Supplemental Online Material for

Discovery of Non-ETS Gene Fusions in Human Prostate Cancer using Next Generation RNA Sequencing.

Dorothee Pflueger[1*], Stéphane Terry[1*], Andrea Sboner[2*], Lukas Habegger[3], Raquel Esgueva[1], Pei-Chun Lin[1], Maria A. Svensson[1], Naoki Kitabayashi[1], Benjamin J. Moss[1], Theresa Y. MacDonald[1], Xuhong Cao[4], Terrence Barrette[4,5], Ashutosh K. Tewari[6], Mark S. Chee[7], Arul M. Chinnaiyan[4,5,8,9,10], David S. Rickman[1], Francesca Demichelis[1,11], Mark B. Gerstein[2,3,12#], and Mark A. Rubin[1#]

[1] Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY, USA
[2] Molecular Biophysics and Biochemistry Department, Yale University, New Haven, CT, USA
[3] Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA
[4] Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, USA
[5] Department of Pathology, University of Michigan, Ann Arbor, USA
[6] Department of Urology, Weill Cornell Medical College, New York, NY, USA
[7] Prognosys Biosciences, Inc., La Jolla, CA, USA
[8] Howard Hughes Medical Institute
[9] Department of Urology, University of Michigan, Ann Arbor, USA
[10] Comprehensive Cancer Center, University of Michigan, Ann Arbor, USA
[11] Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA
[12] Department of Computer Science, Yale University, New Haven, CT, USA
* equal contribution
[#] Shared senior authorship

Supplemental Online Material includes

Supplemental Text

Candidate filtering system in FusionSeq (Sboner et al. IN PRESS):

The filters eliminate candidates artificially created by several sources of errors:

- mis-alignment of the reads: the sequences of the mapped reads and of the genes involved in the fusion are compared to eliminate artificial cases due to sequence similarity. Moreover, candidates supported by reads covering repetitive regions are eliminated;
- random-pairing of transcript fragments during sample preparation: the insert-size of the fusion candidates is compared with the insert-size distribution of the transcriptome norm. Abnormal insert-sizes indicates a potential artifact due to random-pairing since the insert-size of the transcript fragment are supposed to be similar given the size selection step in the experimental protocol;
- combination of random pairing and mis-alignments: the consistency of the expression of the genes generating the fusion transcript is taken into account. Indeed, if a fusion transcript is detected, the two individual genes generating the transcript should be expressed at least at similar levels. Moreover, the similarity of the reads with highly abundant transcripts, such as ribosomal genes, is considered;
- PCR amplification: candidates with supporting inter-transcript reads concentrated on very small regions are discarded because likely product of PCR amplification;
- mis-annotation: the consistency of the gene annotation set is checked for potential errors, such as isoforms of the same genes reported as being part of different genes.

Scoring system in FusionSeq:

The scoring of chimeric transcripts by FusionSeq (Sboner et al. IN PRESS) enables sorting and prioritizing the experimental validation comparison between samples taking into account the variance in the number of mapped reads across samples and the individual genes' expression levels (overall number of intra-gene paired-end reads). The scoring methods provide "confidence values" for the candidate chimeric transcripts. The three scores that FusionSeq provides for each candidate are:

1. SPER (supportive paired-end reads per million mapped reads); a normalized measure of fusion transcript abundance;

2. DASPER (Difference between the observed and analytically calculated expected SPER) which indicates how many normalized inter-transcript paired-end reads more than the expected number are observed. The higher DASPER is, the more likely the candidates is true. This measure helps identify potential artifacts generated by random-pairing of highly expressed genes;.

3. RESPER (Ratio of Empirically computed SPERs) which compares the observed SPER with the SPERs of the other candidates in the same sample, allowing better comparison between samples, especially when different insert-sizes are used.

Altogether, these measurements provide the user with a way to enrich for true candidates and prioritize the validation. More detail can be found in (Sboner et al. IN PRESS).

Exclusion of erroneous fusion candidates:

Class of *cis* fusion transcript candidates. Typically, alignment tools map sequence reads to genomic locations independent of strand orientation. When analyzing paired-end reads,

FusionSeq cannot take into account if the reads were generated by the + or the – DNA strand. Thoroughly, we investigated one cis candidate, *SLC16A8-PICK1*, which was present in several prostate cancers and benign prostate. Guided by the paired-end reads' location, we designed several PCR assays interrogating for a *SLC16A3-PICK1* fusion transcript (**Supplemental Fig. S2A**) but also investigating for the possibility of ongoing transcription beyond the annotated boundaries of both genes (**Supplemental Fig. S2B**). Only when using the primer pairs for ongoing transcription of *PICK1*, specific PCR bands were visible (**Supplemental Fig. S2C**). PCR primer f is located in the breakpoint region as suggested by the Junction Sequence Identifier, but this region contains repetitive sequences and this might be an explanation for the PCR band smear that is visible on the gel. Finally, by sequencing the two strongest bands, we were able to show that *PICK1* (+ strand) has at least three non-annotated additional 3 prime exons that are overlapping with *SLC16A3* 5 prime exons (- strand) (**Supplemental Fig. S2D**). This indicates that both paired-end reads of the *PICK1-SLC16A8* candidate belong to *PICK1* and were inaccurately mapped to *SLC16A3*, currently being the only annotated gene in the region of the second paired-end read. These data also suggest that more cis events may result from gene extending over their annotated gene boundaries**.**

Highly expressed tissue-specific or ubiquitously expressed genes. Both *KLK2* and *HNRNPA2B1* have been reported as 5 prime partners in *ETS* gene fusions in prostate cancer (Kumar-Sinha et al. 2008). The chimeric transcript *HNRNPA2B1-PLA2G2A* was identified as erroneous since the paired-end reads were piling up in one position of the last exon of *HNRNPA2B1* and *PLA2G2A*, respectively, instead of being randomly distributed across the exon. This suggests that a *HNRNPA2B1* and a *PLA2G2A* transcript were artificially connected

during the ligation step in the library preparation and this artificial chimeric transcript was preferentially amplified during subsequent PCR. *KLK2*, and similarly *KLK3* (*PSA*), are nominated in several potential gene fusions with varying gene fusion partners even in the same sample. With the exception of one KLK2 candidate (i.e., *KLK2-ETV1*), they all have low or even negative DASPER scores and do not appear as top candidates in the lists.

FKBP5* as novel 5 prime fusion partner of *ERG:

In one tumor sample the top candidate harboured 371 PE reads joining *FKBP5*, a highly androgen-regulated gene on chromosome 6p21.3, with *ERG* (21q22.2). Interestingly, the second highest fusion candidate in this sample connected *TMPRSS2* and *ERG* (247 PE reads) and the fourth highest candidate involved both of the putative 5 prime partners *FKBP5* and *TMPRSS2* (**Supplemental Fig. S3A**). Since FISH confirmed that only one *ERG* allele was rearranged (**Supplemental Fig. S3B**), we explored for the existence of a more complex triple fusion transcript (**Supplemental Fig. S3C** & **S3D**). Each of the single chimeric candidates suggested by FusionSeq (*FKBP5-ERG*, *TMPRSS2-FKBP5* and "*TMPRSS2-ERG*") was confirmed by RT-PCR (**Supplemental Fig. S3C**). The "*TMPRSS2-ER*G" candidate was explained by the expression of a *TMPRSS2-FKBP5-ERG* triple fusion transcript. Perhaps one of the most appealing aspects of this triple fusion is the indication that *FKPB5* harbors several putative Androgen Responsive Elements (AREs) downstream from the Transcription Start Site. In fact, a recent study demonstrated that at least 3 of these AREs are grouped in the intron 7 of *FKBP5* and *in vitro* settings showed that this region was particularly responsive to androgen exposure (Makkonen et al. 2009). If these intronic sequences are conserved in the *TMPRSS2-FKBP5-ERG* DNA based alteration, they might contribute to the induction of the chimeric transcript in conjunction with

AREs located in *TMPRSS2* in this tumor. Both *ERG* and *FKBP5* are known to exhibit alternative splicing and accordingly, we found at least 4 different fusion transcripts. Two of the transcript variants are predicted to encode 5' truncated ERG. The remaining two retain 5 prime to the truncated ERG a small part of FKBP5 (and TMPRSS2) which encodes for a conserved domain of the FKBP_C superfamily. We believe the first mentioned variants encoding truncated ERG to exhibit similar functional roles to a conventional *TMPRSS2-ERG* gene fusion, which has been shown to increase invasive potential *in vitro* and *in vivo* (Klezovitch et al. 2008; Tomlins et al. 2008). The exact functional role for the latter two variants is still not known and it is unclear if the FKBP5 domain contributes to its putative role in cancer progression. Also of interest is that the distribution of the PE reads for the *TMPRSS2-FKBP5* fusion transcript further suggested that these two 5 prime partners underwent a balanced translocation (for PCR validation see **Supplemental Fig. S3C**) and we validated this fusion at the DNA level using a fusion specific FISH assay (**Supplemental Fig. S3B**). Both 5 prime partners show rearrangement on the genomic level (**Supplemental Fig. S3B**).

mRNA junction sequence (encompassing primers from Supplemental Table S6) and predicted ORF and protein domains of novel gene fusion proteins:

Fusion gene	Predicted ORF
KLK2-ETV1	5' truncated ETV1; fusion protein
TMPRSS2-FKBP5-ERG	5' truncated ERG; fusion proteins
FKBP5-TMPRSS2	fusion protein
TNPO1-IKBKB	wild type IKBKB
CDKN1A-CD9	5' truncated CD9
ALG5-PIGU	fusion proteins
MIER2-RSRC2	fusion protein

KLK2-ETV1:

Variant 1:
GTTCTCTCCATCGCCTTGTCTGTGGGGTGCACTGATTTCGCCGCCAGCTTTCTGAACCCTGTAACTCCTT
TCCTCCTTTGCCGACGATGCCAAGGGAAGGACGTCCTATGTACCAACGCCAGATGTCTGAGCCAAACA
TCCCCTTCCCACCACAAGGCTTTAAGCAGGAGTACCACGACCCAGTGTATGAACACAACACCATGGTT
GGCAGTGCGGCCAGCCAAAGCTTTCCCCCTCCTCTGATGATTAAACAGGAACCCAGAGATTTTGCATA
TGACTCAGAAGTGCCTAGCTGCCACTCCATTTATATGAGGCAAGAAGGCTTCCTGGCTCATCCCAGCA
GAACAGAAGGCTGTATGTTTGAAAAGGGCCCCAGGCAGTTTTATGATGACACCTGTGTTGTCCCAGAA
AAATTCGATGGAGACATCAAACAAGAGCCAGGAATGTATCGGGAAGGA

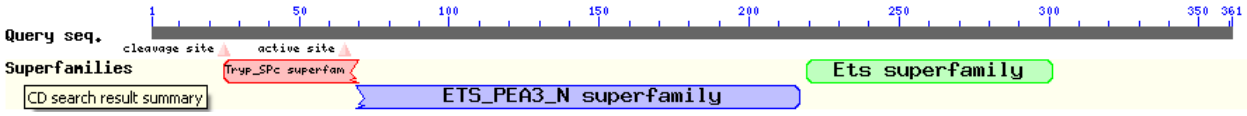
Variant 2:
GTTCTCTCCATCGCCTTGTCTGTGGGGTGCACTGGTGCCGTGCCCCTCATCCAGTCTCGGATTGTGGGA
GGCTGGGAGTGTGAGAAGCATTCCCAACCCTGGCAGGTGGCTGTGTACAGTCATGGATGGGCACACTG
TGGGGGTGTCCTGGTGCACCCCCAGTGGGTGCTCACAGCTGCCCATTGCCTAAAGAAATTTCGCCGCC
AGCTTTCTGAACCCTGTAACTCCTTTCCTCCTTTGCCGACGATGCCAAGGGAAGGACGTCCTATGTACC
AACGCCAGATGTCTGAGCCAAACATCCCCTTCCCACCACAAGGCTTTAAGCAGGAGTACCACGACCCA
GTGTATGAACACAACACCATGGTTGGCAGTGCGGCCAGCCAAAGCTTTCCCCCTCCTCTGATGATTAA
ACAGGAACCCAGAGATTTTGCATATGACTCAGAAGTGCCTAGCTGCCACTCCATTTATATGAGGCAAG
AAGGCTTCCTGGCTCATCCCAGCAGAACAGAAGGCTGTATGTTTGAAAAGGGCCCCAGGCAGTTTTAT
GATGACACCTGTGTTGTCCCAGAAAAATTCGATGGAGACATCAAACAAGAGCCAGGAATGTATCGGG
AAGGA

>KLK2 wild type
MWDLVLSIALSVGCTGAVPLIQSRIVGGWECEKHSQPWQVAVYSHGWAHCGGVLVHPQWVLTAA
HCLKKNSQVWLGRHNLFEPEDTGQRVPVSHSFPHPLYNMSLLKHQSLRPDEDSSHDLMLLRLSE
PAKITDVVKVLGLPTQEPALGTTCYASGWGSIEPEEFLRPRSLQCVSLHLLSNDMCARAYSEKV
TEFMLCAGLWTGGKDTCGGDSGGPLVCNGVLQGITSWGPEPCALPEKPAVYTKVVHYRKWIKDT
IAANP

>ETV1 wild type
MDGFYDQQVPYMVTNSQRGRNCNEKPTNVRKRKFINRDLAHDSEELFQDLSQLQETWLAEAQVP
DNDEQFVPDYQAESLAFHGLPLKIKKEPHSPCSEISSACSQEQPFKFSYGEKCLYNVSAYDQKP
QVGMRPSNPPTPSSTPVSPLHHASPNSTHTPKPDRAFPAHLPPSQSIPDSSYPMDHRFRRQLSE
PCNSFPPLPTMPREGRPMYQRQMSEPNIPFPPQGFKQEYHDPVYEHNTMVGSAASQSFPPPLMI
KQEPRDFAYDSEVPSCHSIYMRQEGFLAHPSRTEGCMFEKGPRQFYDDTCVVPEKFDGDIKQEP
GMYREGPTYQRRGSLQLWQFLVALLDDPSNSHFIAWTGRGMEFKLIEPEEVARRWGIQKNRPAM
NYDKLSRSLRYYYEKGIMQKVAGERYVYKFVCDPEALFSMAFPDNQRPLLKTDMERHINEEDTV
PLSHFDESMAYMPEGGCCNPHPYNEGYVY

>Variant 1: ETV1 truncated
MPREGRPMYQRQMSEPNIPFPPQGFKQEYHDPVYEHNTMVGSAASQSFPPPLMIKQEPRDFAYD
SEVPSCHSIYMRQEGFLAHPSRTEGCMFEKGPRQFYDDTCVVPEKFDGDIKQEPGMYREGPTYQ
RRGSLQLWQFLVALLDDPSNSHFIAWTGRGMEFKLIEPEEVARRWGIQKNRPAMNYDKLSRSLR
YYYEKGIMQKVAGERYVYKFVCDPEALFSMAFPDNQRPLLKTDMERHINEEDTVPLSHFDESMA
YMPEGGCCNPHPYNEGYVY

>Variant 2: KLK2-ETV1
MWDLVLSIALSVGCTGAVPLIQSRIVGGWECEKHSQPWQVAVYSHGWAHCGGVLVHPQWVLTAA
HCLKKFRRQLSEPCNSFPPLPTMPREGRPMYQRQMSEPNIPFPPQGFKQEYHDPVYEHNTMVGS
AASQSFPPPLMIKQEPRDFAYDSEVPSCHSIYMRQEGFLAHPSRTEGCMFEKGPRQFYDDTCVV
PEKFDGDIKQEPGMYREGPTYQRRGSLQLWQFLVALLDDPSNSHFIAWTGRGMEFKLIEPEEVA
RRWGIQKNRPAMNYDKLSRSLRYYYEKGIMQKVAGERYVYKFVCDPEALFSMAFPDNQRPLLKT
DMERHINEEDTVPLSHFDESMAYMPEGGCCNPHPYNEGYVY



TMPRSS2-FKBP5-ERG

Variant 1:
CGCGAGCTAAGCAGGAGGCGGAGGCGGAGGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGCGC
GGCAGTCCACCTGGAAGGCCGCTGTGGTGGAAGGATGTTTGACTGCAGAGATGTGGCATTCACTGTGG
GCGAAGGAGAAGACCACGACATTCCAATTGGAATTGACAAAGCTCTGGAGAAAATGCAGCGGGAAGA
ACAATGTATTTTATATCTTGGACCAAGGAAGCCTTATCAGTTGTGAGTGAGGACCAGTCGTTGTTTGAG
TGTGCCTAC

Variant 2:
CGCGAGCTAAGCAGGAGGCGGAGGCGGAGGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGCGC
GGCAGTCCACCTGGAAGGCCGCTGTGGTGGAAGGATGTTTGACTGCAGAGATGTGGCATTCACTGTGG
GCGAAGGAGAAGACCACGACATTCCAATTGGAATTGACAAAGCTCTGGAGAAAATGCAGCGGGAAGA
ACAATGTATTTTATATCTTGGACCAAGATATGGTTTTGGAGAGGCAGGGAAGCCTAAATTTGGCATTG
AACCTAATGCTGAGCTTATATATGAAGTTACACTTAAGAGCTTCGAAAAGGAAGCCTTATCAGTTGTG
AGTGAGGACCAGTCGTTGTTTGAGTGTGCCTAC

Variant 3:
CGCGAGCTAAGCAGGAGGCGGAGGCGGAGGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGCGC
GGCAGGTCATATTGAACATTCCAGATACCTATCATTACTCGATGCTGTTGATAACAGCAAGATGGCTTT
GAACTCAGGGTCACCACCAGCTATTGGACCTTACTATGAAAACCATGGATACCAACCGGAAAACCCCT
ATCCCGCACAGCCCACTGTGGTCCCCACTGTCTACGAGGTGCATCCGGCTCAGTACTACCCGTCCCCCG
TGCCCCAGTACGCCCCGAGGGTCCTGACGCAGGCTTCCAACCCCGTCGTCTGCACGCAGCCCAAATCC
CCATCCGGGACAGTGTGCACCTCAATCCACCTGGAAGGCCGCTGTGGTGGAAGGATGTTTGACTGCAG
AGATGTGGCATTCACTGTGGGCGAAGGAGAAGACCACGACATTCCAATTGGAATTGACAAAGCTCTG
GAGAAAATGCAGCGGGAAGAACAATGTATTTTATATCTTGGACCAAGGAAGCCTTATCAGTTGTGAGT
GAGGACCAGTCGTTGTTTGAGTGTGCCTAC

Variant 4:
CGCGAGCTAAGCAGGAGGCGGAGGCGGAGGCGGAGGGCGAGGGGCGGGGAGCGCCGCCTGGAGCGC
GGCAGGTCATATTGAACATTCCAGATACCTATCATTACTCGATGCTGTTGATAACAGCAAGATGGCTTT
GAACTCAGGGTCACCACCAGCTATTGGACCTTACTATGAAAACCATGGATACCAACCGGAAAACCCCT
ATCCCGCACAGCCCACTGTGGTCCCCACTGTCTACGAGGTGCATCCGGCTCAGTACTACCCGTCCCCCG
TGCCCCAGTACGCCCCGAGGGTCCTGACGCAGGCTTCCAACCCCGTCGTCTGCACGCAGCCCAAATCC
CCATCCGGGACAGTGTGCACCTCAATCCACCTGGAAGGCCGCTGTGGTGGAAGGATGTTTGACTGCAG
AGATGTGGCATTCACTGTGGGCGAAGGAGAAGACCACGACATTCCAATTGGAATTGACAAAGCTCTG

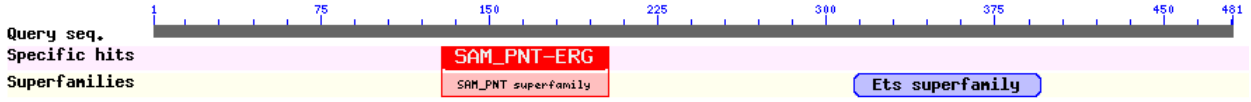
GGAAGCCTAAATTTGGCATTGAACCTAATGCTGAGCTTATATATGAAGTTACACTTAAGAGCTTCGAAAAGGAAGCCTTATCAGTTGTGAGTGAGGACCAGTCGTTGTTTGAGTGTGCCTAC

>TMPRSS2 wild type
MALNSGSPPAIGPYYENHGYQPENPYPAQPTVVPTVYEVHPAQYYPSPVPQYAPRVLTQASNPV
VCTQPKSPSGTVCTSKTKKALCITLTLGTFLVGAALAAGLLWKFMGSKCSNSGIECDSSGTCIN
PSNWCDGVSHCPGGEDENRCVRLYGPNFILQVYSSQRKSWHPVCQDDWNENYGRAACRDMGYKN
NFYSSQGIVDDSGSTSFMKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCIACGVNLNSSRQSRI
VGGESALPGAWPWQVSLHVQNVHVCGGSIITPEWIVTAAHCVEKPLNNPWHWTAFAGILRQSFM
FYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQPEQLCWISGW
GATEEKGKTSEVLNAAKVLLIETQRCNSRYVYDNLITPAMICAGFLQGNVDSCQGDSGGPLVTS
KNNIWWLIGDTSWGSGCAKAYRPGVYGNVMVFTDWIYRQMRADG

>FKBP5 wild type
MTTDEGAKNNEESPTATVAEQGEDITSKKDRGVLKIVKRVGNGEETPMIGDKVYVHYKGKLSNG
KKFDSSHDRNEPFVFSLGKGQVIKAWDIGVATMKKGEICHLLCKPEYAYGSAGSLPKIPSNATL
FFEIELLDFKGEDLFEDGGIIRRTKRKGEGYSNPNEGATVEIHLEGRCGGRMFDCRDVAFTVGE
GEDHDIPIGIDKALEKMQREEQCILYLGPRYGFGEAGKPKFGIEPNAELIYEVTLKSFEKAKES
WEMDTKEKLEQAAIVKEKGTVYFKGGKYMQAVIQYGKIVSWLEMEYGLSEKESKASESFLLAAF
LNLAMCYLKLREYTKAVECCDKALGLDSANEKGLYRRGEAQLLMNEFESAKGDFEKVLEVNPQN
KAARLQISMCQKKAKEHNERDRRIYANMFKKFAEQDAKEEANKAMGKKTSEGVTNEKGTDSQAM
EEEKPEGHV

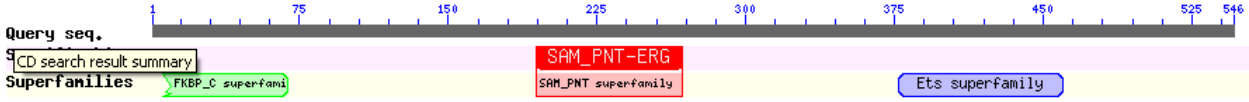
>ERG wild type
MIQTVPDPAAHIKEALSVVSEDQSLFECAYGTPHLAKTEMTASSSSDYGQTSKMSPRVPQQDWL
SQPPARVTIKMECNPSQVNGSRNSPDECSVAKGGKMVGSPDTVGMNYGSYMEEKHMPPPNMTTN
ERRVIVPADPTLWSTDHVRQWLEWAVKEYGLPDVNILLFQNIDGKELCKMTKDDFQRLTPSYNA
DILLSHLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTDLPYEPPRRSAWTGHGHPTPQSKAA
QPSPSTVPKTEDQRPQLDPYQILGPTSSRLANPGSGQIQLWQFLLELLSDSSNSSCITWEGTNG
EFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQP
HPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFFAAPNPYWNSPTGGIYPNTR
LPTSHMPSHLGTYY

>Variant 1: ERG truncated
MYFISWTKEALSVVSEDQSLFECAYGTPHLAKTEMTASSSSDYGQTSKMSPRVPQQDWLSQPPA
RVTIKMECNPSQVNGSRNSPDECSVAKGGKMVGSPDTVGMNYGSYMEEKHMPPPNMTTNERRVI
VPADPTLWSTDHVRQWLEWAVKEYGLPDVNILLFQNIDGKELCKMTKDDFQRLTPSYNADILLS
HLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTGGAAFIFPNTSVYPEATQRITTRPDLPYEP
PRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQRPQLDPYQILGPTSSRLANPGSGQIQLWQFLL
ELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVH
GKRYAYKFDFHGIAQALQPHPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFF
AAPNPYWNSPTGGIYPNTRLPTSHMPSHLGTYY



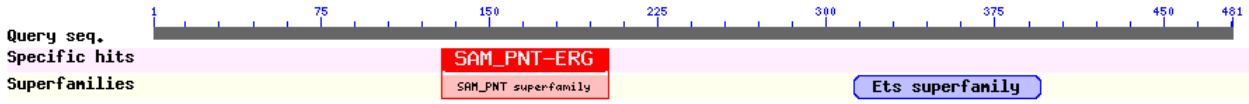
>Variant 2: FKBP5-ERG

MFDCRDVAFTVGEGEDHDIPIGIDKALEKMQREEQCILYLGPRYGFGEAGKPKFGIEPNAELIY
EVTLKSFEKEALSVVSEDQSLFECAYGTPHLAKTEMTASSSSDYGQTSKMSPRVPQQDWLSQPP
ARVTIKMECNPSQVNGSRNSPDECSVAKGGKMVGSPDTVGMNYGSYMEEKHMPPPNMTTNERRV
IVPADPTLWSTDHVRQWLEWAVKEYGLPDVNILLFQNIDGKELCKMTKDDFQRLTPSYNADILL
SHLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTGGAAFIFPNTSVYPEATQRITTRPDLPYE
PPRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQRPQLDPYQILGPTSSRLANPGSGQIQLWQFL
LELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKV
HGKRYAYKFDFHGIAQALQPHPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSF
FAAPNPYWNSPTGGIYPNTRLPTSHMPSHLGTYY



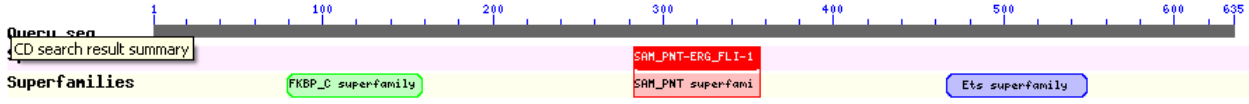
>Variant 3: ERG truncated

MYFISWTKEALSVVSEDQSLFECAYGTPHLAKTEMTASSSSDYGQTSKMSPRVPQQDWLSQPPA
RVTIKMECNPSQVNGSRNSPDECSVAKGGKMVGSPDTVGMNYGSYMEEKHMPPPNMTTNERRVI
VPADPTLWSTDHVRQWLEWAVKEYGLPDVNILLFQNIDGKELCKMTKDDFQRLTPSYNADILLS
HLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTGGAAFIFPNTSVYPEATQRITTRPDLPYEP
PRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQRPQLDPYQILGPTSSRLANPGSGQIQLWQFLL
ELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVH
GKRYAYKFDFHGIAQALQPHPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFF
AAPNPYWNSPTGGIYPNTRLPTSHMPSHLGTYY



>Variant 4: TMPRSS2-FKBP5-ERG

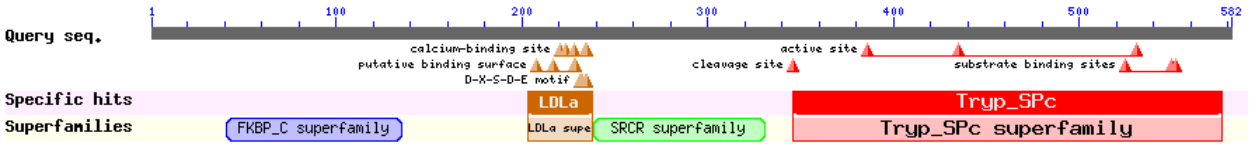
MALNSGSPPAIGPYYENHGYQPENPYPAQPTVVPTVYEVHPAQYYPSPVPQYAPRVLTQASNPV
VCTQPKSPSGTVCTSIHLEGRCGGRMFDCRDVAFTVGEGEDHDIPIGIDKALEKMQREEQCILY
LGPRYGFGEAGKPKFGIEPNAELIYEVTLKSFEKEALSVVSEDQSLFECAYGTPHLAKTEMTAS
SSSDYGQTSKMSPRVPQQDWLSQPPARVTIKMECNPSQVNGSRNSPDECSVAKGGKMVGSPDTV
GMNYGSYMEEKHMPPPNMTTNERRVIVPADPTLWSTDHVRQWLEWAVKEYGLPDVNILLFQNID
GKELCKMTKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTGGAA
FIFPNTSVYPEATQRITTRPDLPYEPPRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQRPQLDP
YQILGPTSSRLANPGSGQIQLWQFLLELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWGERKS
KPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPPESSLYKYPSDLPYMGS
YHAHPQKMNFVAPHPPALPVTSSSFFAAPNPYWNSPTGGIYPNTRLPTSHMPSHLGTYY



<u>FKBP5</u>-TMPRSS2:

ACGGGCAAGGCGGAGAGTGACGGAGTCGGCGAGCCCCCGCGGCGACAGGTTCTCTACTTAAAAGACA
ATGACTACTGATGAAGGTGCCAAGAACAATGAAGAAAGCCCCACAGCCACTGTTGCTGAGCAGGGAG
AGGATATTACCTCCAAAAAAGACAGGGGAGTATTAAAGATTGTCAAAAGAGTGGGGAATGGTGAGGA
AACGCCGATGATTGGAGACAAAGTTTATGTCCATTACAAAGGAAAATTGTCAAATGGAAAGAAGTTTG
ATTCCAGTCATGATAGAAATGAACCATTTGTCTTTAGTCTTGGCAAAGGCCAAGTCATCAAGGCATGG
GACATTGGGGTGGCTACCATGAAGAAAGGAGAGATATGCCATTTACTGTGCAAACCAGAATATGCAT
ATGGCTCGGCTGGCAGTCTCCCTAAAATTCCCTCGAATGCAACTCTCTTTTTTGAGATTGAGCTCCTTG
ATTTCAAAGGAGAGGATTTATTTGAAGATGGAGGCATTATCCGGAGAACCAAACGGAAAGGAGAGGG
ATATTCAAATCCAAACGAAGGAGCAACAGTAGAAAAGACTAAGAAAGCACTGTGCATCACCTTGACC
CTGGGGACCTTCCTCGTGGGAGCTGCGCTGGCCGCTGGCCTACTCTGGAAGTTCATGGGCAGCAAGTG
CTCCAACTCTGGGATAGAGTGCGACTCCTCAGGTACCTGCATCAACCCCTCTAACTGGTGTGATGGCGT
GTCACACTGCCCCGGCGGGGAGGACGAGAATCGGTGTGTTCGCCTCTACGGACCAAACTTCATCCTTC
AGGTGTACTCATCTCAGAGGAAGTCCTGGCACCCTGTGTGCCAAGACGACTGGAACGAGAACTACGG
GCGGGCGGCCTGCAGGGACATGGGCTATAAGAATAATTTTTACTCTAGCCAAGGAATAGTGGATGACA
GCGGA

>FKBP5-TMPRSS2
MTTDEGAKNNEESPTATVAEQGEDITSKKDRGVLKIVKRVGNGEETPMIGDKVYVHYKGKLSNG
KKFDSSHDRNEPFVFSLGKGQVIKAWDIGVATMKKGEICHLLCKPEYAYGSAGSLPKIPSNATL
FFEIELLDFKGEDLFEDGGIIRRTKRKGEGYSNPNEGATVEKTKKALCITLTLGTFLVGAALAA
GLLWKFMGSKCSNSGIECDSSGTCINPSNWCDGVSHCPGGEDENRCVRLYGPNFILQVYSSQRK
SWHPVCQDDWNENYGRAACRDMGYKNNFYSSQGIVDDSGSTSFMKLNTSAGNVDIYKKLYHSDA
CSSKAVVSLRCIACGVNLNSSRQSRIVGGESALPGAWPWQVSLHVQNVHVCGGSIITPEWIVTA
AHCVEKPLNNPWHWTAFAGILRQSFMFYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFN
DLVKPVCLPNPGMMLQPEQLCWISGWGATEEKGKTSEVLNAAKVLLIETQRCNSRYVYDNLITP
AMICAGFLQGNVDSCQGDSGGPLVTSKNNIWWLIGDTSWGSGCAKAYRPGVYGNVMVFTDWIYR
QMRADG

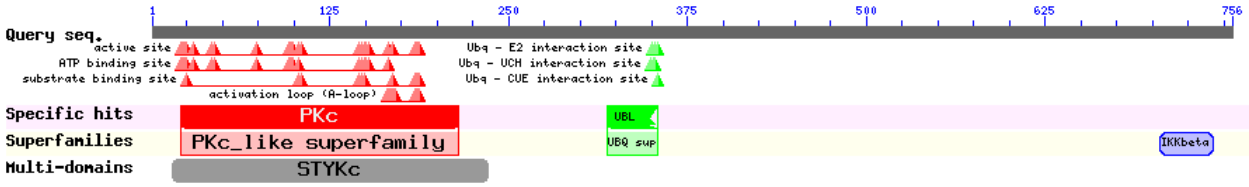


***TNPO1*-IKBKB:**

CAGACACCACCATCCAGAGAACCGTGCAACAAAAACTGGAACAACTTAATCAGTATCCAGACTTTAAC
AACTACTTGATTTTTGTTCTTACAAAATTAAAATCTGAAGAGTTAGCACGACATCAGTATGAGCTGGTC
ACCTTCCCTGACAACGCAGACATGTGGGGCCTGGGAAATGAAAGAGCGCCTTGGGACAGGGGGATTT
GGAAATGTCATCCGATGGCACAATCAGGAAACAGGTGAGCAGAT

>IKBKB wild type
MSWSPSLTTQTCGAWEMKERLGTGGFGNVIRWHNQETGEQIAIKQCRQELSPRNRERWCLEIQI
MRRLTHPNVVAARDVPEGMQNLAPNDLPLLAMEYCQGGDLRKYLNQFENCCGLREGAILTLLSD
IASALRYLHENRIIHRDLKPENIVLQQGEQRLIHKIIDLGYAKELDQGSLCTSFVGTLQYLAPE
LLEQQKYTVTVDYWSFGTLAFECITGFRPFLPNWQPVQWHSKVRQKSEVDIVVSEDLNGTVKFS
SSLPYPNNLNSVLAERLEKWLQLMLMWHPRQRGTDPTYGPNGCFKALDDILNLKLVHILNMVTG
TIHTYPVTEDESLQSLKARIQQDTGIPEEDQELLQEAGLALIPDKPATQCISDGKLNEGHTLDM
DLVFLFDNSKITYETQISPRPQPESVSCILQEPKRNLAFFQLRKVWGQVWHSIQTLKEDCNRLQ
QGQRAAMMNLLRNNSCLSKMKNSMASMSQQLKAKLDFFKTSIQIDLEKYSEQTEFGITSDKLLL
AWREMEQAVELCGRENEVKLLVERMMALQTDIVDLQRSPMGRKQGGTLDDLEEQARELYRRLRE

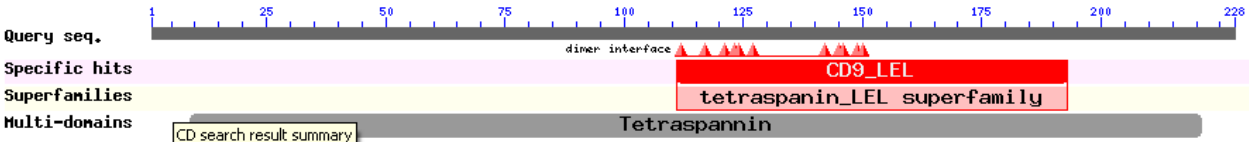
KPRDQRTEGDSQEMVRLLLQAIQSFEKKVRVIYTQLSKTVVCKQKALELLPKVEEVVSLMNEDE
KTVVRLQEKRQKELWNLLKIACSKVRGPVSGSPDSMNASRLSQPGQLMSQPSTASNSLPEPAKK
SEELVAEAHNLCTLLENAIQDTVREQDQSFTALDWSWLQTEEEEHSCLEQAS



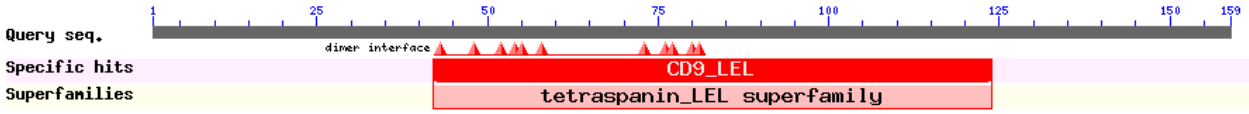
***CDKN1A*-CD9:**

CTGCCGAAGTCAGTTCCTTGTGGAGCCGGAGCTGGGCGCGGATTCGCCGAGGCACCGAGGCACTCAGA
GGAGCTTGCCGGGATTGCTGTCCTTGCCATTGGACTATGGCTCCGATTCGACTCTCAGACCAAGAGCAT
CTTCGAGCAAGAAACTAATAATAATAATTCCAGCTTCTACACAGGAGTCTATATTCTGATCGGAGCCG
GCGCCCTCATGATGCTGGTGGGCTTCCTGGGCTGCTGCGGGGCTGTGCAGGAGTCCCAGTGCATGCTG
GGACTGTTCTTCGGCTTCCTCTTGGTGATATTCGCC

>CD9 truncated long
MPVKGGTKCIKYLLFGFNFIFWLAGIAVLAIGLWLRFDSQTKSIFEQETNNNNSSFYTGVYILI
GAGALMMLVGFLGCCGAVQESQCMLGLFFGFLLVIFAIEIAAAIWGYSHKDEVIKEVQEFYKDT
YNKLKTKDEPQRETLKAIHYALNCCGLAGGVEQFISDICPKKDVLETFTVKSCPDAIKEVFDNK
FHIIGAVGIGIAVVMIFGMIFSMILCCAIRRNREMV



>CD9 truncated short
MMLVGFLGCCGAVQESQCMLGLFFGFLLVIFAIEIAAAIWGYSHKDEVIKEVQEFYKDTYNKLK
TKDEPQRETLKAIHYALNCCGLAGGVEQFISDICPKKDVLETFTVKSCPDAIKEVFDNKFHIIG
AVGIGIAVVMIFGMIFSMILCCAIRRNREMV



***ALG5*-PIGU:**

GAGATCCTGCGTTCACTTATGAAGTGATAGTAGTTGATGATGGCAGTAAAGATCAGACCTCAAAGGTA
GCTTTTAAATATTGCCAGAAATATGGAAGTGACAAAGTACGTGTGATAACCCTGGTGAAGAATCGTGG
AAAAGGTGGAGCGATTAGAATGGGTATATTCAGTTCTCGAGGAGAAAAGATCCTTATGGCAGATGCTG
ATGGAGCCACAAAGTTTCCAGATGTTGAGAAATTAGAAAAGGGGCTAAATGATCTACAGCCTTGGCCT
CGGCAGTACATACCTGTGAAAATGAAGAGCAAAGCCTTCTGGATCTTTTCTTGGGAGTATGCCATGAT
GTATGTGGGAAGCCTAGTGGTAATCATTTGCCTCTCCTTCTTCCTTCTCAGCTCTTGGGATTTCATCCCC
GCAGTCTATGGCTTTATACTTTCTGTTCCAGATCTCACTCCAAACATTGGTCTTTTCTGGTACTTCTTTG
CAGAGATGTTTGAGCACTTCAGCCTCTTCTTTGTATGTGTGTTTCAGATCAACGTCTTCTTCTACACCAT
CCCCTTAGCCATAAAGCTAAAGGAGCACCCCATCTTCTTCATGTTTATCCAGATCGCTGTCATCGCCAT

CTTTAAGTCCTACCCGACAGTGGGGGACGTGGCGCTCTACATGGCCTTCTTCCCCGTGTGGAACCATCT
CTACAGATTCCTGAGAAACATCTTTGTCCTCACCTGCATCATCATCGTCTGTTCCCTGCTCT

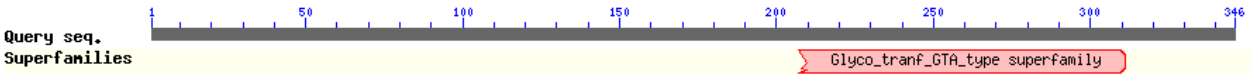
PIGU-ALG5:

TGCGTTACATCCCTTTGAAAGTGGCCCTGTTCTATCTCTTAAATCCTTACACGATTTTGTCTTGTGTTGC
CAAGTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATAAAAGGCAGTGC
TTTCCTCAGTGCTATTTTTCTTGCCTTAGCGACATACCAGTCTCTGTACCCACTCACCTTGTTTGTCCCA
GGACTCCTCTATCTCCTCCAGAATCAAATGGCTATAGCATGTGGATCTCGAGCTCATTTAGAAAAAGA
ATCAATTGCTCAGCGTTCTTACTTCCGTACTCTTCTCATGTATGGGTTCCACTTTCTGGTGTGGTTCCTT
TGTGTCAAAGGAATCAGGGACACACAGTGTGGGTTCAAATTATTTACTCGAGAAGCAGCTTCACGGAC
GTTTTCATCTCTACACGTTGAACGATGGGCATTTGATGTAGAACTACTGTACATAGCACAGTTCTTTAA
AATTCCAATAGCAGAAATTGCTGTCAACTGGACAGAAATTGAAGGTTCTAAATTAGTTCCATTCTGGA
GCTGGCTACAAATGGGTAAAGACCTACTTTTTATACGACTTCGATATTTGACTGGTGCCTGGAG

>PIGU wild type
MAAPLVLVLVVAVTVRAALFRSSLAEFISERVEVVSPLSSWKRVVEGLSLLDLGVSPYSGAVFH
ETPLIIYLFHFLIDYAELVFMITDALTAIALYFAIQDFNKVVFKKQKLLLELDQYAPDVAELIR
TPMEMRYIPLKVALFYLLNPYTILSCVAKSTCAINNTLIAFFILTTIKGSAFLSAIFLALATYQ
SLYPLTLFVPGLLYLLQRQYIPVKMKSKAFWIFSWEYAMMYVGSLVVIICLSFFLLSSWDFIPA
VYGFILSVPDLTPNIGLFWYFFAEMFEHFSLFFVCVFQINVFFYTIPLAIKLKEHPIFFMFIQI
AVIAIFKSYPTVGDVALYMAFFPVWNHLYRFLRNIFVLTCIIIVCSLLFPVLWHLWIYAGSANS
NFFYAITLTFNVGQILLISDYFYAFLRREYYLTHGLYLTAKDGTEAMLVLK

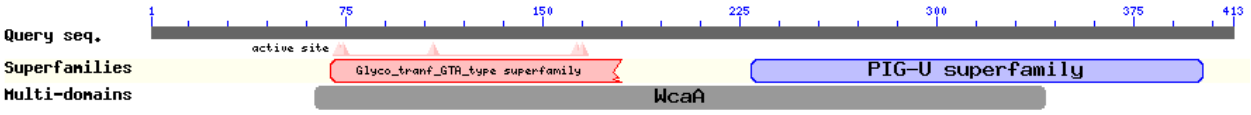
>ALG5 wild type
MAPLLLQLAVLGAALAAAALVLISIVAFTTATKMPALHRHEEEKFFLNAKGQKETLPSIWDSPT
KQLSVVVPSYNEEKRLPVMMDEALSYLEKRQKRDPAFTYEVIVVDDGSKDQTSKVAFKYCQKYG
SDKVRVITLVKNRGKGGAIRMGIFSSRGEKILMADADGATKFPDVEKLEKGLNDLQPWPNQMAI
ACGSRAHLEKESIAQRSYFRTLLMYGFHFLVWFLCVKGIRDTQCGFKLFTREAASRTFSSLHVE
RWAFDVELLYIAQFFKIPIAEIAVNWTEIEGSKLVPFWSWLQMGKDLLFIRLRYLTGAWRLEQT
RKMN

>PIGU-ALG5
MAAPLVLVLVVAVTVRAALFRSSLAEFISERVEVVSPLSSWKRVVEGLSLLDLGVSPYSGAVFH
ETPLIIYLFHFLIDYAELVFMITDALTAIALYFAIQDFNKVVFKKQKLLLELDQYAPDVAELIR
TPMEMRYIPLKVALFYLLNPYTILSCVAKSTCAINNTLIAFFILTTIKGSAFLSAIFLALATYQ
SLYPLTLFVPGLLYLLQNQMAIACGSRAHLEKESIAQRSYFRTLLMYGFHFLVWFLCVKGIRDT
QCGFKLFTREAASRTFSSLHVERWAFDVELLYIAQFFKIPIAEIAVNWTEIEGSKLVPFWSWLQ
MGKDLLFIRLRYLTGAWRLEQTRKMN



>ALG5-PIGU
MAPLLLQLAVLGAALAAAALVLISIVAFTTATKMPALHRHEEEKFFLNAKGQKETLPSIWDSPT
KQLSVVVPSYNEEKRLPVMMDEALSYLEKRQKRDPAFTYEVIVVDDGSKDQTSKVAFKYCQKYG

SDKVRVITLVKNRGKGGAIRMGIFSSRGEKILMADADGATKFPDVEKLEKGLNDLQPWPRQYIP
VKMKSKAFWIFSWEYAMMYVGSLVVIICLSFFLLSSWDFIPAVYGFILSVPDLTPNIGLFWYFF
AEMFEHFSLFFVCVFQINVFFYTIPLAIKLKEHPIFFMFIQIAVIAIFKSYPTVGDVALYMAFF
PVWNHLYRFLRNIFVLTCIIIVCSLLFPVLWHLWIYAGSANSNFFYAITLTFNVGQILLISDYF
YAFLRREYYLTHGLYLTAKDGTEAMLVLK



MIER2-RSRC2:

GACCATCAGTTCAACCTCGCAGAGATCCTGTCACAGAACTACAGTGTTAGGGGGGAGTGCGAGGAGG
CCTCGAGGTGCCCAGACAAGCCCAAGGAGGAGCTGGAGAAGGACTTCATCTCCCAGAGTGAAGATGA
AGCTGGATGTAGCTCAGTTGATGAAGAAAGTTACAAGACTCTGAAGCAGCAGGAAGAAGTATTTCGA
AATTTAGATGCTCAGTATGAAATGGCAAGATCACAAACCCACACACAAAGAGGAATGGGTTTGGGTTT
CACATCTTCAATGCGAGGAATGGATGCAGTTTGAAAATGATCACACTTGTAAAGTTTGGGACTTATAG
ACTTCTTGTTCTGATGTCACGTCCTTGTTCACCAAACAGCTAGCACTCTAGCTTGCATGGGTGTTGCATT
GACTTTAATTTATTGAAAAATACAAATTTTTGTAAATATCAGATCAGTGATACTGGTGTTAGTGTTGTA
ATCAGGTTAAACCCACTTCCATTAAACTTGACAGGACTATAGAAGGATAATATTTTTTAGTTCATGAAT
TCTACTTTTCAAATATATAAAAGCTGCAGGTGGGGATAAAATCTCATACATGGATTTTTTCGTGTCCGC
TGTCTTGTGTACTT

>MIER2 wild type
MAEASSLGRQSPRVVSCLEHSLCPGEPGLQTTAVVSMGSGDHQFNLAEILSQNYSVRGECEEAS
RCPDKPKEELEKDFISQSNDMPFDELLALYGYEASDPISDRESEGGDVAPNLPDMTLDKEQIAK
DLLSGEEEEETQSSADDLTPSVTSHEASDLFPNRSGSRFLADEDREPGSSASSDTEEDSLPANK
CKKEIMVGPQFQADLSNLHLNRHCEKIYENEDQLLWDPSVLPEREVEEFLYRAVKRRWHEMAGP
QLPEGEAVKDSEQALYELVKCNFNVEEALRRLRFNVKVIRDGLCAWSEEECRNFEHGFRVHGKN
FHLIQANKVRTRSVGECVEYYYLWKKSERYDYFAQQTRLGRRKYVPSGTTDADQDLDGSDPDGP
GRPRPEQDTLTGMRTDPLSVDGTAGGLDEPGVASDGLPSSEPGPCSFQQLDESPAVPLSHRPPA
LADPASYQPAVTAPEPDASPRLAVDFALPKELPLISSHVDLSGDPEETVAPAQVALSVTEFGLI
GIGDVNPFLAAHPTCPAPGLHSEPLSHCNVMTC

>RSRC2 wild type
MAASDTERDGLAPEKTSPDRDKKKEQSEVSVSPRASKHHYSRSRSRSRERKRKSDNEGRKHRSR
SRSKEGRRHESKDKSSKKHKSEEHNDKEHSSDKGRERLNSSENGEDRHKRKERKSSRGRSHSRS
RSRERRHRSRSRERKKSRSRSRERKKSRSRSRERKKSRSRSRERKRRIRSRSRSRSRHRHRTRS
RSRTRSRSRDRKKRIEKPRRFSRSLSRTPSPPPFRGRNTAMDAQEALARRLERAKKLQEQREKE
MVEKQKQQEIAAAAATGGSVLNVAALLASGTQVTPQIAMAAQMAALQAKALAETGIAVPSYYNP
AAVNPMKFAEQEKKRKMLWQGKKEGDKSQSAEIWEKLNFGNKDQNVKFRKLMGIKSEDEAGCSS
VDEESYKTLKQQEEVFRNLDAQYEMARSQTHTQRGMGLGFTSSMRGMDAV

>MIER2-RSRC2
MAEASSLGRQSPRVVSCLEHSLCPGEPGLQTTAVVSMGSGDHQFNLAEILSQNYSVRGECEEAS
RCPDKPKEELEKDFISQSEDEAGCSSVDEESYKTLKQQEEVFRNLDAQYEMARSQTHTQRGMGL
GFTSSMRGMDAV

No putative conserved domains have been detected

References

Sboner A, Habegger L, Pflueger D, Terry S, Chen D, Rozowsky J, Tewari AK, Kitabayashi N, Moss BJ, Chee MS et al. IN PRESS. FusionSeq: a modular framework for finding gene fusions by analyzing Paired-End RNA-Sequencing data *Genome Biol*.

Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. 2008. Recurrent gene fusions in prostate cancer. *Nature reviews* **8**(7): 497-511.

Makkonen H, Kauhanen M, Paakinaho V, Jaaskelainen T, Palvimo JJ. 2009. Long-range activation of FKBP51 transcription by the androgen receptor via distal intronic enhancers. *Nucleic Acids Res* **37**(12): 4135-4148.

Klezovitch O, Risk M, Coleman I, Lucas JM, Null M, True LD, Nelson PS, Vasioukhin V. 2008. A causal role for ERG in neoplastic transformation of prostate epithelium. *Proc Natl Acad Sci U S A* **105**(6): 2105-2110.

Tomlins SA, Laxman B, Varambally S, Cao X, Yu J, Helgeson BE, Cao Q, Prensner JR, Rubin MA, Shah RB et al. 2008. Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* **10**(2): 177-188.

Supplemental Tables

Supplemental Table S1. Inter- and intra-chromosomal gene fusions in human prostate cancer and prostate cancer cell lines discovered by previous RNA-Sequencing studies and validated with FISH and RT-PCR. Intra-chromosomal gene fusions are colored in blue and inter-chromosomal gene fusions are colored in red.

Sample Type	**Previously reported ETS fusions**	**Validated Novel Fusion**	**Study**
VCaP PCA cell line	*TMPRSS2-ERG* (1)	*TMPRSS2-ERG* (1); *USP10-ZDHHC7*; *HJURP-INPP4A*; *EIF4E2-HJURP*; *ZDHHC7-ABCB9*; *TIA1-DIRC2*	(2, 3)
LNCaP PCA cell line	*MIPOL1-ETV1*	*MIPOL1-DGKB*	(3)
RWPE Benign prostate cell line	Negative	-	(3)
VCaP-Met	*TMPRSS2-ERG* (1)	*RC3H2-RGS3*; *LMAN2-AP3S1*	(3)
Met-3	Negative	*STRN4-GPSN2*	(3)
Localized PCA aT64	Negative	*HERPUD1-ERG*	(2)
Localized PCA aT52	Negative	*AX747630-ETV1*	(2)
Localized PCA (N=3)	Negative	*NDRG1-ERG*	(4)
Locally advanced PCA (PCA3)	Negative	*SLC45A3-BRAF*	(5)
Locally advanced PCA (PCA17)	Negative	*ESRP1-RAF1*	(5)

1. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005; 310:644-8.
2. Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, Lonigro RJ, Schroth G, Kumar-Sinha C, Chinnaiyan AM. Chimeric transcript discovery by paired-end transcriptome sequencing. Proc Natl Acad Sci U S A 2009; 106:12353-8.
3. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM. Transcriptome sequencing to detect gene fusions in cancer. Nature 2009; 458:97-101.
4. Pflueger D, Rickman DS, Sboner A, Perner S, LaFargue CJ, Svensson MA, Moss BJ, Kitabayashi N, Pan Y, de la Taille A, Kuefer R, Tewari AK, Demichelis F, Chee MS, Gerstein

MB, Rubin MA. N-myc downstream regulated gene 1 (NDRG1) is fused to ERG in prostate cancer. Neoplasia 2009; 11:804-11.
5. Nallasivam Palanisamy, Bushra Ateeq, Shanker Kalyana-Sundaram, Dorothee Pflueger, Kalpana Ramnarayanan, Sunita Shankar, Bo Han, Qi Cao, Xuhong Cao, Khalid Suleman, Chandan Kumar-Sinha, Saravana M Dhanasekaran, Ying-bei Chen, Raquel Esgueva, Samprit Banerjee, Christopher J LaFargue, Javed Siddiqui, Francesca Demichelis, Peter Moeller, Tarek A Bismar, Rainer Kuefer, Douglas R Fullen, Timothy M Johnson, Joel K Greenson, Thomas J Giordano, Patrick Tan, Scott A Tomlins, Sooryanarayana Varambally, Mark A Rubin, Christopher A Maher, Arul M Chinnaiyan. Rearrangements of the *RAF* kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med 2010 Jul;16:793-8.

Supplemental Table S2. Fraction of candidates removed by FusionSeq filtering system

				# of candidates removed by filters:									
				Mis-alignment filters			Random pairing filter	Mis-alignment and random pairing filters		Others			
Sample ID	# of original candidates	# of remaining candidates after filtering	percent removed	Large scale homology	Small scale homology	Repeat masker	Artifactual insert size	Ribosomal	Expression consistency	PCR	Annotation consistency	Proximity	Black list
STID3023_B62	22586	1069	95.3%	41	34	7584	104	34	3879	9453	88	277	23
STID2620_D	8455	208	97.5%	43	65	3717	79	15	408	3601	91	211	17
STID2858_C	3006	154	94.9%	38	397	640	126	24	652	778	60	121	16
STID580_Ba	5791	170	97.1%	20	38	3144	70	38	783	1308	58	145	17
STID581_Dt	13683	645	95.3%	31	30	3185	24	8	1734	7637	186	190	13

The filtering system of FusionSeq removes on average 96% artefactual candidates of the original list in the 5 samples featuring novel gene fusions.

Supplemental Table S3. Validation of chimeric candidates in the read-through and cis category.

Type	Candidate				Experimental validation	Candidate detected in n number of cancer tissue samples by FusionSeq	Candidate detected in benign prostate tissue by FusionSeq
	Gene A		Gene B		RT-PCR		
read-through	*ZNF649* *	19q13.3	*ZNF577* *	19q13.3	pos	22	yes
	SMG5	1q22	*PAQR6*	1q22	pos	13	yes
	VMAC *	19p13.3	*CAPS* *	19p13.3	pos	15	yes
	AK311452 *	19q12	*AK094188* *	19q12	pos	22	yes
	ZNF772	19q13.4	*VN1R1*	19q13.4	pos	11	no
	ST6GALNAC6	9q34.1	*Ak1*	9q34.1	pos	7	no
	HARS2	5q31.3	*ZMAT2*	5q31.3	pos	10	yes
	SLC45A3 *	1q32.1	*ELK4* *	1q32.1	pos	14	yes
	AZGP1	7q22.1	*GJC3*	7q22.1	pos	15	yes
	TYMP	22q13.3	*SCO2*	22q13.3	neg	16	yes
	ABCB8 *	7q36.1	*ACCN3* *	7q36.1	neg	8	no
cis	*PICK1*	22q13.1	*SLC16A8*	22q13.1	neg	13	yes
	BX647715	6p21.1	*SUPT3H*	6p21.1	neg	8	yes

$n_{total}=25$

Eleven read-through candidates and two cis candidates were chosen for validation by the RT-PCR technique. The table shows if experimental validation was successful along with what the overall frequency of the respective candidate was in prostate cancer tissues and if it was nomated in adjacent benign prostate tissue. *Candidates marked have been previously reported elsewhere.

Supplemental Table S4. Clinical characteristics of patients represented on tissue micro arrays (TMA)

		Prostate Cancer TMA (n = 88)
Age [years], median (range)		61 (42-75)
Gleason score	6	22 (25%)
	7	59 (67%)
	8-9	7 (8%)
Tumor stage	pT2a-c	67 (76%)
	pT3a	16 (18%)
	pT3b	3 (3%)
	pT4	2 (2%)
Lymph node status	N1	1 (1%)
	N0	87 (99%)
Preoperative PSA [ng/ml], median (range)		5.1 (1.1-15.7)
PSA biochemical recurrence	Yes	6 (7%)
	No	82 (93%)
Surgical margine	Yes	10 (11%)
	No	78 (89%)
Vascular invasion	Yes	6 (7%)
	No	82 (93%)

Demographics of 88 patients suffering from localized prostate cancer. This cohort was used for screening rearrangements in the newly discovered fusion genes.

Supplemental Table S5. Clinical characteristics of patients used in RNA-Seq

Sample	Age [years]	Gleason score	Tumor stage	Lymph node status	Preoperative PSA [ng/ml]	PSA biochemical recurrence	Surgical margin	Vascular invasion
STID410_C	60	7	pT2c	N0	4.52	no	no	no
STID420_D	51	7	pT2c	N0	13.7	yes	no	no
STID580_B	56	7	pT2c	N0	3	no	no	no
STID581_D	69	9	pT3b	N0	9.2	yes	no	no
STID1024_D	73	9	pT3b	N0	7.2	no	no	no
STID1043_B	59	6	pT2c	N0	4.9	no	no	no
STID1700_D	57	8	pT3a	N0	2.94	yes	yes	no
STID1783_B	66	8	pT2c	N0	9.8	no	no	no
STID2525_A	59	7	pT2c	N0	5.56	yes	yes	no
STID2620_D	69	6	pT2c	N0	9	no	no	no
STID2660_B	59	7	pT3a	N0	5	no	no	no
STID2661_D	59	7	pT2c	N0	4.52	no	no	no
STID2740_A	58	7	pT2c	N0	2.8	no	no	no
STID2743_D	65	7	pT2c	N0	8.1	no	no	no
STID2762_D	68	7	pT2c	N0	5.5	N/A	N/A	N/A
STID2849_D	52	7	pT2c	N0	4	no	no	no
STID2858_C	52	7	pT3a	N0	5.3	no	yes	yes
STID2872_D	62	7	pT3a	N0	6	N/A	yes	yes
STID3023_B62	67	8	pT2c	N0	5	no	no	no
STID3027_B57	66	8	pT3b	N0	10.2	no	yes	no
STID3042_H51	67	7	pT2c	N0	3.3	no	no	no
STID3043_B56	70	7	pT2c	N0	7.2	no	no	no
STID3071_B51	55	7	pT3a	N0	3	no	no	no
STID3127_B56	55	7	pT2c	N0	4.2	no	no	no
STID3134_B58	64	7	pT3a	N0	5.5	no	no	no

Demographics of 25 patients whose prostate cancer has been subjected to RNA-Seq. (N/A; missing data point)

Supplemental Table S6. Primer sequences.

Target	sense (5' --> 3')	antisense (5' --> 3')
RNA chimera:		
KLK2-ETV1 ex9	GTTCTCTCCATCGCCTTGTCT	CACTGGGTCGTGGTACTCCT
KLK2-ETV1 ex12	GTTCTCTCCATCGCCTTGTCT	TCCTTCCCGATACATTCCTG
FKBP5-ERG	TGGCATTCACTGTGGGCGAAGG	GTAGGCACACTCAAACAACGACTGG
TMPRSS2-FKBP5	TAGGCGCGAGCTAAGCAGGAG	CGCCCACAGTGAATGCCACATC
FKBP5 ex3a-TMPRSS2 ex6	ACGGGCAAGGCGGAGAGTG	GACACGCCATCACACCAGTTAGAG
FKBP5 ex6-TMPRSS2 ex8	TGGCATTCACTGTGGGCGAAGG	TCCGCTGTCATCCACTATTCCTTG
TMPRSS2-(FKBP5-)ERG	TAGGCGCGAGCTAAGCAGGAG	GTAGGCACACTCAAACAACGACTGG
CDKN1A-CD9	CTGCCGAAGTCAGTTCCTTGT	GGCGAATATCACCAAGAGGA
TNPO1-IKBKB	CAGACACCACCATCCAGAGAA	ATCTGCTCACCTGTTTCCTGAT
ALG5-PIGU 1	GAGATCCTGCGTTCACTTATGAAGT	TGTCGGGTAGGACTTAAAGATGGC
ALG5-PIGU 2	CCCTGGTGAAGAATCGTGGAAAAG	AGAGCAGGGAACAGACGATGATG
PIGU-ALG5 1	TGCGTTACATCCCTTTGAAAGTGG	AAATGCCCATCGTTCAACGTGTAG
PIGU-ALG5 2	TGCCAAGTCTACCTGTGCCATC	CTCCAGGCACCAGTCAAATATCG
MIER2-RSRC2	GACCATCAGTTCAACCTCGCAGAG	AAGTACACAAGACAGCGGACACG
ZNF646-ZNF577	CACACGGGAGAGAGACCCTA	CCTTCCCAGAAGTGGTGGT
SMG5-PAQR6	GTCTCCCTCTGTCTAGCTTCCA	TGCACAGGAAGGAGTTGAGTG
VMAC-CAPS	GAGGCAAACGCACACCTG	GTCCCACTTCCTGCACACA
AK311452-AK094188	GATTTCTCGCAGACCAGGAGGAC	AGTGTTTCCAGCCTGAGACTTCTG
ZNF772-VN1R1	TTCCTCTACCATCCTCGGGTGGAC	GGGCTGCGTGCTGGGAAACA
ST6GALNAC6-Ak1	CTCAGTGGGGTGTGTTTGTCT	CTTCTCCATGATTTCCGACAG
HARS2-ZMAT2	TCCTTGCCTGATTCCCATC	CTGCACTGGTTTTCCATCTTTC
SLC45A3-ELK4	CGCTGGCTCCGGGTGACA	TGCCCATCATTAGAGGTCCAACAG
AZGP1-GJC3	GGTGGAAGGAATGGAGGATT	CTGGGTACAGCCACCAAGAT
PICK1-SLC16A8:		
ab	CTCGTGTCCACCATGTCCAAGTAC	AAGAAGCTGTGTTCCCAAGTCACTG
ac	CTCGTGTCCACCATGTCCAAGTAC	GCGAGCCCACAGAACCAGAAATAG
de	GAGCACAGGTGGTACAGGTTCATG	CGCTCGGGTGTCTTGGTCTTTATC
af	CTCGTGTCCACCATGTCCAAGTAC	GAGGTTTCACCGTGTTAGCCAGG
ge	GGACCGATTGTGCCACTTTGGAG	CGCTCGGGTGTCTTGGTCTTTATC
qPCR:		
ALG5	GACTCACCTACCAAACAACTTT	GCTTCATCCATCATCACAGG
PIGU	CCCTATGGAAATGCGTTACA	GTTGATGGCACAGGTAGAC
TNPO1	GAGGCGTCTGGGATGGTGTG	GGTTCTCTGGATGGTGGTGTC
CDKN1A	CTGCCGAAGTCAGTTCCTTGT	CATGGGTTCTGACGGACATC
PSA	GAGCACCCCTATCAACCCCCTATT	AGCAACCCTGGACCTCACACCTAA
TBP	TGCCTCCAGAATATGCCTCT	CAATGGTTTTCAAGCTTTCCA
CD9	GACTGTTCTTCGGCTTCCTCTTG	ACTCCTGGACTTCCTTAATCACCTC
ALG5-PIGU	GCAGATGCTGATGGAGCCACAAAG	GCCATAGACTGCGGGGATGAAATC
PIGU-ALG5	ACCTGTGCCATCAACAACACC	CGGAAGTAAGAACGCTGAGCA

Supplemental Table S7. BAC probes for FISH.

Assay	BACs	
PIGU break-apart	5' RP11-45A2	3' RP11-318N1
ALG5 break-apart	5' RP11-270O5	3' RP11-297M16
CDKN1A break-apart	5' RP11-110F15	3' RP11-52A1
CD9 break-apart	5' CTD-2002E14	3' CTD-2079K14
CDKN1A-CD9 fusion	RP11-110F15	CTD-2079K14
TNPO1 break-apart	5' RP11-746O5	3' RP11-354N11
IKBKB break-apart	5' RP11-589C21	3' RP11-213C20
TNPO1-IKBKB fusion	RP11-351B18	RP11-357H6
KLK locus break-apart	5' RP11-933E3	3' RP11-26P14
ETV1 break-apart	5' RP11-661L15	3' RP11-79G16
TMPRSS2 break-apart	5' RP11-891L10	3' RP11-354C5
FKBP5 break-apart	5' CTD-2288M12	3' RP11-22L3
FKBP5-TMPRSS2 fusion	CTD-2288M12	RP11-120C17
ERG break-apart	5' RP11-372O17	3' RP11-24A11

Figure Legends

Supplemental Fig. S1. Schematic of the study design. Prostate cancers were characterized for *ERG* rearrangement by FISH and *ERG* fusion status by RT-PCR. Samples were subjected to PE RNA-seq and resulting data was processed through FusionSeq. FusionSeq is a compendium of modules/filters that map and filter PE reads. From this, chimeric candidates are nominated and further classified. The Junction Sequence Identifier is a module that aims to identify the mRNA breakpoints and single reads that map specifically to this breakpoint to further support the candidate. In a subsequent step, scores are attributed in order to nominate and sort chimeric candidates depending on their likelihood of being real and in order to prioritize experimental validation.

Supplemental Fig. S2. Uncovering the chimeric cis candidate *PICK1-SLC16A8* as erroneous. **(A)** Schematic of a possible *PICK1-SLC16A8* gene fusion and primers used to evaluate a possible fusion transcript. (**B**) schematic of actual structure of the genomic region encompassing *PICK1* and *SLC16A8* with primers indicated that were used to unravel the transcriptionally active gene. (**C**) Five different combinations of primers were used. PCR products are only visible with primer combinations that indicate extended *PICK1* expression. (**D**) Sanger sequencing of the PCR products reveals that the *PICK1* gene locus has at least three additional 3' exons.

Supplemental Fig. S3. Unravelling a complex fusion event involving *TMPRSS2, ERG* and *FKBP5***.** (**A**) The alignment of the PE-reads to *TMPRSS2, FKBP5* and *ERG* indicates a complex fusion event. *FKBP5* has connecting reads to both *TMPRSS2* and *ERG* (red). (**B**) A fusion specific FISH assay also confirmed the presence of a FKBP5-TMPRSS2 gene fusion allele in

cancer (upper left) but not in adjacent benign tissue (upper right). Below, FISH experiments also confirm the rearrangement of all three involved genes (*TMPRSS2*, *FKBP5* and *ERG*) in this cancer focus. (**C**) RT-PCR and subsequent Sanger sequencing confirms the presence of a *TMPRSS2-FKBP5-ERG* triple fusion resulting in various fusion transcript isoforms (1-4). All resulting connections were confirmed by additional RT-PCR assays. The reciprocal balanced translocation between *FKBP5* and *TMPRSS2* is transcriptionally active as shown by a specific *FKBP5-TMPRSS2* fusion transcript. (**D**) Schematic figure of the complex triple fusion and reciprocal balanced translocation involving *TMPRSS2*, *FKBP5* and *ERG*.

Supplemental Fig. S4. (**A**) Expression of truncated CD9 protein in HEK293 cells. Western blot shows strong expression of FLAG-tagged CD9 in WT-*CD9* transfected cells but not from untransfected cell or cells transfected with *LacZ* control expression vector. A lower and less intense band is detected from *CDKN1A-CD9* fusion transfected cells likely representing expression of a truncated version of CD9 originating from the fusion. (**B**) Expression of truncated CD9 protein in benign prostate epithelial cell line RWPE shows absence of membranous staining and reduced expression when compared to WT-CD9 transfected RWPE.

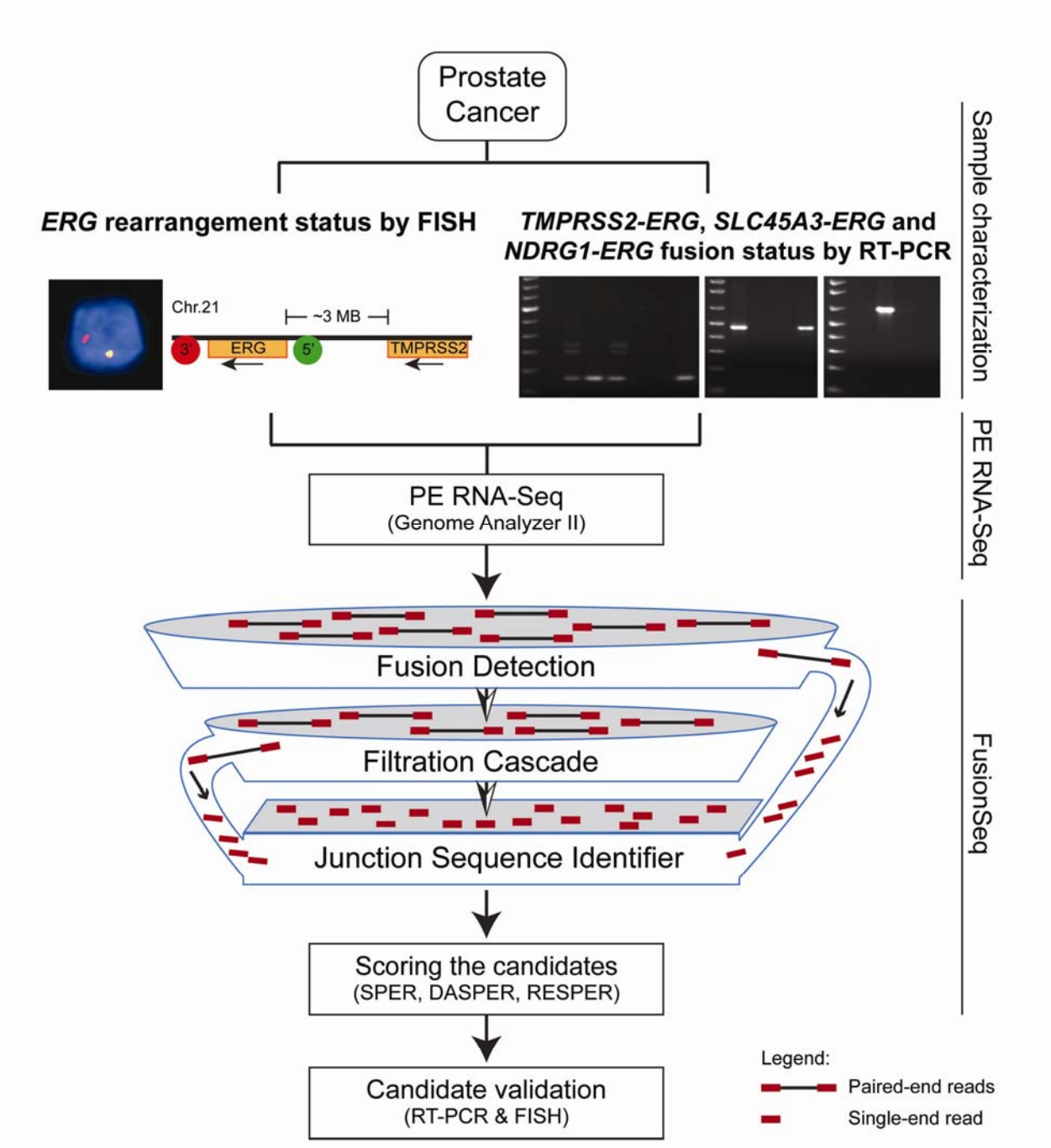
Supplemental Fig. S5: (**A**) Immunofluorescent detection of ALG5-PIGU protein (green). Nuclei were stained with DAPI (blue). Objective magnification for immunofluorescence images, x63. (**B**) *PIGU* expression levels across a set of prostate cancers with the *PIGU* translocation positive sample indicated in black. (**C**) Effect of siRNA-mediated dowregulation of *PIGU* expression on proliferation of prostate epithelial cells and prostate cancer cells. Cells were transfected with siRNAs against *PIGU (gray columns),* MYC *(Black columns)* or control non targeting *(White*

columns) siRNAs. At 72h, cell growth was assessed by means of a WST-1 assay and compared to J0. *Columns*, means of five individual measurements; *bars*, SE. MYC a potent oncogene was used as a positive control. *PIGU* silencing using *PIGU* siRNAs reduce the proliferation of LNCaP, 22Rv1 and RWPE cells but has a lower effect compared to the silencing of *MYC*. (**D**) FISH break apart assays of the *PIGU* and *ALG5* genes in benign tissue. Pictures were taken from benign epithelial cells adjacent to the cancer in the patient with *ALG5-PIGU* translocation.

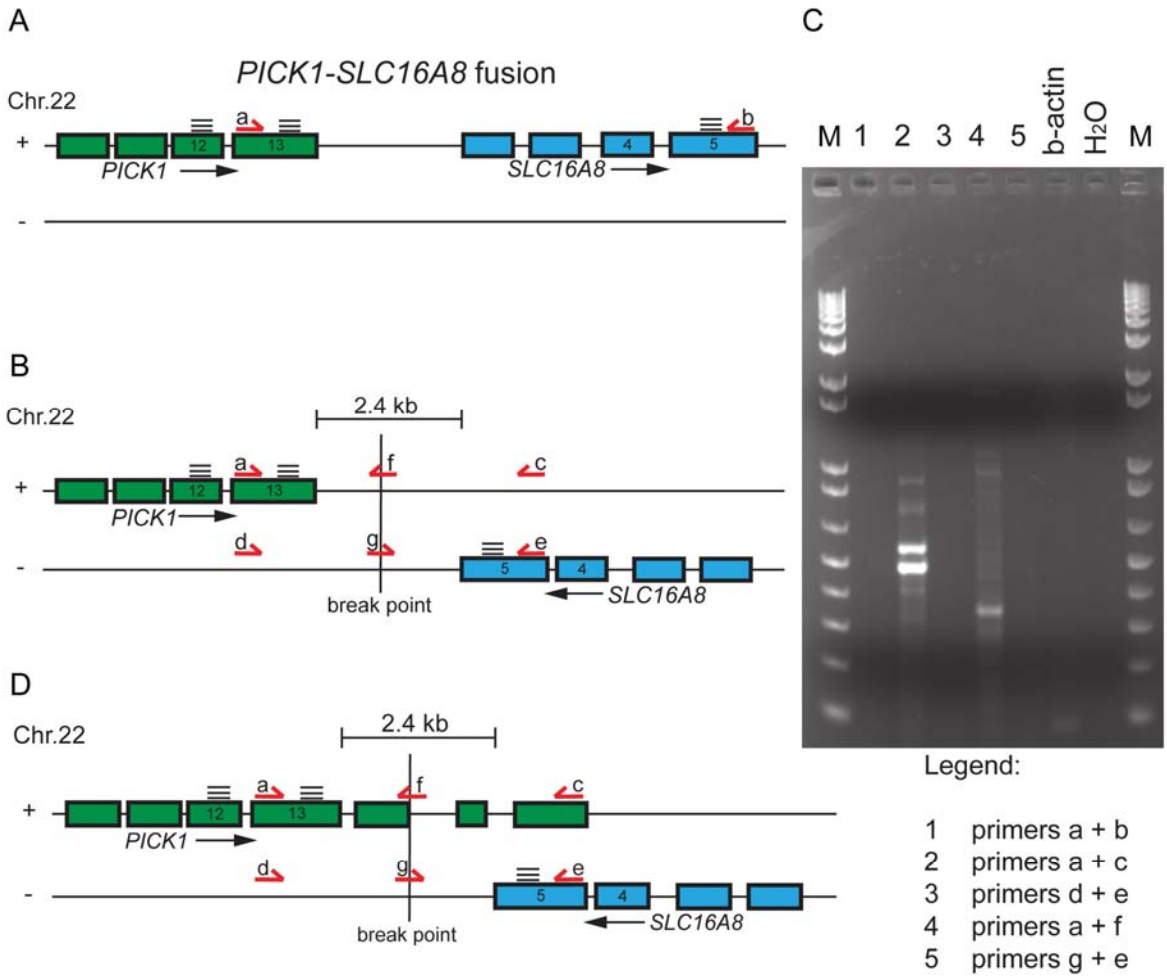
Supplemental Fig. S6. Androgen regulation of novel gene fusion 5' partners. Androgen-sensitive LNCaP and VCaP cells were exposed to the synthetic androgen R1881 for 24h and expression levels of *PIGU, ALG5, CDKN1A, TNPO1* as well as the androgen-induced gene *KLK3* were assessed using RT-PCR. No apparent changes are observed for *PIGU, ALG5, CDKN1A* and *TNPO1* in contrast to that found for *KLK3* expression suggesting that these genes may not be directly regulated by androgens.

Supplemental Fig. S7. Characterization of the *MIER2-RSRC2* gene fusion candidate. (**A**) Experimental validation of the gene fusion transcript by RT-PCR and subsequent Sanger sequencing. (**B**) Predicted open reading frame (ORF) of the *MIER2-RSRC2* fusion (upper sequence). For comparison, primary sequences of wilt type *MIER2* and *RSRC2*, are given below. (**C**) *MIER2* and *RSRC2* expression levels in a set of 25 prostate cancers including the fusion positive sample (black).

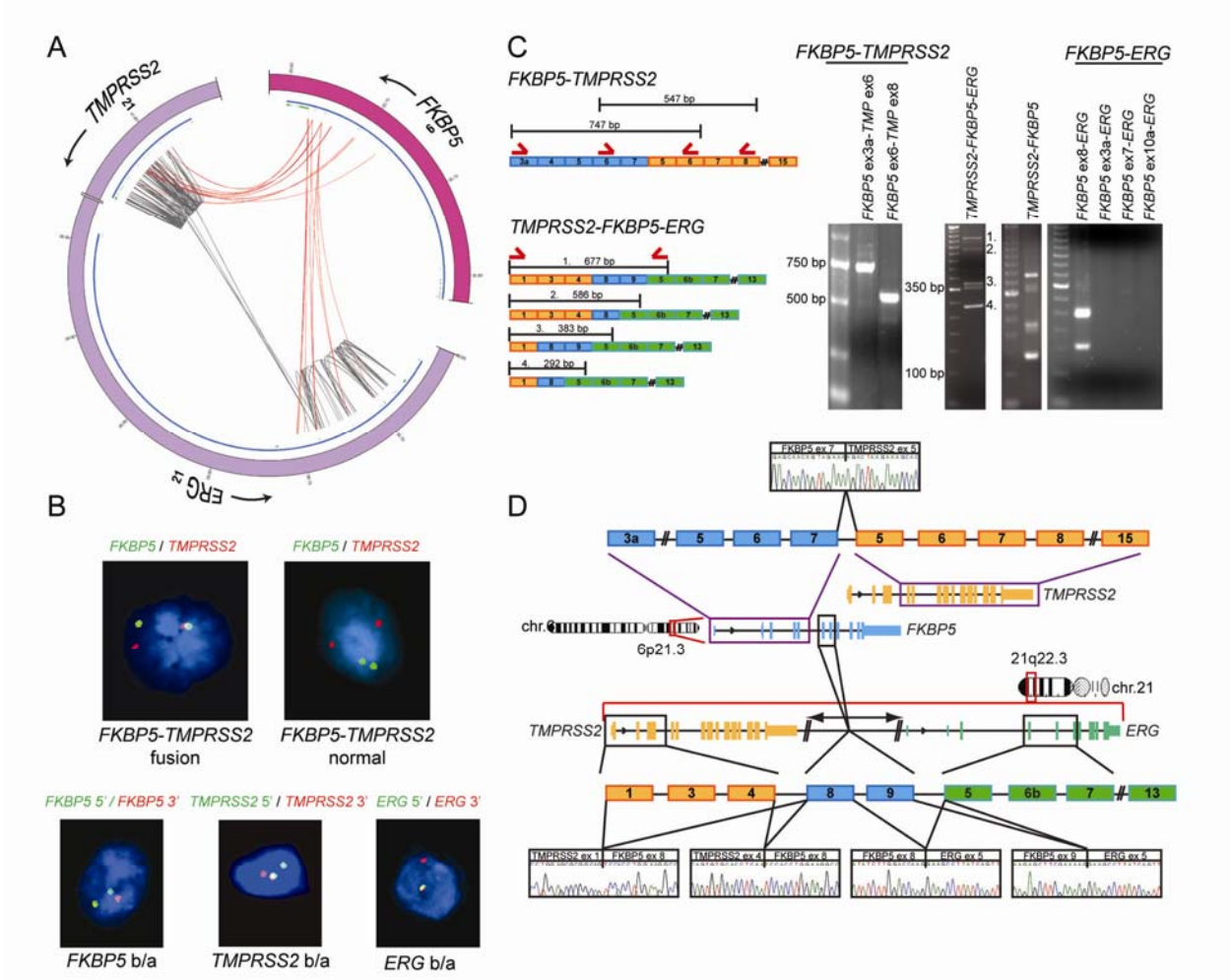
Supplemental Figure S8. Characterization of the *RYBP-FOXP1* gene fusion candidate in NCI-H660 cells. (**A**) Histogram showing candidates nominated by FusionSeq from RNA-Seq data including *TMPRSS2-ERG* and *RYBP-FOXP1*. (**B**) Experimental validation of the *RYBP-FOXP1* fusion by RT-PCR, subsequent Sanger sequencing, and schematic prediction of the intra-chromosomal rearrangement leading to *RYBP-FOXP1*. (**C**) Corresponding mRNA sequence of the PCR amplicon shown in (B).



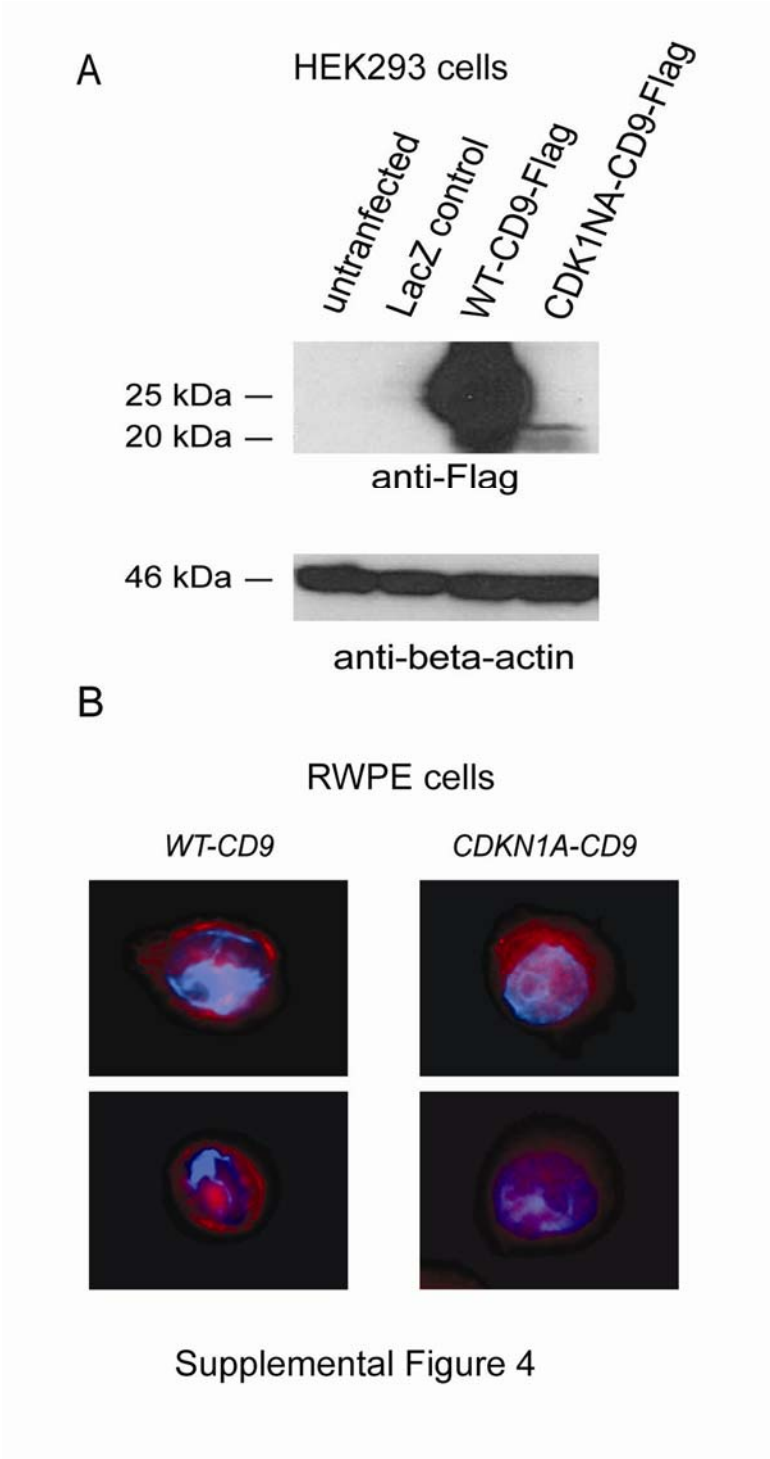
Supplementary Figure 1



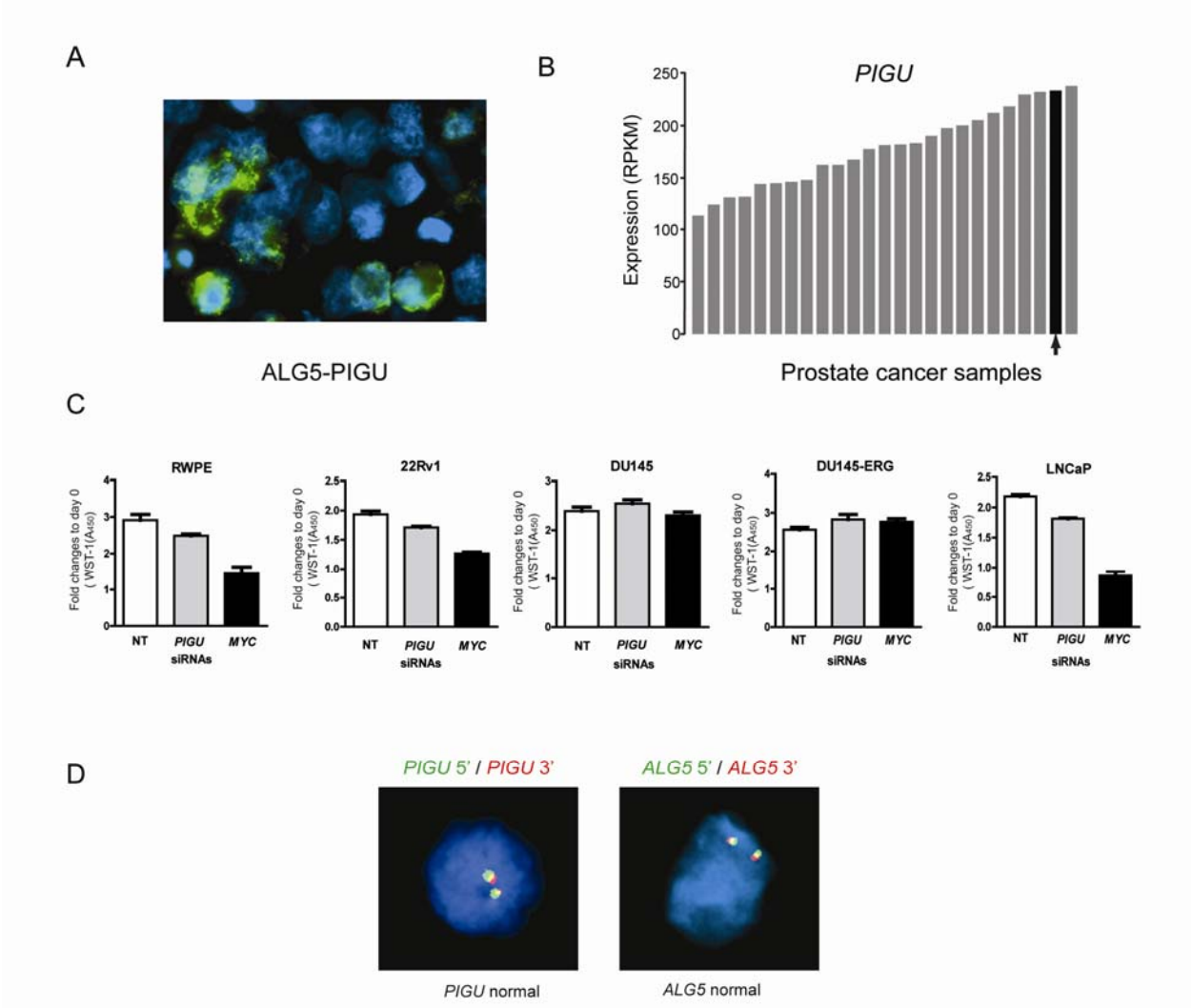
Supplemental Figure 2



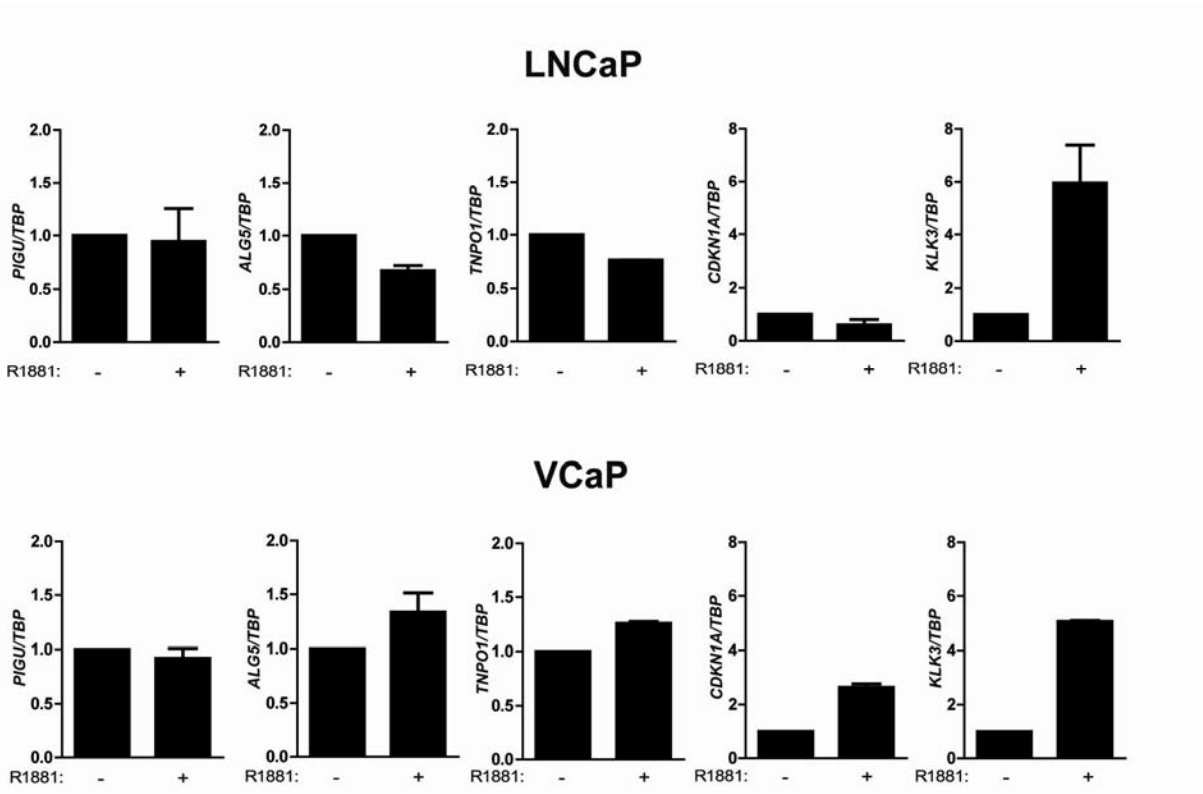
Supplemental Figure 3



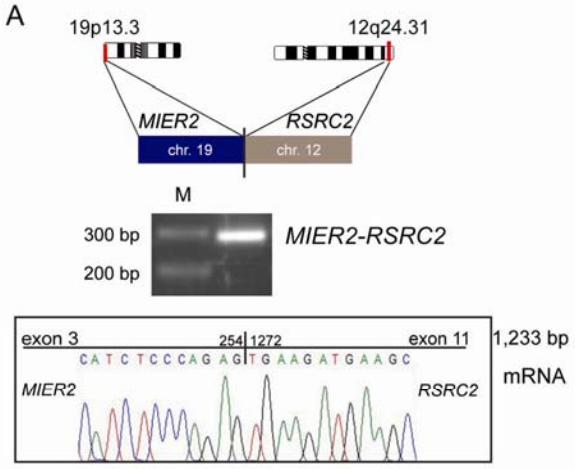
Supplemental Figure 4

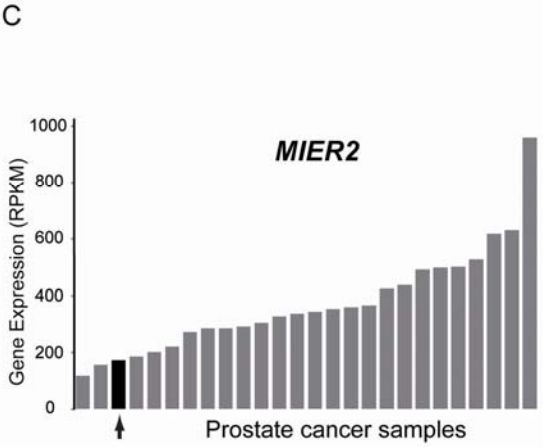


Supplemental Figure 5



Supplementary Figure 6



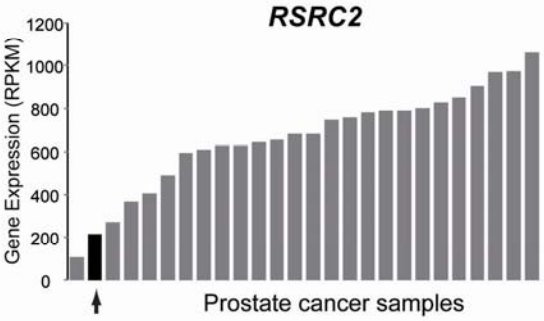


B

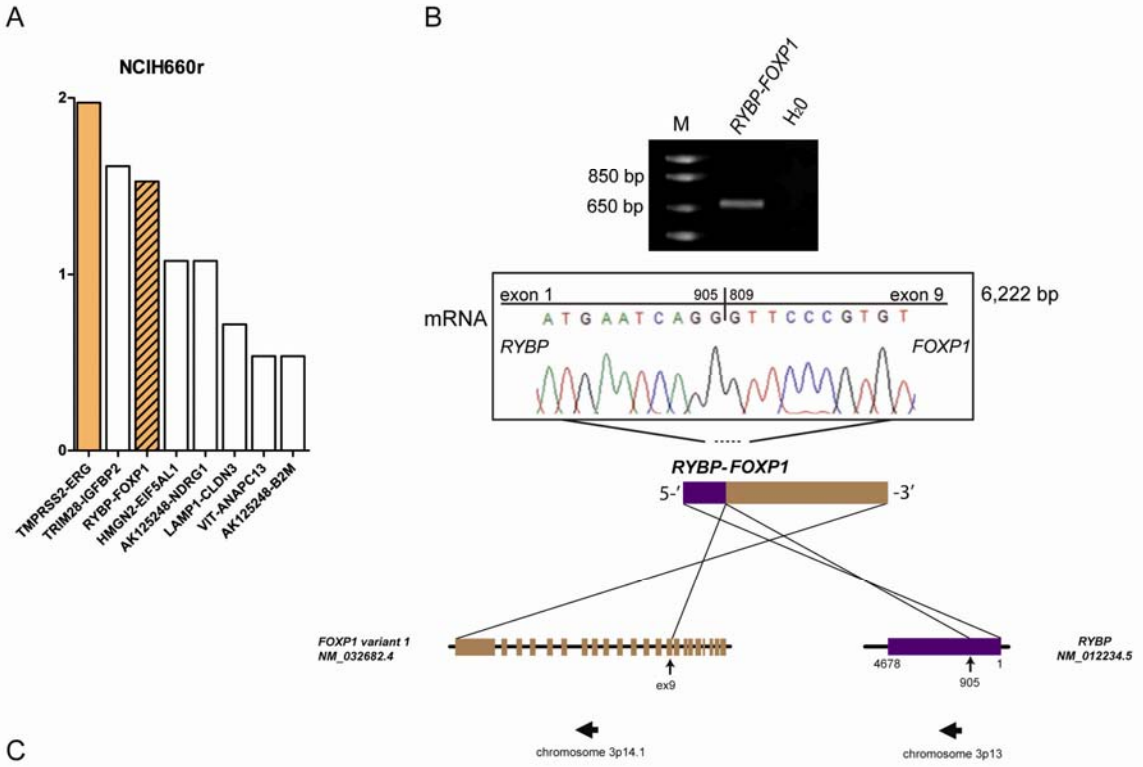
>MIER2-RSRC2
MAEASSLGRQSPRVVSCLEHSLCPGEPGLQTTAVVSMGSGDHQFNLAEILSQNYSVRGEC
EEASRCPDKPKEELEKDFISQSEDEAGCSSVDEESYKTLKQQEEVFRNLDAQYEMARSQT
HTQRGMGLGFTSSMRGMDAVStop

>MIER2
MAEASSLGRQSPRVVSCLEHSLCPGEPGLQTTAVVSMGSGDHQFNLAEILSQNYSVRGEC
EEASRCPDKPKEELEKDFISQSNDMPFDELLALYGYEASDPISDRESEGGDVAPNLPDMT
LDKEQIAKDLLSGEEEEETQSSADDLTPSVTSHEASDLFPNRSGSRFLADEDREPGSSAS
SDTEEDSLPANKCKKEIMVGPQFQADLSNLHLNRHCEKIYENEDQLLWDPSVLPEREVEE
FLYRAVKRRWHEMAGPQLPEGEAVKDSEQALYELVKCNFNVEEALRRLRFNVKVIRDGLC
AWSEEECRNFEHGFRVHGKNFHLIQANKVRTRSVGECVEYYYLWKKSERYDYFAQQTRLG
RRKYVPSGTTDADQDLDGSDPDGPGRPRPEQDTLTGMRTDPLSVDGTAGGLDEPGVASDG
LPSSEPGPCSFQQLDESPAVPLSHRPPALADPASYQPAVTAPEPDASPRLAVDFALPKEL
PLISSHVDLSGDPEETVAPAQVALSVTEFGLIGIGDVNPFLAAHPTCPAPGLHSEPLSHC
NVMTCStop

>RSRC2
MAASDTERDGLAPEKTSPDRDKKKEQSEVSVSPRASKHHYSRSRSRSRERKRKSDNEGRK
HRSRSRSKEGRRHESKDKSSKKHKSEEHNDKEHSSDKGRERLNSSENGEDRHKRKERKSS
RGRSHSRSRSRERRHRSRSRERKKSRSRSRERKKSRSRSRERKKSRSRSRERKRRIRSRS
RSRSRHRHRTRSRSRTRSRSRDRKKRIEKPRRFSRSLSRTPSPPPFRGRNTAMDAQEALA
RRLERAKKLQEQREKEMVEKQKQQEIAAAAATGGSVLNVAALLASGTQVTPQIAMAAQMA
ALQAKALAETGIAVPSYYNPAAVNPMKFAEQEKKRKMLWQGKKEGDKSQSAEIWEKLNFG
NKDQNVKFRKLMGIKSEDEAGCSSVDEESYKTLKQQEEVFRNLDAQYEMARSQTHTQRGM
GLGFTSSMRGMDAVStop



Supplementary Figure 7



C

RYBP-FOXP1 mRNA:

CGGATCAATTCTCAGCTGGTGGCACAACAAGTGGCACAACAGTATGCCACCCCACCACCCCCTAAAAAGGAGAAGAAGGAGAAAGTTGAAAAGC
AGGACAAAGAGAAACCTGAGAAAGACAAGGAAATTAGTCCTAGTGTTACCAAGAAAAATACCAACAAGAAAACCAAGCCCCGGCTGAAAAACGT
GGACAGGAGCACTGCACAGCAGTTGGCAGTAACTGTGGGCAACGTCACCGTCATTATCACAGACTTTAAGGAAAAGACTCGCTCCTCATCGACATC
CTCATCCACAGTGACCTCCAGTGCAGGGTCAGAACAGCAGAACCAGAGCAGCTCGGGGTCAGAGAGCACAGACAAGGGCTCCTCCCGTTCCTCCA
CGCCAAAGGGCGACATGTCAGCAGTCAATGATGAATCTTTCTGAAATTGCACATGGAATTGTGAAAACTATGAATCAGGGTTCCCGTGTCAGTGGCTA
TGATGACACCTCAAGTTATCACTCCCCAGCAAATGCAGCAGATCCTCCAGCAACAAGTGCTGAGCCCTCAGCAGCTCCAGGTTCTCCTCCAGCAGCAG
CAGGCCCTCATGCTTCAACAGCAGCAGCTTCAAGAGTTTTATAAAAAACAACAGGAACAGTTGCAGCTTCAACTTTTACAACAACAACATGCTGGAAA
ACAGCCTAAAGA

Supplemental Figure 8