of rs140189706 and rs67816797 (whole blood)	–	Top 10 ($p = 1.8 \times 10^{-4}$)	–	Not tested [b]
FAHD1 (16p13.3)	rs67816797; downstream gene (at ~120 kb)			133 LD proxies of rs140189706 and rs67816797 (whole blood)	–	–	Ankle injury	Significant association ($p_{FDR} = 5.4 \times 10^{-5}$); direction of effect consistent with lead SNP effect and eQTL results
JPT2 (=HN1L) (16p13.3)	rs67816797; upstream gene (at ~4 kb)			129 LD proxies of rs140189706 and rs67816797 (whole blood)	–	Top 10 ($p = 6.4 \times 10^{-5}$)	Intraocular pressure, snoring	–
IGFALS (16p13.3)	rs67816797; upstream gene (at ~83 kb)			3 LD proxies of rs140189706 and rs67816797 (whole blood)	–	–	–	Not tested [b]
MRPS34 (16p13.3)	rs67816797; upstream gene (at ~65 kb)			35 LD proxies of rs140189706 and rs67816797 (whole blood)	–	–	–	Significant association (p_{FDR} = 0.0085); inconsistent directions between lead SNP effect, eQTL analysis results, and correlation of gene expression with PFS
NME3 (16p13.3)	rs67816797; upstream gene (at ~63 kb)			132 LD proxies of rs140189706 and rs67816797 (whole blood)	–	–	–	Significant association (p_{FDR} = 0.009); direction of effect consistent with lead SNP effect and eQTL results
SLC9A3R2 (16p13.3)	rs67816797; downstream gene (at ~320 kb)			rs67816797, rs140189706 (thyroid carcinoma)	–	–	6 traits, including systolic blood pressure and cardiovascular disease	–
KRTAP19-1, KRTAP19-6, KRTAP19-3, KRTAP19-2, KRTAP13-4, KRTAP19-4, KRTAP8-1, KRT71, KRTAP15-1, KRT17, KRTAP6-3, KRTAP4-11, KRTAP13-2 (21q22.11)	rs2065281: downstream flanking gene is *KRTAP19-4* (at ~2 kb) and upstream flanking gene is *KRTAP19-5* (at ~3 kb)	Directly genotyped in NBCS, GUB-1, GUB-2, and BCPP; high imputation quality in Barcelona and Sheffield (>0.97); no effect heterogeneity	Very strong regional association	KRTAP13–2: rs2065281 (prostate adenocarcinoma)	–	KRTAP19-6 is the strongest association ($p = 3.9 \times 10^{-5}$); KRTAP-19-1, KRTAP-19-3, and KRTAP-13-4 are among the top 10 ($p = 2.6 \times 10^{-4}$, 3.2×10^{-4}, and 3.3×10^{-4}, respectively)	Only KRTAP19-8: PR interval in *Tripanosoma cruzi* seropositivity; response to platinum-based chemotherapy (cisplatin) and 3-hydroxypropylmercapturic acid levels in smokers	Significant association for KRT17 (p_{FDR} = 0.0014); inconsistent direction of effect between correlation of gene expression with PFS and effects in literature. No results for KRTAP19-6, KRTAP19-2, KRTAP19-4, KRTAP8-1, and KRTAP4-11 [b]
HSF1 [c] (8q24.3)	NA	NA	NA	–	rs76607989	–	Heel bone mineral density, type 2 diabetes	–
ZNF263 [c] (16p13.3)	NA	NA	NA	–	rs76607989	–	Lung function (FVC)	Significant association ($p_{FDR} = 5.4 \times 10^{-5}$); inconsistent direction of effect between correlation of gene expression with PFS and effects in literature

Table 6 (*Continued*)

Gene (location)	Lead SNP + annotation	Lead SNP association LocusZoom plot	eQTL analysis	Lead SNP influences a TFBS	Gene-based association signal	Previously published gene associations in GWAS catalog	Association with PFS in UROMOL (N = 530 with 65 events)
SP1 [c] (12q13.13)	NA	NA	NA	–	rs76607989	Nominally significant	19 traits, including erythrocyte count and body mass index	–
MZF1/ MZF1_1-4 [c] (19q13.43)	NA	NA	NA	–	rs11151504	–	–	–

BCPP = Bladder Cancer Prognosis Programme (Birmingham, UK); eQTL = expression quantitative trait locus; FDR = false discovery rate; FVC = forced vital capacity; GUB = Genito-Urinary BioBank (Toronto, Canada); GWAS = genome-wide association study; LD = linkage disequilibrium; NA = not applicable; NBCS = Nijmegen Bladder Cancer Study (Nijmegen, The Netherlands); PFS = progression-free survival; SNP = single nucleotide polymorphism; TFBS = transcription-factor binding site.
Lead SNPs were defined as the SNPs with the smallest association p value in top associated loci (regions with at least one SNP with $p < 1 \times 10^{-5}$).
Chromatin interaction mapping is not included in the table as it revealed no candidate genes.
[a] LD proxy is defined as an SNP with $r^2 > 0.8$ with a lead SNP.
[b] Not tested, expression levels were too low (log expression [log cpm] below 2).
[c] Transcription factors that were identified based on lead SNPs influencing their binding site.

cancer cells: its expression is required to maintain high sumoylation levels and to maintain homeostasis and survival in cancer cells, while a lack of UBC9 contributes to spectacular inflammation activation, which promotes cancer progression via epithelial-mesenchymal transition and stem cell–like population formation [23]. Here, we found a protective effect of higher expression of *UBE2I* on the risk of NMIBC progression.

Many of the other identified candidate genes have previously been implicated in bladder cancer or other types of cancer and tumorigenesis (Supplementary Tables 23 and 24, respectively), and/or in prognostic cancer outcomes (Supplementary Table 25). Finally, our results point toward several transcription factors, most of which have previously been implicated in bladder cancer or other cancer types (Supplementary Table 26).

Our study has some important strengths. We used the largest collection of NMIBC cases to date with both genome-wide SNP and outcome data available. We performed strict quality control procedures in our statistical analyses and used the same cleaning and analysis pipeline for all cohorts. We reduced the likelihood of false positives by combining GWAS results from six independent cohorts. Finally, we included a functional validation step by testing associations between expression of prioritized candidate genes in NMIBC tumors and RFS and PFS in an additional independent cohort.

Some study limitations exist, however. The included cohorts are all from European ancestry, and our results might therefore not be applicable to other populations. In addition, definitions used for RFS and PFS differ slightly between cohorts, which might have resulted in heterogeneity in effect estimates between cohorts and false-negative findings. Similar is the case for treatment guidelines between hospitals and countries. Owing to a lack of detailed information on (adequacy of) treatment, we did not adjust for treatment or focus on identification of single nucleotide variants that are relevant for specific treatment subgroups only. We also used a broad definition of progression that included progression in grade and stage, as well as metastatic progression. Furthermore, it should be emphasized that only two of our lead SNPs reached genome-wide statistical significance. Although the reported associations are based on a meta-analysis of six cohorts and can therefore be considered as being statistically replicated to some extent, there is still a chance that the reported associations are false-positive findings. Similarly, we might have found false-negative results, as our study suffered from limited power to detect SNPs with small effect sizes and/or a low MAF. Assuming a recurrence risk of 49% (the average 5-yr recurrence risk among the six cohorts), we had >80% power at $\alpha = 5 \times 10^{-8}$ to detect SNPs with an allelic HR of 1.3 and a MAF of at least 0.2, but we could detect only SNPs with an HR of at least 1.6 for a MAF of 0.05. For an assumed progression risk of 20%, the effect sizes we could detect are larger. Finally, a prevalent case bias might play a role in our study, as time between NMIBC diagnosis and invitation to the study was relatively long for some cohorts. This may have resulted in an NMIBC study population with relatively more favorable survival.

5. Conclusions

We identified several SNPs that are reproducibly associated with RFS and PFS. These loci point toward 54 candidate genes including six transcription factors. For five of these genes, we could confirm directionally consistent associations between tumor gene expression and RFS and PFS outcomes in an additional independent NMIBC cohort. The putative candidate genes can be used as a resource for future functional studies into biological mechanisms of recurrence and progression development in NMIBC.

***Author contributions*:** Tessel E. Galesloot had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Kiemeney, Vermeulen.
Acquisition of data: Aben, Bryan, Catto, Cheng, Conroy, Dyrskjøt, Fleshner, James, Malats, Mengual, Verhaegh, Zeegers, Kiemeney, Vermeulen.
Analysis and interpretation of data: Galesloot, Grotenhuis, Lamy, Lindskrog, Kiemeney, Vermeulen.
Drafting of the manuscript: Galesloot.
Critical revision of the manuscript for important intellectual content: Grotenhuis, Kolev, Aben, Bryan, Catto, Cheng, Dyrskjøt, Fleshner, James, Lamy, Lindskrog, Malats, Mengual, Verhaegh, Zeegers, Kiemeney, Vermeulen.
Statistical analysis: Galesloot, Grotenhuis, Lamy, Lindskrog, Kolev.
Obtaining funding: None.
Administrative, technical, or material support: None.
Supervision: Kiemeney, Vermeulen.
Other: None.

Financial disclosures: Tessel E. Galesloot certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: James W.F. Catto has received reimbursement for consultancy from Astra Zeneca, Ferring, Roche, and Janssen; speaker fees from BMS, MSD, Nucleix, and Roche; and honoraria for membership of advisory boards for Ferring, Roche, BMS, QED Therapeutics, and Janssen. Richard T. Bryan has contributed to advisory boards for Olympus Medical Systems and Janssen, UroGen Pharma, and QED Therapeutics. Lars Dyrskjøt has sponsored research agreements with C2i-genomics, Natera, AstraZeneca, and Ferring, and has received honoraria for advisory/consulting role at Ferring and speaker fee from Roche.

***Funding/Support and role of the sponsor*:** The University of Sheffield collection was partly funded by Yorkshire Cancer Research (program grant number S305PA: Genetic instability and death in cancer cells). James W.F. Catto is funded by an NIHR research professorship. Sita H. Vermeulen is supported by a grant from the Netherlands Organization for Scientific Research (NWO Vidi 91717334). The University of Birmingham collection was funded by Cancer Research UK (C1343/A5738), the University of Birmingham, and the Birmingham and The Black Country and West Midlands North and South Comprehensive Local Research Networks, UK. UROMOL sample genotyping was funded by EU-FP7-HEALTH-F2-2008 #201663 and by the Spanish ISCIII-FIS #PI18/01347 grants.

Acknowledgements: This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. We would like to thank deCODE genetics for genotyping the samples. We also thank the registration team of the Netherlands Comprehensive Cancer Organisation (IKNL) for the collection of data for the Netherlands Cancer Registry as well as IKNL staff for scientific advice.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.euo.2021.07.001.

References

[1] Cumberbatch MGK, Jubber I, Black PC, et al. Epidemiology of bladder cancer: a systematic review and contemporary update of risk factors in 2018. Eur Urol 2018;74:784–95.
[2] Sylvester RJ, van der Meijden AP, Oosterlinck W, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. Eur Urol 2006;49, 466–5; discussion 475–477.
[3] Sylvester RJ, Rodriguez O, Hernandez V, et al. European Association of Urology (EAU) prognostic factor risk groups for non-muscle-invasive bladder cancer (NMIBC) incorporating the WHO 2004/2016 and WHO 1973 classification systems for grade: an update from the EAU NMIBC Guidelines Panel. Eur Urol 2021;79:480–8.
[4] Leal J, Luengo-Fernandez R, Sullivan R, Witjes JA. Economic burden of bladder cancer across the European Union. Eur Urol 2016;69:438–47.
[5] Chen M, Hildebrandt MA, Clague J, et al. Genetic variations in the sonic hedgehog pathway affect clinical outcomes in non-muscle-invasive bladder cancer. Cancer Prev Res (Phila) 2010;3:1235–45.
[6] Lipunova N, Wesselius A, Cheng KK, et al. Systematic review: genetic associations for prognostic factors of urinary bladder cancer. Biomark Cancer 2019;11, 1179299X19897255.
[7] Grotenhuis AJ, Dudek AM, Verhaegh GW, et al. Independent replication of published germline polymorphisms associated with urinary bladder cancer prognosis and treatment response. Bladder Cancer 2016;2:77–89.
[8] Wu C, Kraft P, Stolzenberg-Solomon R, et al. Genome-wide association study of survival in patients with pancreatic adenocarcinoma. Gut 2014;63:152–60.
[9] Li W, Middha M, Bicak M, et al. Genome-wide scan identifies role for AOX1 in prostate cancer survival. Eur Urol 2018;74:710–9.
[10] Rizvi A, Karaesmen E, Morgan M, et al. gwasurvivr: an R package for genome-wide survival analysis. Bioinformatics 2019;35:1968–70.
[11] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1826.
[12] Gong J, Mei S, Liu C, et al. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. Nucleic Acids Res 2018;46:D971–6.
[13] Lindskrog SV, Prip FF, Lamy P, et al. An integrated multi-omics analysis identifies clinically relevant molecular subtypes of non-muscle-invasive bladder cancer. Nat Commun 2021;12:2301.
[14] Kumar S, Ambrosini G, Bucher P. SNP2TFBS—a database of regulatory SNPs affecting predicted transcription factor binding site affinity. Nucleic Acids Res 2017;45:D139–44.

[15] de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11:e1004219.
[16] Lenfant L, Cancel-Tassin G, Gazut S, et al. Genetic variability in 13q33 and 9q34 is linked to aggressiveness patterns and a higher risk of progression of non-muscle-invasive bladder cancer at the time of diagnosis. BJU Int 2021;127:375–83.
[17] Laufman O, Kedan A, Hong W, Lev S. Direct interaction between the COG complex and the SM protein, Sly1, is required for Golgi SNARE pairing. EMBO J 2009;28:2006–17.
[18] Dascher C, Balch WE. Mammalian Sly1 regulates syntaxin 5 function in endoplasmic reticulum to Golgi transport. J Biol Chem 1996;271:15866–9.
[19] Bando Y, Katayama T, Taniguchi M, et al. RA410/Sly1 suppresses MPP + and 6-hydroxydopamine-induced cell death in SH-SY5Y cells. Neurobiol Dis 2005;18:143–51.
[20] Rimkus C, Martini M, Friederichs J, et al. Prognostic significance of downregulated expression of the candidate tumour suppressor gene *SASH1* in colon cancer. Br J Cancer 2006;95:1419–23.
[21] Ren X, Liu Y, Tao Y, et al. Downregulation of SASH1 correlates with tumor progression and poor prognosis in ovarian carcinoma. Oncol Lett 2016;11:3123–30.
[22] Burgess JT, Bolderson E, Saunus JM, et al. SASH1 mediates sensitivity of breast cancer cells to chloropyramine and is associated with prognosis in breast cancer. Oncotarget 2016;7:72807–18.
[23] Huang X, Tao Y, Gao J, et al. UBC9 coordinates inflammation affecting development of bladder cancer. Sci Rep 2020;10:20670.