Integrative analysis of 1q23.3 copy number gain in metastatic urothelial carcinoma - Supplemental Information - 1q23.3 Analysis

Markus Riester, Lillian Werner, Joaquim Bellmunt, Shamini Selvarajah, Elizabeth A. Guancial, Barbara A. Weir, Edward C. Stack, Rachel S. Park, Robert O'Brien, Fabio A. B. Schutz, Toni K. Choueiri, Sabina Signoretti, Josep Lloreta, Luigi Marchionni, Enrique Gallardo, Federico Rojo, Denise I. Garcia, Yvonne Chekaluk, David Kwiatkowski, Bernard Bochner, William C. Hahn, Azra H. Ligon, Justine A. Barletta, Massmio Loda, David M. Berman, Philip Kantoff, Franziska Michor, Jonathan Rosenberg

August 15, 2013

Contents

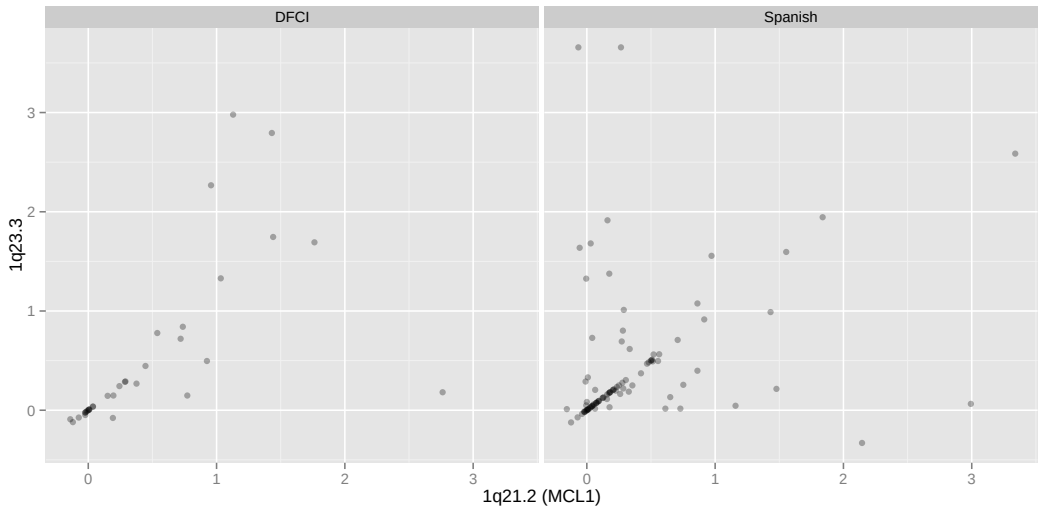


Figure S7: Comparison of copy number ratios at the 1q21.2 (MCL1) locus and 1q23.3. As shown in Supplemental Table S3, although many patients have the same copy numbers at both loci, there are more amplifications at 1q23.3.

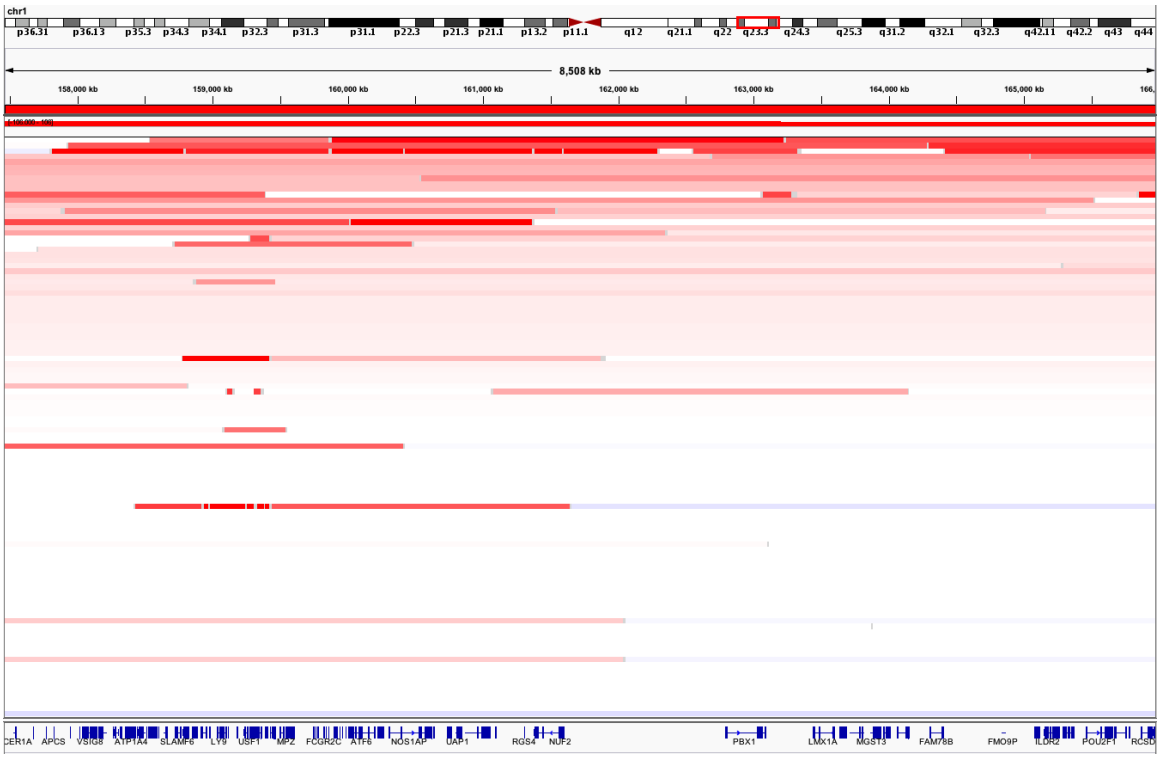
1 1q23.3 Region Analysis

We then tested for all focal GISTIC events (residual q-value < 0.05) whether high gains ($\log_2$ copy number ratio > 0.9) or high deletions ($\log_2 < -1.3$) were associated with overall survival in the Spanish cohort. There were 63 significant GISTIC peaks with at least one high-level amplification or deletion. Table S3 lists all peaks with significant survival association ($p < 0.05$), ranked by GISTIC q-value. Only peaks on chromosome arm 1q (hg18 121.02 − 167.67 Mb) were significant, with 1q23.3 showing the lowest GISTIC q-value. The 1q21.2 loci harbors MCL1. MCL1 and 1q23.3 copy numbers are compared in Supplemental Figure S7. In all analyses of this study, copy numbers of regions were estimated by the GISTIC software and correspond to the maximum copy numbers within wide peak regions.

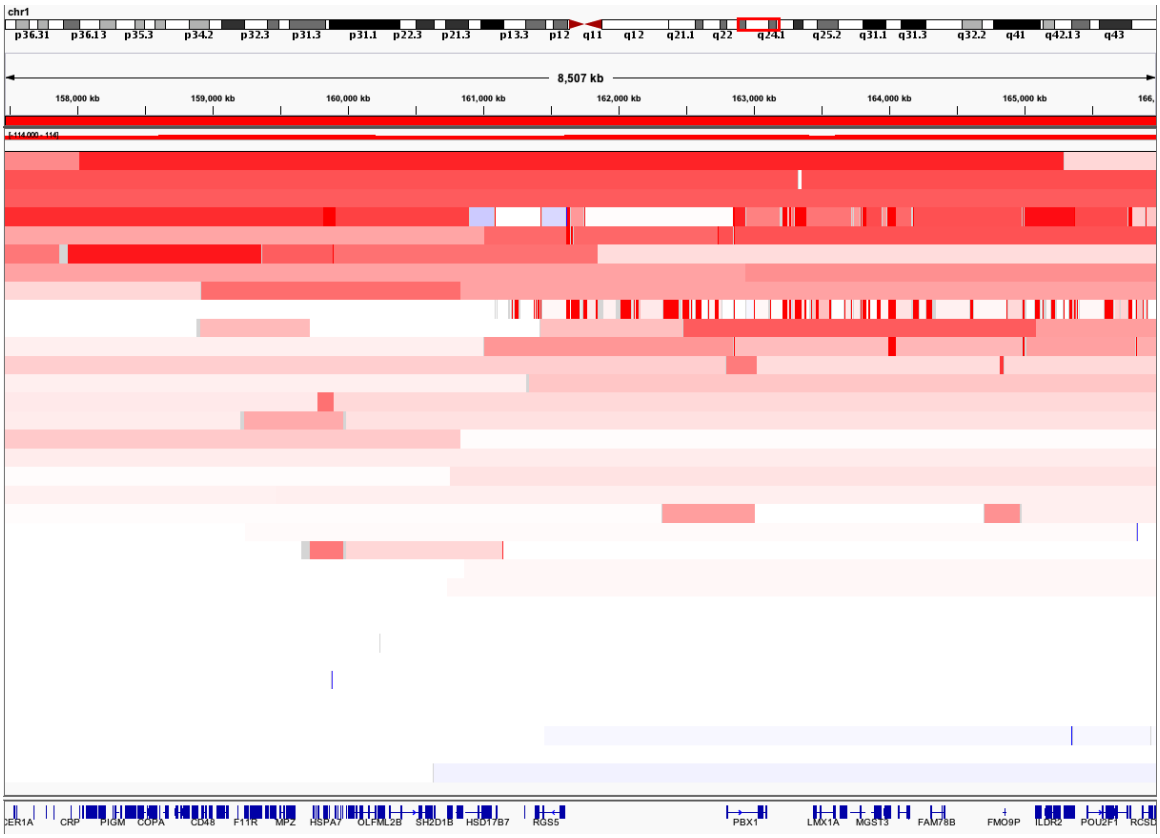
Table S3: Associations of GISTIC peaks with overall survival after start of chemotherapy, adjusted for ECOG performance status and visceral disease (Spanish cohort).

Chr	Start	End	Type	GISTIC q-value	adj. Cox p-value	adj. Cox FDR	High Gains
1q23.3	159.257	159.407	Amp	< 0.001	0.007	0.117	15
1q21.2	148.748	149.236	Amp	< 0.001	0.005	0.117	10
1q21.2	147.645	148.138	Amp	< 0.001	0.003	0.117	9

We used the GISTIC algorithm to identify focal recurrent alterations and their drivers. GISTIC determines *wide peak* regions that contain the target of the focal copy number gain with a specified confidence level. With a level of 0.95, this algorithm identified 3 peaks in 1q23.3 in 3 bladder cancer cohorts (Fig. 3 in the paper, Supplemental Figure S8). In our discovery cohort, only one peak had a significant residual q-value (< 0.05). The residual q-value is the q-value adjusted for overlapping neighboring peaks. Table S4 lists the frequencies of 1q23.3 gains in all three bladder cancer cohorts and all three peaks with default GISTIC cutoffs. Here we repeat the survival analysis for the two alternative peaks.



(a) Spanish



(b) DFCI

Figure S8: IGV screenshots of the 1q23.3 amplified region (Figure 3) for both cohorts.

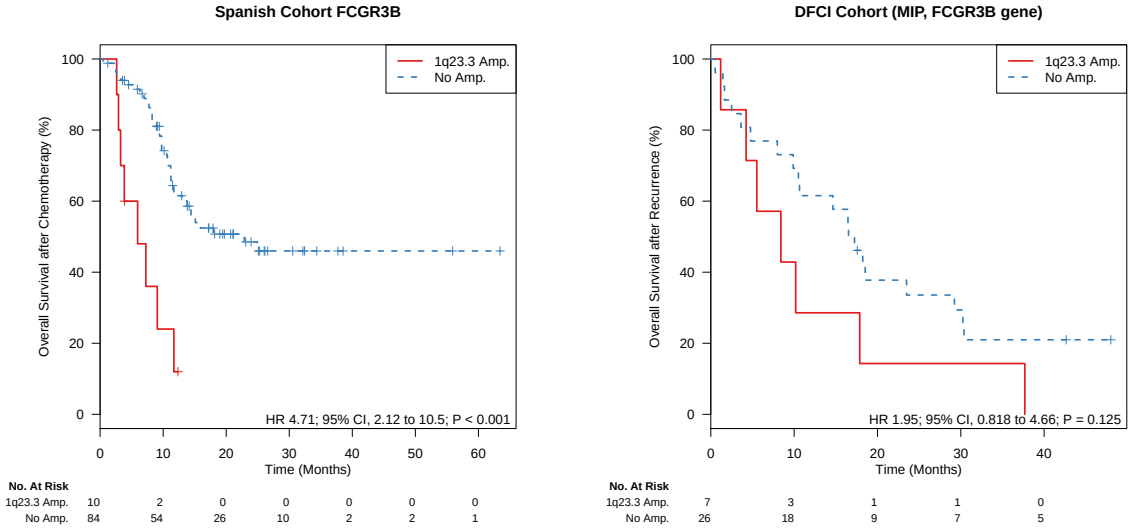
Table S4: Copy numbers in all bladder cancer cohorts at all three 1q23.3 GISTIC peaks. Numbers of patients in the TCGA cohort correspond to patients with clinical information.

Cohort	Peak	High Deletion	Deletion	Normal	Gain	High Gain
Spanish	1	0 (0%)	1 (1.1%)	40 (42.6%)	38 (40.4%)	15 (16%)
Spanish	2	0 (0%)	1 (1.1%)	43 (45.7%)	40 (42.6%)	10 (10.6%)
Spanish	3	0 (0%)	5 (5.3%)	43 (45.7%)	39 (41.5%)	7 (7.4%)
DFCI	1	0 (0%)	0 (0%)	22 (64.7%)	8 (23.5%)	4 (11.8%)
DFCI	2	0 (0%)	1 (2.9%)	18 (52.9%)	7 (20.6%)	8 (23.5%)
DFCI	3	0 (0%)	0 (0%)	15 (44.1%)	12 (35.3%)	7 (20.6%)
TCGA	1	0 (0%)	2 (2.2%)	42 (46.2%)	35 (38.5%)	12 (13.2%)
TCGA	2	0 (0%)	12 (13.2%)	28 (30.8%)	35 (38.5%)	16 (17.6%)
TCGA	3	0 (0%)	6 (6.6%)	42 (46.2%)	31 (34.1%)	12 (13.2%)

1.1 Peak 2 (FCGR3B gene)

Table S5: Associations of ECOG PS, visceral disease, 1q23.3 amplification (peak 2, FCGR3B gene) and OS. The concordance provided in the right-most column is the probability that in a random pair of patients, the patient with higher risk had shorter survival.

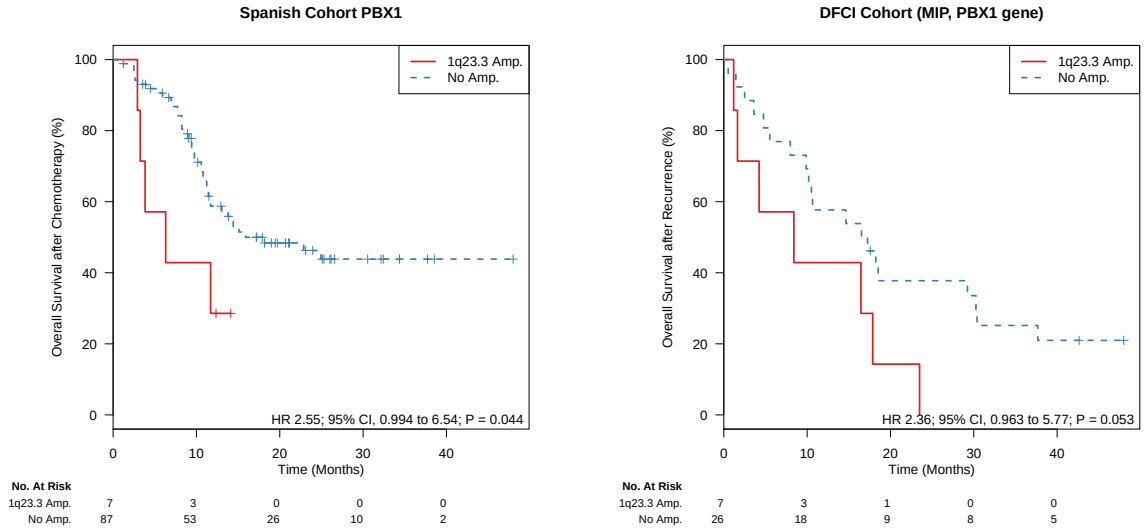
Cohort	Characteristic	N	Events	HR	P	Concordance
Spanish	1q23.3 Amp (log2 $>$ 0.9)	94	46	4.71 (2.12-10.46)	$<$ 0.001	0.58 (+-0.02)
Spanish		94	46		$<$ 0.001	0.68 (+-0.04)
	1q23.3 Amp (log2 $>$ 0.9)			5.14 (2.18-12.15)	$<$ 0.001	
	ECOG PS $>$ 0			1.86 (0.94-3.66)	0.073	
	visceral disease			2.27 (1.23-4.18)	0.008	
DFCI	1q23.3 Amp (log2 $>$ 0.9)	33	29	1.95 (0.82-4.66)	0.132	0.56 (+-0.04)
DFCI		32	28		0.224	0.64 (+-0.06)
	1q23.3 Amp (log2 $>$ 0.9)			1.8 (0.74-4.37)	0.197	
	ECOG PS $>$ 0			1.65 (0.75-3.62)	0.216	
	visceral disease			1.39 (0.65-2.99)	0.397	



(a) Spanish

(b) DFCI

Figure S9: Kaplan-Meier curves for the second 1q23.3 GISTIC peak, located around *FCGR3B*.



(a) Spanish

(b) DFCI

Figure S10: Kaplan-Meier curves for the *PBX1* GISTIC peak at 1q23.3 (peak 3).

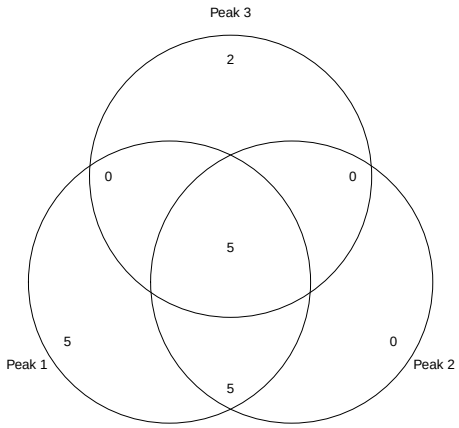
1.2 Peak 3 (PBX1 gene)

Table S6: Associations of ECOG PS, visceral disease, 1q23.3 amplification (peak 3, PBX1 gene) and OS.

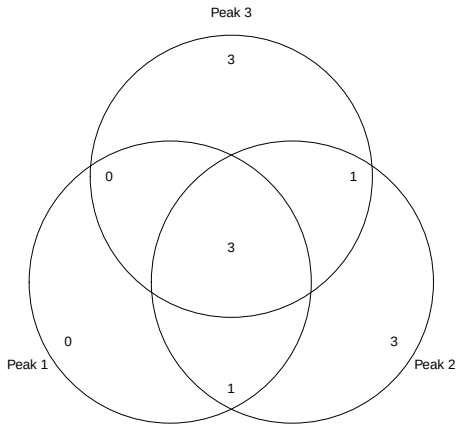
Cohort	Characteristic	N	Events	HR	P	Concordance
Spanish	1q23.3 Amp (log2 $>$ 0.9)	94	46	2.55 (0.99-6.54)	0.051	0.54 (+-0.02)
Spanish		94	46		0.008	0.67 (+-0.04)
	1q23.3 Amp (log2 $>$ 0.9)			1.92 (0.73-5.03)	0.183	
	ECOG PS $>$ 0			2.03 (1.04-3.99)	0.039	
	visceral disease			1.84 (1.03-3.3)	0.04	
DFCI	1q23.3 Amp (log2 $>$ 0.9)	33	29	2.36 (0.96-5.77)	0.06	0.57 (+-0.04)
DFCI		32	28		0.198	0.64 (+-0.06)
	1q23.3 Amp (log2 $>$ 0.9)			1.9 (0.74-4.85)	0.179	
	ECOG PS $>$ 0			1.53 (0.68-3.46)	0.308	
	visceral disease			1.27 (0.59-2.72)	0.546	

1.3 Nanostring Data

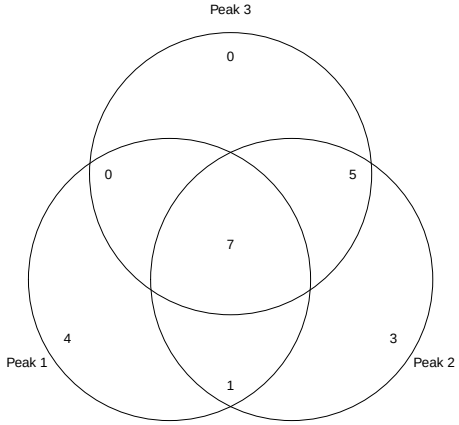
For 79 patients in the Spanish cohort and 24 patients in the DFCI cohort, we had matched Nanostring read count data available for all protein coding genes in the 1q23.3 amplified region (Figure 3 of the main paper). The following Supplementary Table S7 lists the correlation coefficients and their statistical significance (false discovery rates) in both cohorts. For each of the genes in 1q23.3, we further tested in both cohorts whether increased mRNA expression is associated with poor prognosis (columns FDR OS). In Supplemental Figure S12, the correlations of mRNA expression and copy number are shown for genes in the GISTIC peak 1.



(a) Spanish cohort



(b) DFCI cohort



(c) TCGA cohort

Figure S11: Venn diagram of patients with amplification ($\log_2 > 0.9$) of the three GISTIC peaks in the (a) Spanish, (b) DFCI and (c) TCGA cohorts.

Table S7: For the genes in 1q23.3, this table lists the correlations of copy number and mRNA expression and the correlations of overall survival and mRNA expression. These correlations are reported for both the Spanish and the DFCI cohorts. Column 1 and 4 is the Spearman correlation of Nanostring read counts and aCGH/MIP copy number. Columns 2 and 5 list the corresponding false discovery rates (FDRs). Columns 3 and 6 report the significance of the association of mRNA expression with overall survival (OS) in the Spanish and DFCI cohorts, respectively. The sample size for the DFCI cohort was too small for adjusting for multiple testing and the unadjusted p-value instead of the FDR is shown. Finally, the right-most column indicates whether the gene is located in the GISTIC peak 1.

Gene	Spanish Cohort ($n = 79$)			DFCI Cohort ($n = 24$)			Peak 1
	Correlation	FDR	FDR OS	Correlation	FDR	P-value OS	
DEDD	0.7557	< 0.001	0.089	0.7757	< 0.001	0.01	Yes
F11R	0.6687	< 0.001	0.002	0.8009	< 0.001	0.02	Yes
SDHC	0.6989	< 0.001	< 0.001	0.7313	< 0.001	0.03	No
NIT1	0.7324	< 0.001	0.014	0.6139	< 0.001	0.05	Yes
DUSP12	0.6283	< 0.001	0.045	0.7861	< 0.001	0.10	No
PVRL4	0.5981	< 0.001	0.229	0.8287	< 0.001	0.03	Yes
PFDN2	0.6017	< 0.001	0.003	0.7991	< 0.001	0.04	Yes
NCSTN	0.5854	< 0.001	0.016	0.8278	< 0.001	0.03	No
NDUFS2	0.6312	< 0.001	< 0.001	0.7278	< 0.001	0.06	No
USP21	0.6559	< 0.001	0.011	0.6704	< 0.001	0.10	Yes
IGSF9	0.6143	< 0.001	1	0.7026	< 0.001	0.01	No
PPOX	0.6184	< 0.001	0.015	0.6696	< 0.001	0.06	Yes
ATF6	0.5309	< 0.001	0.029	0.8035	< 0.001	0.10	No
USF1	0.5577	< 0.001	0.041	0.7357	< 0.001	0.05	Yes
UHMK1	0.5742	< 0.001	0.067	0.6957	< 0.001	0.08	No
B4GALT3	0.5982	< 0.001	0.007	0.6339	< 0.001	0.07	No
DUSP23	0.4851	< 0.001	1	0.8296	< 0.001	0.04	No
TOMM40L	0.6293	< 0.001	0.04	0.5661	< 0.001	0.41	No
COPA	0.466	< 0.001	< 0.001	0.7922	< 0.001	0.04	No
TAGLN2	0.4987	< 0.001	0.141	0.7061	< 0.001	0.01	No
PIGM	0.4762	< 0.001	0.649	0.7435	< 0.001	0.16	No
UAP1	0.4847	< 0.001	0.091	0.7174	< 0.001	0.14	No
HSD17B7	0.5825	< 0.001	0.982	0.5017	< 0.001	0.31	No
PEX19	0.4255	< 0.001	0.002	0.733	< 0.001	0.04	No
UFC1	0.4847	< 0.001	0.054	0.6235	< 0.001	0.14	Yes
IGSF8	0.4533	< 0.001	1	0.5565	< 0.001	0.10	No
PEA15	0.3917	< 0.001	0.901	0.6009	< 0.001	0.03	No
CCDC19	0.4305	< 0.001	1	0.4765	0.01784	0.39	No
ARHGAP30	0.2134	0.009603	1	0.5174	< 0.001	0.00	Yes
NOS1AP	0.2132	0.00939	1	0.4852	0.01846	0.02	No
CASQ1	0.2776	< 0.001	1	0.26	0.2204	0.05	No
ATP1A4	0.3647	< 0.001	1	0.08	0.4614	0.08	No
VSIG8	0.2505	0.003912	1	0.2852	0.1784	0.78	No
PBX1	0.2982	< 0.001	0.015	0.1922	0.3263	0.20	No
APCS	0.2876	< 0.001	1	0.2096	0.3099	0.30	No
VANGL2	0.2119	0.009186	0.045	0.1435	0.3746	0.10	No
SLAMF9	0.2197	0.01031	1	0.1043	0.4361	0.59	No
ITLN1	0.1589	0.02657	< 0.001	0.18	0.3284	0.13	No
HSPA6	0.0816	0.07042	0.984	0.2861	0.184	0.49	No
LMX1A	0.2996	< 0.001	1	-0.2457	0.2379	0.99	No
SH2D1B	0.2502	0.003807	1	-0.1565	0.3568	0.41	No
C1orf192	0.1306	0.0382	1	0.04696	0.5012	0.28	No
FCRL6	0.1017	0.05679	1	0.05913	0.4913	0.20	No

ITLN2	0.216	0.009826	1	-0.1652	0.3416	0.82	No
RGS4	0.06776	0.08078	1	0.08609	0.4601	0.41	No
CD244	0.2101	0.00899	1	-0.1922	0.3186	0.02	No
APOA2	0.09354	0.06067	0.109	-0.0313	0.5187	0.92	No
ATP1A2	0.1339	0.03885	1	-0.1087	0.4342	0.88	No
C1orf111	0.1365	0.03706	1	-0.1278	0.4014	0.62	No
RGS5	0.152	0.02817	1	-0.2017	0.3077	0.53	No
CRP	0.07067	0.07988	1	-0.05739	0.4834	0.24	No
SLAMF8	-0.04394	0.1006	1	0.1461	0.3714	0.75	No
KCNJ9	0.09381	0.06162	1	-0.1313	0.3988	0.91	No
C1orf110	0.09418	0.0626	1	-0.1704	0.3449	0.67	No
CD84	-0.1607	0.02709	1	0.2278	0.2748	0.38	No
DDR2	-0.1433	0.0327	1	0.03913	0.5101	0.72	No
SLAMF6	-0.2224	0.007042	1	0.06609	0.4817	0.54	No
MPZ	-0.1132	0.0508	1	-0.1835	0.3361	0.98	No
SLAMF1	-0.1751	0.01972	1	-0.07478	0.4717	0.75	No
FCGR2B	-0.05752	0.08777	1	-0.2948	0.1726	0.52	No
CD48	-0.2017	0.01174	1	-0.09826	0.4476	0.23	No
SLAMF7	-0.219	0.01006	1	-0.0935	0.4492	0.43	No
FCER1G	-0.1989	0.0115	1	-0.1661	0.3491	0.47	No
LY9	-0.1574	0.02608	1	-0.2504	0.2294	0.56	No

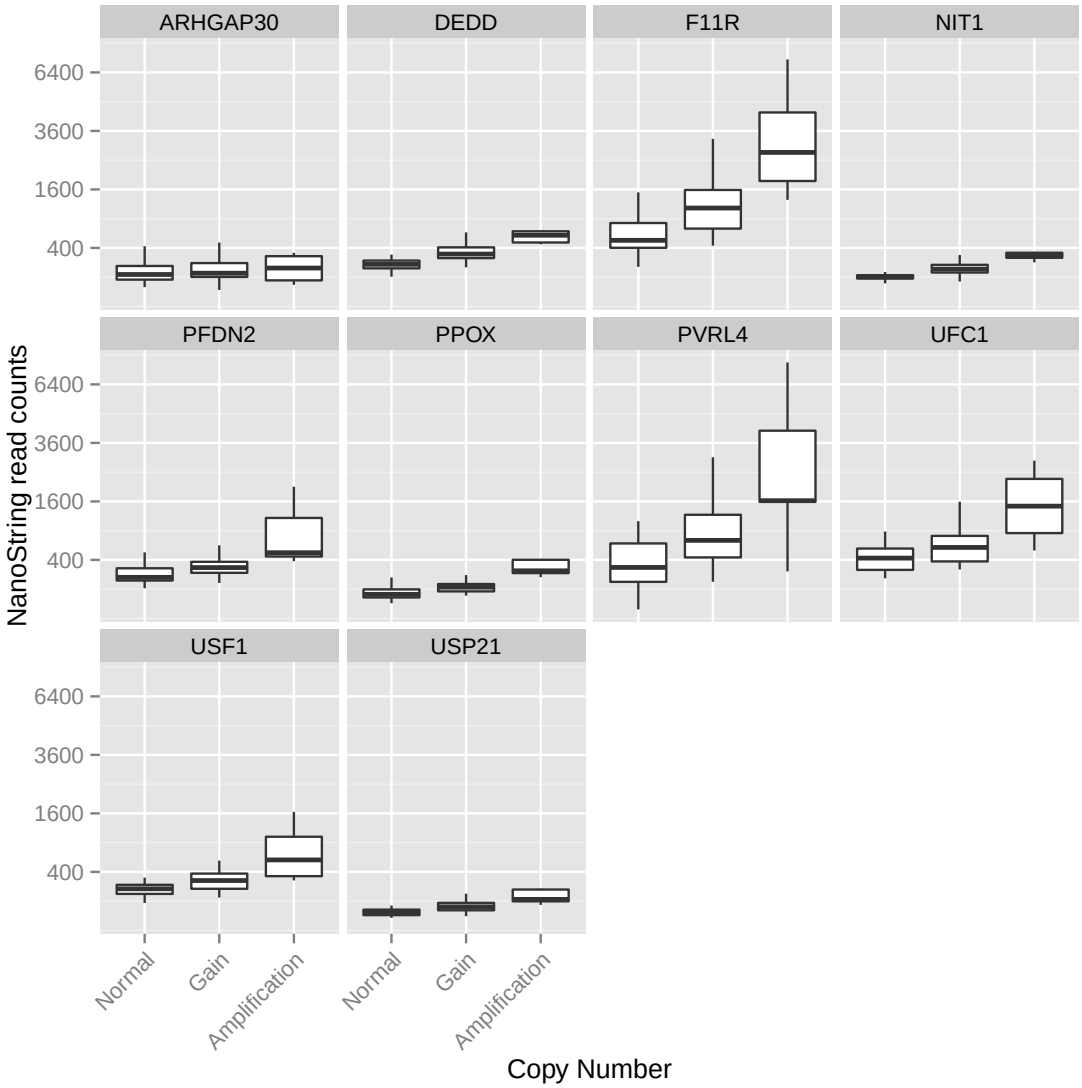


Figure S12: Boxplots visualizing the correlation of mRNA expression and copy number in the Spanish cohort for all genes in the 1q23.3 GISTIC peak 1. Shown on the y-axis are square-root transformed NanoString read counts for samples with normal copy number ($-0.15 > \log_2$ copy number ratio < 0.15), copy number gain ($0.15 > \log_2$ copy number ratio < 0.9), and copy number amplification (> 0.9).

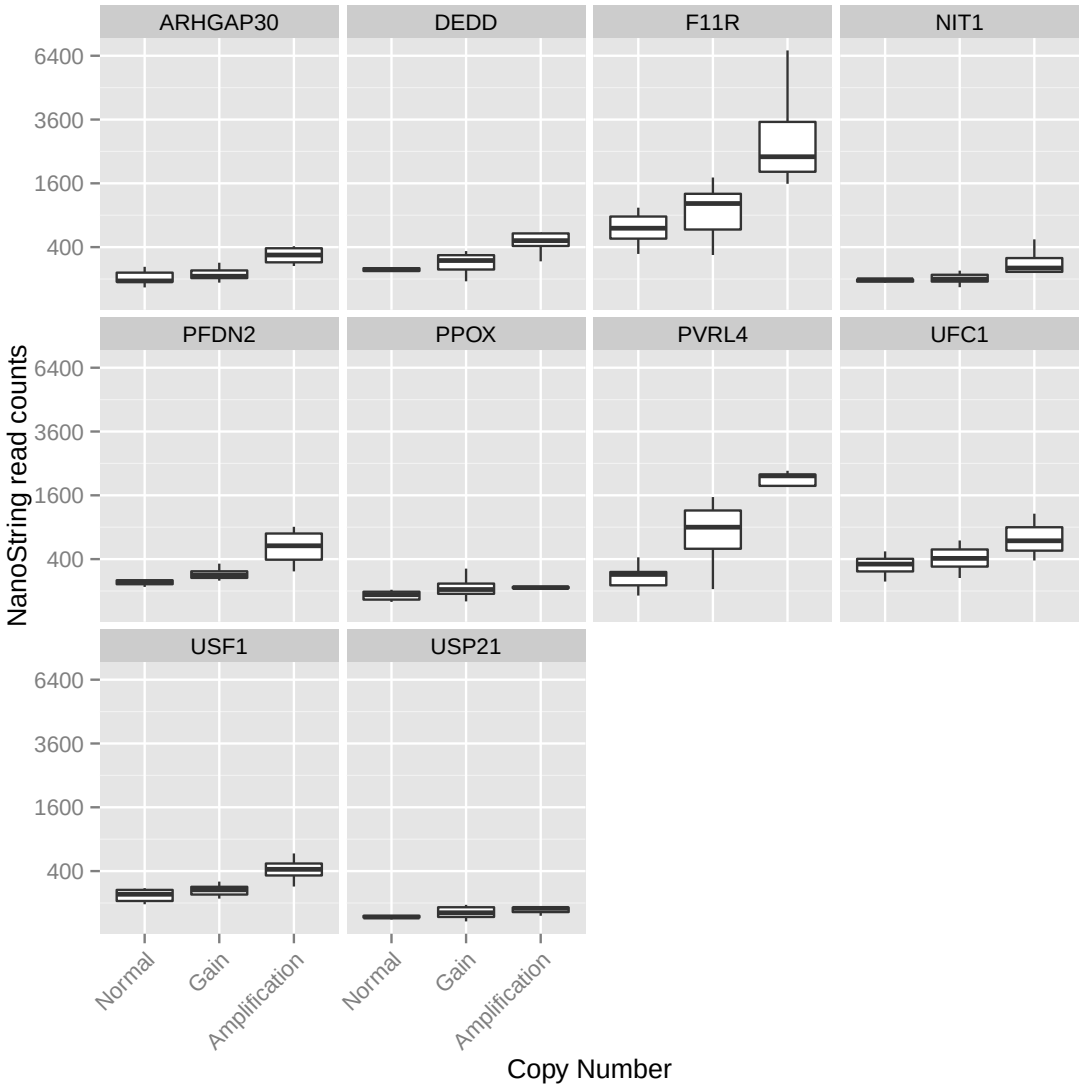
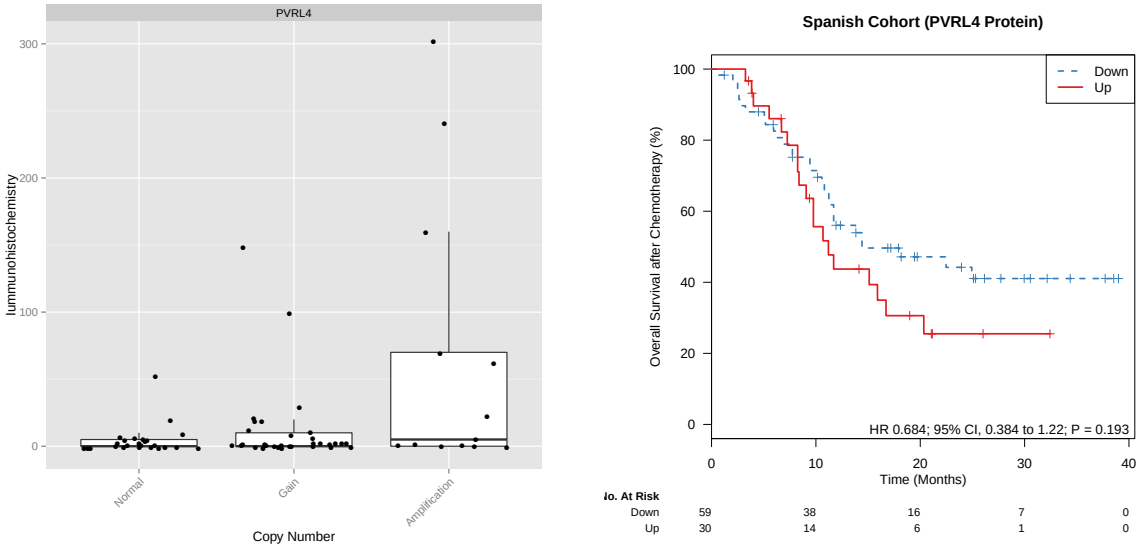


Figure S13: Boxplots visualizing the correlation of mRNA expression and copy number in the DFCI cohort for all genes in the 1q23.3 GISTIC peak 1. Shown on the y-axis are square-root transformed NanoString read counts for samples with normal copy number ($-0.25 > \log_2$ copy number ratio < 0.25), copy number gain ($0.25 > \log_2$ copy number ratio < 0.9), and copy number amplification (> 0.9).

1.4 Immunohistochemistry analysis of PVRL4

Similarily as for the NanoString data, Supplemental Figure S14 shows the correlation of copy number and IHC staining. IHC staining was only performed for the *PVRL4* gene in the Spanish cohort.

Methods: Tissue microarrays from the Spanish cohort tumors were stained with anti-PVRL4 (Thermo, PA5-30837). Antigen retrieval was performed in citrate buffer using a microwave set on high for five minutes, repeated three times. Following antigen retrieval, slides were transferred to an i6000 automated staining deck (Biogenix), rinsed in a PBS-t wash for 15 minutes, incubated in a commercial peroxidase blocking solution (Dako) for 30 minutes to suppress endogenous peroxidase, and then immediately incubated with protein block (Dako) for 20 minutes to suppress non-specific background staining. The slides were then incubated with the primary antisera to PVRL4 for one hour. The primary antisera was visualized using a peroxidase-based detection kit (Dako Envision), following the manufacturers protocol. The slides were counterstained with hematoxylin (Biogenix). Individual TMA cores were evaluated by a single pathologist (J.A.B.).



(a) (b)

Figure S14: Panel (a) shows the correlation of copy number and Immunohistochemistry (IHC) staining of the *PVRL4* gene in the Spanish cohort. (b) Kaplan-Meier plot, in which patients where stratified into two groups based on the median IHC score.

References

[1] C Huttenhower, M Schroeder, M D Chikina, and O G Troyanskaya. The sleipnir library for computational functional genomics. *Bioinformatics*, 24(13):1559–1561, Jul 2008.

[2] M N McCall, B M Bolstad, and R A Irizarry. Frozen robust multiarray analysis (frma). *Biostatistics*, 11(2):242–253, Apr 2010.

[3] M Riester, J M Taylor, A Feifer, T Koppie, J E Rosenberg, R J Downey, B H Bochner, and F Michor. Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. *Clin Cancer Res*, 18(5):1323–1333, Mar 2012.