Supplementary Materials

Nucleotide-resolution analysis of TMPRSS2 and ERG rearrangements in prostate cancer

Christopher Weier, Michael C. Haffner, Timothy Mosbruger, David M. Esopi, Jessica Hicks,

Qizhi Zheng, Helen Fedor, William B. Isaacs, Angelo M. De Marzo, William G. Nelson,

Srinivasan Yegnasubramanian

	Page
Supplementary Methods	2
Supplementary Figure 1	3
Supplementary Figure 2	4
Supplementary Figure 3	5
Supplementary Figure 4	6
Supplementary Figure 5	7
Supplementary Figure 6	8
Supplementary Figure 7	9
Supplementary Figure 8	10
Supplementary Figure 9	11
Supplementary Figure 10	12
Supplementary Table 1	13
Supplementary Table 2	14
Supplementary Table 3	15
Supplementary Table 4	16

Table of Contents

Additional Supplementary Material not included in this PDF:

Supplementary File containing chromosomal coordinates of RNA Baits used in geBACS.

SUPPLEMENTARY METHODS

PCR reactions for validation and case-specific breakpoint assignment

PCR reactions consisted of 40 µl reactions containing 50 ng of DNA, 1× Platinum Taq buffer (Invitrogen, Carlsbad, CA), 1.5 U Platinum Taq polymerase (Invitrogen), 250 µM each dNTPs, 1.5 mM MgCl₂, 0.25 µg/µl BSA, 2 µl dimethyl sulfoxide, 400 nM each forward and reverse primers. A touchdown PCR approach was utilized as follows: 95°C for 3 min, 4 cycles of denaturing at 94°C for 30 s, annealing at 70°C for 30 s and extension at 72°C for 45 s, repeating with successive annealing temperatures of 68°C, 65°C, and 60°C for 4 cycles each and finally 36 cycles at an annealing temperature of 55°C, followed by a 7 min extension step at 72°C.

Α

Case 45



Case 66



Case 77



В

Characteristics of rearrangement junctions identified in reference samples.*

Reference Case #	Junctions	Gene 1	Gene 2	Orientation	^A Microhomogy	^B Reads	^c Junction Reads
Case 45	TMPRSS2-ERG	chr21:41800806	chr21:38776561	s/s	-	18	5
Case 66	TMPRSS2-ERG	chr21:41795687	chr21:38804491	s/s	А	24	11
Case 77	TMPRSS2-ERG	chr21:41799367	chr21:38790118	s/s	-	21	12

*, Conventions are same as for Supplementary Table 2.

A. Bases at the junction that are shared by both genes in the rearrangement.

B. Number of paired-end reads flanking or directly overlapping the rearrangement junction.

C. Number of reads directly overlapping the rearrangement junction.

Supplementary Figure 1. Detection of rearrangement junctions in reference samples.

(A) Paired-end read distribution surrounding 3 knownTMPRSS2-ERG rearrangements from 3 reference samples allowing assessment of effectiveness of the geBACS strategy. The position of the breakpoint is indicated by the red line and by the Chr21 positions of the TMPRSS2 and ERG breakpoints at the top of each panel. (B) Characteristics of TMPRSS2-ERG breakpoints of reference samples.



geBACS sensitivity = 100% geBACS specificity = 87.5%

	TMPRSS2-	-ERG FISH
	positive	negative
istochemistry	92%	0%
positive	11/12	0/8
ERG immunoh	0%	100%
_{negative}	1/12	_{8/8}

ERG IHC sensitivity = 92% ERG IHC specificity = 100%

Patient Bank #	Junctions	ERG Stain	FISH
4	TMPRSS2-ERG	positive	D
372	TMPRSS2-ERG	negative	N
	SLC45A3-ERG		
535	TMPRSS2-ERG	positive	D
652	TMPRSS2-ERG	positive	Т
	ERG-ERG		
675	TMPRSS2-ERG	positive	D
682	TMPRSS2-ERG	positive	D
808	TMPRSS2-ERG	positive	Т
	TMPRSS2-ERG		
814	TMPRSS2-ERG	positive	Т
816	TMPRSS2-ERG	positive	Т
	TMPRSS2-TMPRSS2		
981	TMPRSS2-ERG	positive	Т
989	FAM177A-ERG	positive	D
995	TMPRSS2-ERG	positive	Т
	ERG-ERG		
	TMPRSS2-TMPRSS2		
	TMPRSS2-TMPRSS2		
	TMPRSS2-TMPRSS2		
1273	ELK4-ERG	negative	T, ERG*
170	-	negative	Ν
178	-	negative	N
700	-	negative	Ν
913	-	negative	N
994	-	negative	N
1048	-	negative	N
1508	-	negative	N

В

Rearrangement status by FISH; rearrangement with deletion (D), rearrangement with translocation (T), no rearrangement (N). *, Case 1273 showed translocation only of the 3' ERG probe.

Supplementary Figure 2. Sensitivity and specificity determination of geBACS by FISH.

(A) Specificity and sensitivity of the geBACS pipeline in identifying TMPRSS2-ERG rearrangements, determined by using TMPRSS2-ERG FISH. (B) Correlation between ERG IHC and TMPRSS2-ERG FISH. (C) Table summarizing IHC and FISH data.



Supplementary Figure 3. In situ RNA detection of ETV4 and ETV5 in cases harboring TMPRSS2-ETV4 and TMPRSS2-ETV5 rearrangements. Case 1538 with geBACS-confirmed ETV4 rearrangement shows positive reactivity for ETV4 expression in neoplastic cells (arrows) (A), but no reactivity with ETV5 specific probes (C). Case 1164 with geBACS confirmed TMPRSS2-ETV5 rearrangement shows positive reactivity for ETV5 expression in tumor cells (arrows) (B), but no reactivity with ETV4 probes (D). Arrows indicate representive prostate cancer cells, and Arrowheads indicate representative normal prostate epithelial cells.





Supplementary Figure 4. The most common TMPRSS2-ERG rearrangement involves the first intron of TMPRSS2 and the third intron of ERG resulting in an exon 1 – exon 4 TMPRSS2-ERG fusion transcript. (A) RT-PCR showing the expected exon 1 – exon 4 fusion previously noted in the VCaP cell line. Lane 2 shows no amplification of the transcript from the rearrangement negative LAPC4 cell line while lanes 3 and 4 show a similar exon 1 – exon 4 transcript in Cases 808 and 652 respectively, in which geBACS identified rearrangements occurring in intron 1 of TMPRSS2 and intron 3 or ERG. Bands were excised and sequenced to verify the expected rearrangement transcript. (B) Detailed schematic showing the configuration of the most common TMPRSS2-ERG rearrangement sites identified by geBACS and the structure of the expected fusion transcript sequence.



Supplementary Figure 5. Inter- and Intra- genic rearrangements in TMPRSS2 and ERG are only detected in tumor, and not in normal adjacent tissue. All inter- and intra- genic rearrangements identified by geBACS for case 816 (A) and case 995 (B) were detected by PCR using primers flanking the rearrangement breakpoint only in tumor, but not adjacent normal tissues. (A,B) Primers targeting a TMPRSS2 sequence not involved in rearrangement (Wild type TMPRSS2) were used as an input control for the PCR reactions.



Supplementary Figure 6. Detailed characterization of rearrangements in Case 995. (A) Mono-allelic rearrangement associated with translocation by FISH. White arrowheads indicate normal alleles, white arrows indicate rearranged alleles. (B) Circos plot depicting complex interand intra- genic rearrangements involving TMPRSS2 and ERG. (C) A schematic representing the location and orientation of a complex rearrangement in case 995 involving inversion of an ERG segment, inversion of a TMPRSS2 segment, and the rearrangement of TMPRSS2 with ERG is shown. A PCR amplicon spanning multiple contiguous complex rearrangement junctions was sequenced and used to generate this schematic (see Supplementary Table 3 for the PCR primers).



Supplementary Figure 7. Detailed characterization of rearrangements in VCaP cell line. (A). TMPRSS2-ERG FISH of VCaP cells confirmed the presence of a rearrangement, and revealed a bi-allelic rearrangement for TMPRSS2-ERG associated with translocation of

revealed a bi-allelic rearrangement for TMPRSS2-ERG associated with translocation of intervening sequences. (B) Circos plot depicting inter- and intra-genic rearrangements in VCaP. (C) Genomic sequence of the TMPRSS2-ERG fusion in VCaP, as identified by geBACS and confirmed by PCR and Sanger sequencing.



Supplementary Figure 8. Genomic rearrangements involving TMPRSS2, SLC45A3, ERG, ETV4 and ETV5. (A) Circos plots showing architecture of genomic rearrangements in cases harboring TMPRSS2-ETV4 (case 1538), TMPRSS2-ETV5 (case 1165) and SLC45A3-ERG (case 372) fusions. (B) Nucleotide resolution of breakpoint sequence of the above cases. Sequences showing flanking mircohomologies are highlighted in boxes. Note that case 1165 shows extensive microhomology overlapping and flanking the fusion breakpoint.





- 1 CCTCCTTCTTCCAGAGACTAAGCAATCCAGTTTACACAACAGGCTCTACAGT 2 GCCAGTGCATTAGGGCAGCGCTGACTGGTTTCCTGTTAAGCCCTCGC
- 3 GAAGCCGGCCCCACTCTACTTCAGGATACCTAACACAGGTGAGCCCCACC
- 4 CCCCTCTCCTAAGGGCCATCCTTGGTG AGGACGCTGAGGCAaGAGAATTA



case 558

- 1 GTAAAATACATATAGTGTGCTAGATATGAGTATTGATTCTTCACTCCTTT
- 2 TACAATAATAGGTCCTTCCTGACACCT/AATAAGTTTTAAAGGAAGAGGAA
- 3 CTAGAAACTGACACATGCTGAACATaaaagGAGCATAAGGTCTTCAGCACT





- 1 AGAAGGGGAAGATGTGGGGCTGGTGGGGCCCTGCAAACATCACAAAGAGCACT
- 2 GAGGAAGGTCCCCAGGGTCAAGGTGAGATGTTTAATACCTACAAATACAG
- 3 TATCTACGTTATTAAGACAACTGCCTGACGCTCAGTGAAATAATTCAGGT



- 1 ACACAGCTGCCCAGGTGAGTCGCAAGAGCAGACTGAGATAGGCTTCCCGA
- 2 GTGGCTCGTAGCCACTGTACTGACTCAGGAATTTTCAGGGACAAACCTGC
- 3 AAATGTTTAAGTCAGTACAGTGGCcTGAGGACGGGCATCCTGGACCCATG
- 4 TGGACCCATGGTGGCCACATCTAAGC/AAATGAGTAAGATGAAATTTAGCA

Supplementary Figure 9. Cases showing complex genomic rearrangements involving TMPRSS2 and ERG. Representative circos plots of various cases exhibiting complex intra- and inter-chromosomal rearrangements involving TMPRSS2 or ERG. The sequences of the breakpoint for the junctions from each case are shown below.

case 816



Fraction of rearrangement breakpoints showing microhomology

Supplementary Figure 10. Strong enrichment of microhomologies at breakpoint junctions.

Groups of 22 randomly generated rearrangements between the introns of TMPRSS2 and ERG were simulated 10,000 times. The red distribution shows the fraction of rearrangements in each simulated dataset displaying microhomologies (1+ nucleotides) at the junction. The dashed blue line indicates the fraction of rearrangement breakpoints with microhomologies in our observed dataset, and shows that our observed dataset is significantly enriched for microhomology compared to what would be expected by random chance (p = 0.0087).

Median Age (range)	61 (34-72)
Ethnicity	N (%)
Caucasian	74 (89.2)
African American	6 (7.2)
Hispanic	1 (1.2)
Other	1 (1.2)
N/A	1 (1.2)
Gleason Sum*	N (%)
5	1 (1.3)
6	11 (14.1)
7	43 (55.1)
8	11 (14.1)
9	12 (15.4)
Surgical Margins* Positive Negative	N (%) 28 (36) 50 (64)
Median Gland Weight (range)*	55.05 (27-155)
TNM*'**	N (%)
T1	0 (0)
Т2	6 (7.7)
Т3	71 (91)
Τ4	0 (0)
ТХ	1 (1.3)
NO	62 (79.5)
N1	16 (20.5)

Supplementary Table 1: Clinical and pathological characteristics of study samples

* Data based on 78 specimens; surgical data unavailable for 5 specimens

**Distant metastases were not assessed at time of prostatectomy

Supplementary Table 2: D	Detailed clini	ical and patholo	gical characteristics of	cohort			
Patient Bank #	Age	Ethnicity	MarginsPositive	GleasonSum	Primary	Nodes	Tumor Cellularity %
4	59	W	Negative	7	T3	NO	80
10	62	W	Negative	7	13	N1	NA
13	64	W	Negative	5	13	NU	NA
70	6/	vv	Positive	6	13	NU	NA NA
82	/1	vv	Negative	/	13	NU	IVA
96	48	W	Negative	9	13	N1	NA
116	64	W	Negative	6	13	NU	NA
134	60	W	Negative	6	13	NO	NA
144	59	W	Negative	7	T3	NO	NA
146	63	W	Negative	7	T3	NO	NA
178	51	W	Negative	7	T3	NO	NA
179	63	W	Negative	9	13	N1	NA
199	70	W	Negative	7	T3	N1	NA
202	58	W	Negative	6	T3	N1	NA
206	57	W	Negative	9	T3	NO	NA
214	69	W	Positive	7	Т3	NO	65
219	66	W	Positive	8	T3	NO	NA
243	52	W	Negative	7	Т3	N1	NA
270	61	W	Negative	7	Т3	NO	80
327	62	W	Positive	7	Т3	NO	75
341	66	W	Negative	7	T2	NO	80
353	65	W	Negative	6	Т3	N0	85
357	57	W	Positive	7	T3	N0	70
358	57	W	Negative	7	T3	N1	65
362	55	W	Negative	9	T3	N0	70
372	68	W	Positive	7	T3	N0	80
388	72	W	Negative	7	Т3	N0	70
460	68	W	Negative	9	Т3	N0	75
535	54	W	Negative	8	T3	N0	70
541	67	W	Negative	8	Т3	N0	65
558	62	W	Negative	6	T3	N0	NA
580	52	W	Negative	7	T3	N0	NA
594	54	W	Positive	6	T3	NO	NA
652	66	W	Positive	7	Т3	N1	85
667	65	W	Positive	7	T3	N1	75
675	57	W	Negative	6	T3	N0	70
682	56	W	Positive	7	T3	N0	65
700	66	W	Negative	7	T3	N1	75
731	52	W	Negative	9	T3	N0	NA
733	67	W	Positive	9	Т3	N0	65
735	62	В	Positive	9	T3	NO	90
738	55	W	Negative	7	Т3	N0	70
741	66	W	Negative	9	T2	NO	NA
759	68	W	Positive	7	ТΧ	NO	NA
774	71	W	Negative	7	T3	NO	NA
780	53	W	Negative	8	T3	N1	NA
782	58	W	Negative	7	T3	NO	NA
808	57	W	Negative	7	T3	NO	70
814	67	W	Positive	7	T3	NO	65
816	62	W	Negative	6	T2	NO	70
823	70	W	Positive	7	T3	NO	70
913	57	W	Positive	7	T2	NO	NA
926	66	В	Negative	7	T3	NO	70
938	53	w	Positive	7	T3	NO	80
981	48	н	Positive	7	T3	NO	75
982	61	W	Positive	7	Т3	N1	70
986	66	В	Negative	7	T3	NO	65
989	58	W	Positive	7	T3	NO	70
994	53	В	Positive	7	T3	NO	NA
995	54	W	Negative	7	T3	NO	65
1009	51	W	Positive	7	T3	NO	80
1048	43	W	Negative	6	T3	NO	70
1070	66	W	Positive	8	T3	NO	75
1138	49	W	Negative	8	T3	N1	70
1164	57	W/	Negative	8	та	NO	65
1203	50	\M/	Negative	6	T2	NO	N/A
1203	62	\A/	Positivo	Q	T2	NO	70
1273	50	14/	Negativo	0	T2	NO	65
1422	59	VV	Desitive	0	15	NO	70
1422	64	vv	Positive	9	13	NU	/0
1437	48	w	Negative	8	13	NU	80
1508	61	0	Negative	/	13	NU	70
1518	51	W	Positive	9	13	N1	/5
1538	45	W	Negative	7	13	NO	65
1665	61	В	Negative	7	T3	N1	70
1704	65	W	Positive	9	T3	N1	65
1756	51	W	Negative	8	13	N0	90
1863	42	В	Positive	7	T3	NO	80
4441	55	W	Negative	7	T2	NO	NA

* Data for 78 specimens

Supplementary Table 3: Junction Specific Primer Sets

^A Patient Bank #	Blunctions	^C Gene 1	DGene 2	^E Forward	F Reverse
	TMPRSS2_FRG	chr21://179/326	chr21.38753597	GCAGGCTGTTGGGGTTTTAT	
10	TMPRSS2-FRG	chr21:41792775	chr21:38792757	GGTCAAAATGGTTGGCTCTG	
70	TMPRSS2-MPRI 12	chr21:41783383	chr17:77283404		
	HAP1*-TMPRSS2	chr17:37144837	chr21:41779769	GAGAATGGTTCTGCCCAAAG	AGACTGTGCTGGAGGAAAGG
	TMPRSS2-HAP1*	chr21:41794230	chr17:37149531	CTGACCCCCACTCTGAAGTT	CCTGTTGGGTGAGGTTTCTC
	TMPRSS2-SWI5	chr21:41779767	chr9:130085490	CTCCCTCAAGTGCAGTCCTAA	AGTGGCATGATTCTGGCTCA
82	EIE3K-TMPRSS2	chr19:43815389	chr21:41788584	GGAATCACCACTGGACACCT	AGGTCCAATAGCTGGTGGTG
	NF1-TMPRSS2	chr17:26590306	chr21:41774186	TCTGACCAGATGACCACAGC	TCTTCAGTGGTGCCCAGACT
134	TMPRSS2-ERG	chr21:41795069	chr21:38803391	TACTGGCAGCATCACTCTGG	CAGCCAAACAAACAACTTGC
	ACPP-TMPRSS2	chr3:133565489	chr21:41795096	CCAGGGGCCAGTTTTTAGAG	CCAGAGGCAGTCACAGGATA
144	TMPRSS2-ERG	chr21:41779065	chr21:38796877	AAATGGGTTTCATCAATGTTGTC	CCGGCCAATAACCTTTTGTA
	TMPRSS2-TMPRSS2	chr21:41789960	chr21:41781176	TCATACCTGTGCCAGTCTCG	AAGGAAAGGAAATCCGTGTG
199	TMPRSS2-ERG	chr21:41786586	chr21:38748712	CATGTAGGTTCATATTCTTCCCTTA	TGGATTTTCAAACCAAGTATCTACC
206	TMPRSS2-ERG	chr21:41793812	chr21:38763138	CATGGAGAGTGAGTGGCTGA	AGTCCAGCCTCATCCTCAGA
214	TMPRSS2-ERG	chr21:41797467	chr21:38749676	TGGAGAAATGATCCTCCAGTC	CTGGCAAGCAAAGATAGGG
341	TMPRSS2-ERG	chr21:41785537	chr21:38798302	AGATGCCTGTCCAGCAAAGT	TTTCATAAGCATCCCACACG
353	TMPRSS2-ERG	chr21:41789358	chr21:38799621	AGAGCTTTGCTGCCCTTGAT	AATCCTGCCTCAAGGATTTT
357	TMPRSS2-ERG	chr21:41794131	chr21:38797829	CATGTTGGTAGCCTGGGAAT	CCAGCACTCCATGGAACTTT
372	TMPRSS2-ERG	chr21:41797622	chr21:38784847	TAAGGCCATTGCTTTCCAAG	GGATGCCACGAAGAAATACAG
	SLC45A3-ERG	chr1:203908669	chr21:38798718	TTGGTAGGAGGCCATGAAAG	TCTGAGTCGGGGGTAGAGTG
535	TMPRSS2-ERG	chr21:41790462	chr21:38748584	ATGGGGGCCCTTAGATACAG	GGACAGTTAAATGGGCCAAA
558	TDRD7*-TMPRSS2	chr9:99213679	chr21:41794880	CCTCAGGCTTGACAACACAA	TCGTGTGCTAGGCACTATCCT
	TMPRSS2-TMPRSS2	chr21:41771407	chr21:41780399	CAACCTCACCAACTCCTGTT	GAGCGACGGACGTTTCTTTA
	MX1-TMPRSS2	chr21:41752881	chr21:41775109	GCTGGGTGAGCAGAATAGGAT	GGGATTGAAGGATGCTGTCT
580	MX1-TMPRSS2	chr21:41752819	chr21:41775553	GCTGGGTGAGCAGAATAGGAT	GGGATTGAAGGATGCTGTCT
652	TMPRSS2-ERG	chr21:41792424	chr21:38775810	ACTCACACATCGACACTTCCAG	ATGCAAATCCTGGATAAATGCT
	ERG-ERG	chr21:38830023	chr21:38872708	ATGAGCAATTAGCCCCCTCT	ATTCAGTTGCTCCTGGGAAG
675	TMPRSS2-ERG	chr21:41798800	chr21:38789416	GAAAGTGGCCTTGCTAGTGG	TGGCTCTCTCCACATCCTTC
682	TMPRSS2-ERG	chr21:41796887	chr21:38791968	AGCTTGAGGCACCTGGACT	CAGGCACCGTTAGGGATAGT
733	TMPRSS2-ERG	chr21:41794661	chr21:38816717	CTTCTGAGCTGGGAGAGCTG	TCAAGCCAGTTCCAAAGACA
738	THSD7A-TMPRSS2	chr7:11743984	chr21:41788390	GTTCCTTGATGAGTCATGGCTTATATT	AGACGACGGGGTTGGAAG
780	TMPRSS2-ERG	chr21:41792243	chr21:38787029	GGTTAAGTCCAGCAGGATGC	CCCATTGGCTCAATAATTCC
808	TMPRSS2-ERG	chr21:41790468	chr21:38780232	ATGGGGGCCCTTAGATACAG	TATTCAGCCTGGTCACATGC
	TMPRSS2-ERG	chr21:41783369	chr21:38805184	TCCCTACATATGTGCAGTCTAATTTC	GCTCAGAGCACTGGGGACT
	SPATA5L1*-TMPRSS2	chr15:43481074	chr21:41771055	GATGGATTTGTAGCACACCAA	GGGAGTCAAACATCCCAGGT
814	TMPRSS2-ERG	chr21:41792783	chr21:38784111	TGCTTTTCATAGATTCCCTGATT	AGGAACTGAGGAAGAGCAGGTG
	MORC3-TMPRSS2	chr21:36673099	chr21:41762272	ACTGAGGAGGCTCTGCTGAC	CACCACCCAAAAACTACCCA
816	TMPRSS2-ERG	chr21:41791718	chr21:38797360	TTTGTTCCCTTGTCCCTTCTG	AGAAAACCCAGAAACGCTCA
	C1qTNF1-TMPRSS2	chr17:74540999	chr21:41791286	GCACACAGAGGGAAAACACA	TATGGCAGCTGCTAGTCACG
	C1qTNF1-TMPRSS2	chr17:74540985	chr21:41764084	TCTTTCTGCCTGTCCTGGTC	TGCAGATTGCAAAGGAAGTG
	TMPRSS2-TMPRSS2	chr21:41764044	chr21:41796946	GTTTAGAGCTGCCCTGGAGA	AGCTTGAGGCACCTGGACT
981	TMPRSS2-ERG	chr21:41795091	chr21:38796469	TACAGGCACCCACCATCATA	TGGCAGCAGAACAGTATTGG
989	FAM177A1-ERG	chr14:34584342	chr21:38756847	TCCTTTCTCTTCTGTGCCAAA	CCAGGGCTCTCTAACACTGC
995	TMPRSS2-ERG '	chr21:41780434	chr21:38793491	-	CCTITAAAACTTATTTCCTGGCC
	TMPRSS2-TMPRSS2	chr21:41762098	chr21:41773713	TCCCATGGAAAACAAAGTGG	CGGTTTTTCAGAGGGTGTGT
	TMPRSS2-TMPRSS2	chr21:41/80464	chr21:41/90293		GATAGGCACCAGCIGAAGGA
	TMPRSS2-TMPRSS2	chr21:41/86495	chr21:41//3659	GATGCCAACAGTGTTTTCTTAAA	CGTAGCTCTGAGGGAAGGTG
1151		chr21:38/934/2	chr21:38/93604	AGAGCIGGACGIICAGIAAAIG	
1164	TIMPRSS2-ETV5	chr21:41/9/113	chr3:18/2/8460		
1273	ELK4-EKG	chr1:203858916	chr21:38/46/45		
1355	TIMPRSSZ-EKG	chr21:41788584	chr21:38/99/38		
1422	TIMPRSSZ-EKG	chr21:41/95553	chr17:28078855		
1550	TMDDCC2 EDC	chr21.41/95/42	chr21-29720722	GCCCGTTTGCCTTATACCTA	
1963	TMDRCC2_TMADDCC2	chr21.41/91298	chr21.30/29/32		
	TMDDCC2 EDC	chr21.41701944	chr21.41/7/074		
VCAF	TMPRSS2-ENG	chr21.41//3033	chr21.30/30223	GAGTGGCTGAGCCTGAGTTT	CACGTGGAATGGCTTTTTCT
Case 45	TMPR\$\$2-FRG	chr21.41793023	chr21.41/75567		
Case 66	TMPRSS2-FRG	chr21:41795687	chr21:38804491	CCAGGAATGGAGCCTGAGT	CCTTGTGCTTCTTTGTCAGT
Case 77	TMPRSS2-ERG	chr21:41799367	chr21:38790118	TGCTCCCTCTAGCAGTATGACA	CCAAGGGAAAATGATGGTAAGA

¹Complex rearrangement containing a TMPRSS2-ERG junction and an ERG-ERG junction were discovered with a single set of primers.

A. Patient Bank # indicates the subject identification number used for each patient specimen.

B. Genes involved in rearrangements with 5' rearrangement partner listed first.

C. Genomic position of the rearrangement breakpoint for the 5' partner.

D. Genomic position of the rearrangement breakpoint for the 3' partner.

E,F. Forward and Reverse primer sequences respectively, shown 5' to 3'.

* Indicates nearest gene when rearrangement occurs in intergenic space

Supplementary Table 4. Characteristics of rearrangement junctions identified in patient, reference, and cell line samples

Patient	Junctions

^A Patient Bank #	^B Junctions	^C Gene 1	DGene 2	^E Orientation	FIntron	^G Reading Frame	^H FISH	ERG Stain	Reads	^k Junction Reads	LSequence Flanking Breakpoint
4	TMPRSS2-ERG	chr21:41794326	chr21:38753597	s/s	1-3	IF	D	+	2	1	CGTGGTGGGCTTCATGTTGATTAATTGATTTTCAAAAATTATCTGTGGCT
10	TMPRSS2-ERG	chr21:41792775	chr21:38792757	s/s	1-3	IF	n/a	n/a	6	1	GGGTGGCTACTGGCAGTGCCTGTGCCTGGCTGTTACTGAACATGGCCATC
70	TMPRSS2-MPRL12	chr21:41783383	chr17:77283404	s/a		ND	n/a	n/a	6	3	CCTCCTTCTTCCAGAGACTAAGCAATCCAGTTTACACAACAGGCTCTACAGT
	HAP1*-TMPRSS2	chr17:37144837	chr21:41779769	a/s		ND			19	5	GCCAGTGCATTAGGGCAGCGCTGACTGGTTTCCTGTTAAGCCCTCGC
	TMPRSS2-HAP1*	chr21:41794230	chr17:37149531	s/s		ND			12	4	GAAGCCGGCCCCACTCTACTTCAGGATACCTAACACAGGTGAGCCCCACC
	TMPRSS2-SWI5	chr21:41779767	chr9:130085490	s/s		ND			5	2	CCCCTCTCCTAAGGGCCATCCTTGGTG AGGACGCTGAGGCAaGAGAATTA
82	EIF3K-TMPRSS2	chr19:43815389	chr21:41788584	a/s		ND	n/a	n/a	5	2	TTTACTGCTGCACGCCCAGTTTAGCTGTAGTACCTTTTTAATATCTTGAACCCG
	NF1-TMPRSS2	chr17:26590306	chr21:41774186	a/s		ND			13	4	GTCCATTTTCCATGAAAGCCCTTAGCATATCGGCACATGAGTGCTGGGGA
134	TMPRSS2-ERG	chr21:41795069	chr21:38803391	s/s	1-3	IF	n/a	n/a	8	2	CTCTGGTCTTCCATTCCAAAGTCCATGTTAGGGATGGGAGAGGAGGGTA
	ACPP-TMPRSS2	chr3:133565489	chr21:41795096	a/s		ND			6	2	CCCACTCTGTGTTTATTTTATTTTTACTCTGGTCTTCCATTCCAAAGTCCATGG
144	TMPRSS2-ERG	chr21:41779065	chr21:38796877	s/s	5-3	ND	n/a	n/a	7	5	TTATGTTACCTGAAGAGTTGCTCTCTATATTTTTACTTTGCCTTTTTATT
	TMPRSS2-TMPRSS2	chr21:41789960	chr21:41781176	s/s		ND			5	0	AAAGACTGGAGGAACTAAAAGTtcgagcaAAGATTTTGTGACCTCTGACC
199	TMPRSS2-ERG	chr21:41786586	chr21:38748712	s/a	3-3	ND	n/a	n/a	6	3	CGTGGAAATGTGTCCATCATGTGGACTGGACCAGGGACCAAAAAGGCTATAA
206	TMPRSS2-ERG	chr21:41793812	chr21:38763138	s/s	1-3	IF	n/a	n/a	9	0	TGGGAAATATTCTTCTGTGGtTAGACTACTCATGTCTTGTAAGAGGCGTT
214	TMPRSS2-ERG	chr21:41797467	chr21:38749676	s/s	1-3	IF	n/a	n/a	3	1	TTCGGTACATTCTAAGGTAGCTCCAGTGGGGGGGGGGGG
341	TMPRSS2-ERG	chr21:41785537	chr21:38798302	s/s	3-3	ND	n/a	n/a	4	1	ACTCTCCTGGTGAGAACTGTACAGATTGTTTAAAAGTTATTTTAAGCTAA
353	TMPRSS2-ERG	chr21:41789358	chr21:38799621	s/s	2-3	IF	n/a	n/a	1	1	AGTGACGTGATCTCCATTCACTGCAACACGACTCATGTCTGGTTTCTAGCTGGCA
357	TMPRSS2-ERG	chr21:41794131	chr21:38797829	s/s	1-3	IF	n/a	n/a	4	2	TCTTTTCCTTCAGCAGGTTGTTAAATACACACCCTCATACACACAC
372	TMPRSS2-ERG	chr21:41797622	chr21:38784847	s/s	1-3	IF	-	-	1	0	AGTTCACTGGTGATGATGCATGGTTTTGCATTGCGTAGTCCTCTTTGG
	SLC45A3-ERG	chr1:203908669	chr21:38798718	s/s		IF			1	1	TGAACATGAACCCTTCCTGAATGTCATGTATAATTCATAAAATAGTGACAGAT
535	TMPRSS2-ERG	chr21:41790462	chr21:38748584	s/s	2-3	IF	D	+	1	1	AGATGTGGGCTGGTGGGGCCCAAGTGATAATTTGCACACAAGAGACTGTG
558	TDRD7*-TMPRSS2	chr9:99213679	chr21:41794880	s/s		ND	n/a	n/a	2	0	GTAAAATACATATAGTGTGCTAGATATGAGTATTGATTCTTCACTCCTTT
	TMPRSS2-TMPRSS2	chr21:41771407	chr21:41780399	s/s		ND			7	1	TACAATAATAGGTCCTTCCTGACACCT/AATAAGTTTTAAAGGAAGAGGAA
	MX1-TMPRSS2	chr21:41752881	chr21:41775109	s/s		ND			6	3	CTAGAAACTGACACATGCTGAACATaaagGAGCATAAGGTCTTCAGCACT
580	MX1-TMPRSS2	chr21:41752819	chr21:41775553	s/s		ND	n/a	n/a	2	1	AAGGGATTTTCAGCCCTCAGACTTATTGAGATCTAATTTATGTACCATAA
652	TMPRSS2-ERG	chr21:41792424	chr21:38775810	s/s	1-3	IF	Т	+	4	1	TGTTCTGTATAGGGGATATCACAGCAACCAAACACCACTTTCATTTTAT
	ERG-ERG	chr21:38830023	chr21:38872708	s/s		ND			2	1	TCTCCAACCAAACCTGACTTATACTTAG/CCAGTGATAAAAGGAGACTCAAAC
675	TMPRSS2-ERG	chr21:41798800	chr21:38789416	s/s	1-3	IF	D	+	13	0	TGAAGGCCCAGGTGCATTTCTGCTCTCTGGTTCCCCGGACCTTTTAAGGAG
682	TMPRSS2-ERG	chr21:41796887	chr21:38791968	s/s	1-3	IF	D	+	6	1	GCAAATTGCATTGTATTTCAAATTTtAAACTTCAAACTCAGAAAAAAATT
733	TMPRSS2-ERG	chr21:41794661	chr21:38816717	s/s	1-3	IF	n/a	n/a	7	3	TCTGTCCTGCTATGAGACAAGAATGCGGACTTTGTTTTGTTATTCTGTCTCCA
738	THSD7A-TMPRSS2	chr7:11743984	chr21:41788390	s/s		ND	n/a	-	4	2	ATTAGTACAATATTTTTGAAAATTTTCAACTGTTTAGGGGTCACCACCAG
780	TMPRSS2-ERG	chr21:41792243	chr21:38787029	s/s	1-3	IF	n/a	n/a	1	0	CCTGGCCGCTGCACTTACAATTGCACaGCTGCTTGCGATGTTTCACTCAT
808	TMPRSS2-ERG	chr21:41790468	chr21:38780232	s/s	2-3	IF	т	+	38	13	AGAAGGGGAAGATGTGGGCTGGTGGGGCCCTGCAAACATCACAAAGAGCACT
	TMPRSS2-ERG	chr21:41783369	chr21:38805184	a/a	Exon4-Intron3	ND			10	5	GAGGAAGGTCCCCAGGGTCAAGGTGAGATGTTTAATACCTACAAATACAG
	SPATA5L1*-TMPRSS2	chr15:43481074	chr21:41771055	a/s		ND			9	1	TATCTACGTTATTAAGACAACTGCCTGACGCTCAGTGAAATAATTCAGGT
814	TMPRSS2-ERG	chr21:41792783	chr21:38784111	a/s	1-3	ND	т	+	9	0	TCCCCGAGAACGTCCACAGGCACATTTTTCTTCTAAGACTAAAAT
	MORC3-TMPRSS2	chr21:36673099	chr21:41762272	s/a		ND			12	0	TGTAAGGACTGAGACAAATGTCAGCgtGCAGATATGTCTATGACAACCTG
816	TMPRSS2-ERG	chr21:41791718	chr21:38797360	s/s	2-3	IF	Т	+	10	4	ACACAGCTGCCCAGGTGAGTCGCAAGAGCAGACTGAGATAGGCTTCCCGA
	C1qTNF1-TMPRSS2	chr17:74540999	chr21:41791286	s/s		ND			6	2	GIGGCICGIAGCCACIGIACIGACICAGGAAIIIIICAGGGACAAACCIGC
	C1qTNF1-TMPRSS2	chr17:74540985	chr21:41764084	a/s		ND			9	2	AAATGTTTAAGTCAGTACAGTGGCCTGAGGACGGGCATCCTGGACCCATG
	TMPRSS2-TMPRSS2	chr21:41764044	chr21:41796946	s/a		ND			13	3	TGGACCCATGGTGGCCACATCTAAGC/AAATGAGTAAGATGAAATTTAGCA
981	TMPRSS2-ERG	chr21:41795091	chr21:38796469	s/s	1-3	IF	D	+	15	3	GAATGTTACTGGCAGCATCACTCTGAAACAGGCCCTTTGGAGAGGGG
989	FAM177A1-ERG	chr14:34584342	chr21:38756847	a/s		ND	D	+	5	1	GCCAAAGTTCTTCTCCTTACCCGGACACACTGGGAGATAGGCAGGGAGTC
995	TMPRSS2-ERG	chr21:41780434	chr21:38793491	a/a	5-3	ND	Т	+	5	0	TAAAGGAAGAGGAAGAAAAGTAAAAGGAGGAGAAAGGAGAAAGATAAGGGG
	TMPRSS2-TMPRSS2	chr21:41762098	chr21:417/3/13	a/s		ND			12	3	CCGAGCIGGCICCGIGICACIGA/CAGIGAIICCIAACGGGGGIGGGGCGG
	TMPRSS2-TMPRSS2	chr21:41780464	chr21:41790293	s/a		ND			10	3	GTACGGAATGGGGCTCTGCAGATGGCAG/TGTGGCCATACATAAATGTGAA
	TMPRSS2-TMPRSS2	chr21:41786495	chr21:41773659	s/s		ND			2	1	GCGGTGGTTACCATTTACCAGCACTGCCT/TGCTTTTATAAGGGAAGAATATGA
	ERG-ERG	chr21:38/934/2	chr21:38793604	a/s		ND		6 mm m	5	0	
1164	TMPRSS2-ETV5	chr21:41797113	chr3:18/2/8460	s/s	1-7	IF	T, TMPRSS2	+ for ETV5	11	1	GLAGIGGCIIGIGIGIGAIILLILLILLIGAAGILIGLLIGIGLAIIIIGGA
1273	ELK4-ERG	chr1:203858916	chr21:38746745	a/s	2.2	ND	T,ERG	-	- 1	0	
1355	TMPRSS2-ERG	chr21:41788584	chr21:38/99/38	s/s	2-3	IF	n/a	+	1	0	
1422	TMPRSS2-ERG	chr21:41795553	chr21:38788443	a/s	1-3	IF	n/a	+	7	2	
1538	TMPRSS2-ETV4	chr21:41795742	chr17:38978855	s/s	1-3	IF	T,TMPRSS2	+ tor ETV4	17	0	
1665	TMPRSS2-ERG	cnr21:41791298	cnr21:38729732	s/s	2-4	IF.	n/a	n/a	1	0	
1863	HMPKSSZ-HMPKSSZ	cnr21:41/61944	cnr21:41///674	a/s		NU	n/a	n/a	5	U	
VCAP Cell Line Junctions											
VCAP	TMPRSS2-ERG	chr21:41779893	chr21:38798223	a/s	5-3	ND	т	+	28	4	CTCCAGGAGGTTAGGACTGCATACATTGACCTATTTGGAGTCCTTGATAA
	TMPRSS2-TMPRSS2	chr21:41793823	chr21:41779387	s/a		ND			19	7	GCTTTGTGCCAGCAACTGGGAAATA/AGTGAGCCATGGTCATGCCTCTGCA
Reference Case Junctions											
Case 45	TMPRSS2-ERG	chr21:41800806	chr21:38776561	s/s	1-3	IF	n/a	+	18	5	CGAGGAATCCTGGTGGCTTGTTTGGGAGAGGGCAGAGGTTGCAGTCGAG
Case 66	TMPRSS2-ERG	chr21:41795687	chr21:38804491	s/s	1-3	IF	n/a	+	24	11	TGAGAGCCATCATCCCGGTCCTTTAAGCCTCAACTTCCATGGGTCACTAC
Case 77	TMPRSS2-ERG	chr21:41799367	chr21:38790118	s/s	1-3	IF	n/a	+	21	12	TATGACATGGCCAGAGTGCCTGGAACTATTTCCAGTCTCAATGCACTCTGTTTTT

A. Patient Bank # indicates the subject identification number used for each patient specimen.

B. Genes involved in rearrangements with 5' rearrangement partner listed first.

C. Genomic position of the rearrangement breakpoint for the 5' partner.

D. Genomic position of the rearrangement breakpoint for the 3' partner.

E. Refers to the strand orientation of Gene1/Gene2 in the rearrangement; s = sense and a = antisense.

F. Indicates the introns involved in a TMPRSS2-ETS gene rearrangement.

G. Reading frame denoted as being In-Frame (IF) or Non-determinable (ND)

H. Rearrangement status by FISH; Rearrangement with Deletion (D) or with Translocation (T)

I. ERG staining status by IHC; additional ETS rearrangement staining by CISH

J. Number of paired-end reads flanking or directly overlapping the rearrangement junction.

K. Number of reads directly overlapping the rearrangement junction.

L. Confirmed nucleotide resolution of rearrangemnt junctions with relevant architecture; TMPRSS2, ERG, and

Partner Genes are shown with microhomologies and insertions at or near the junction. Microhomologies were

defined as being one or more ambiguously assignable nucleotides at the junction and three or more

nucleotides in sequences flanking the junction that had identity in both partner sequences.

* Indicates nearest gene when rearrangement occurs in intergenic space