Supplementary Data

Supplementary Table S1. Independent Validation of Molecular Markers by Immunohistochemistry and Real-Time RT-PCR Analyses

Gene name	Gene symbol[a]	Tumor		FNA
		QRT-PCR[b] (n = *107*)	*IHC*[c] (n = *154*)	*QRT-PCR*[c] (n = *95*)
Overexpressed in malignant tumors				
High mobility group AT-hook 2	*HMGA2*	✓	✓	✓
Kallikrein 7 (chymotryptic, stratum corneum), transcript variant 1	*KLK7*	–	✓	–
Mannose receptor C2	*MRC2/Endo 180*	–	✓	✓
Leucine-rich repeat kinase 2	*LRRK2*	✓	–	–
Pleomorphic adenoma gene 1	*PLAG1*	✓	✓	✓
Cytochrome P450, family 1, subfamily B, polypeptide 1	*CYP1B1*	–	✓	–
Dipeptidyl-peptidase 4	*DPP4/CD26*	✓	✓	✓
Ffibronectin type III	*FNDC4*	–	–	✓
Cadherin 3 type1, type 1, P-cadherin (placental)	*CDH3/P-Cadherin*	✓	✓	✓
Carcinoembryonic antigen-related cell adhesion molecule 6	*CEACAM6*	✓	–	✓
Protease serine 3 (PRSS3/mesotrypsin)	*PRSS3*	✓	–	✓
Phosphodiesterase 5 A, cGMP-specific, transcript variant 3	*PDE5A*	✓	–	–
Sparc/osteonectin CWCV, and kazal-like domain proteoglycan	*SPOCK1*	✓	–	✓
DIRAS family, GTP-binding RAS-like 3	*DIRAS3/ARH1*	–	✓	–
Stratifin	*SFN/14-3-3σ*	–	✓	✓
Underexpressed in malignant tumors				
Recombination activating gene 2	*RAG2*	✓	–	–
Angiotensin II receptor, type 1, transcript variant 5	*AGTR1*	✓	–	–
v-Kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homologue	*KIT*	–	✓	–
Thyroid peroxidase, transcript variant 5	*TPO5*	✓	–	–

[a] Gene list based on rank order from our microarray analysis (19).
[b] Data previously published (19).
[c] Data shown in present study.
FNA, fine needle aspiration; IHC, immunohistochemistry; QRT-PCR, quantitative real-time RT-PCR.

Supplementary Table S2. Primer Sequence, Annealing Condition, and Amplicon Size

Gene	*Forward sequence*	*Reverse sequence*	*Annealing*	*Amplicon*
HMGA2	5′-GAGTCCCTCTAAAGCAGCTCA-3′	5′-CTAGTCCTCTTCGGCAGACTCT-3′	56	156 bp
MRC2	5′-AAGTGGTCAGAGGCACAGTTCTCC-3′	5′-CCAGTTGGCATACATCAAAGGC-3′	55	189 bp
PLAG1	5′-ATGGAGTTACAAGGTGGCGTGC-3′	5′-CCAATGCTGGTGTGTTTAGGGG-3′	56	169 bp
DPP4	5′-TGGCGTGTTCAAGTGTGGAATAG-3′	5′-CAAGGTTGTCTTCTGGAGTTGGG-3′	55	110 bp
FNDC4	5′-TCACAGTGGAAGGTCTGGATGG-3′	5′-GGGATTGTTGTTGGAGTCATTGTC-3′	55	149 bp
CDH3	5′-TGAGAACCTGAAGGCGGCTAAC-3′	5′-AATCTTGGTCTTGGTCGGAGGC-3′	56	140 bp
CEACAM6	5′-GGGGAAAATCTGAACCTCTCCTG-3′	5′-CAGTGGCTGAGTTATGGGCTTG-3′	55	166 bp
PRSS3	5′-ATCAAACTCTCCTCACCTGCCGTC-3′	5′-TCGTCTGGGTAGTCAGCACCAAAG-3′	56	140 bp
SPOCK1	5′-GCTATTACAAAGCCACACAGTGCC-3′	5′-AAATCCCCTGAGGTTTCCTGC-3′	55	136 bp
SFN	5′-CCAGTATTGAGCAGAAAAGCAACG-3′	5′-TCGTAGTGGAAGACGGAAAAGTTC-3′	55	358 bp
GAPDH	5′-TGACAACTTTGGTATCGTGGAAGG-3′	5′-AGGCAGGGATGATGTTCTGGAG-3′	56	134 bp

Supplementary Table S3. Correlation Matrix of Protein Expression Among Gene Pairs

	HMGA2	*PLAG1*	*KLK7*	*CDH3*	*Stratifin*	*CYP1B1*	*DIRAS3*	*DPP4*	*MRC2*	*c-KIT*
HMGA2	1	0.36	0.26	0.39	0.47	0.33	0.45	0.25	0.35	0.17
PLAG1	0.36	1	0.52	0.59	0.35	0.34	0.47	0.41	0.56	0.43
KLK7	0.26	0.52	1	0.53	0.42	0.43	0.49	0.49	0.52	0.5
CDH3	0.39	0.59	0.53	1	0.23	0.36	0.64	0.43	0.56	0.5
Stratifin	0.47	0.35	0.42	0.23	1	0.32	0.55	0.4	0.48	0.18
CyP1B1	0.33	0.34	0.43	0.36	0.32	1	0.38	0.37	0.25	0.32
DIRAS3	0.45	0.47	0.49	0.64	0.55	0.38	1	0.45	0.58	0.47
DPP4	0.25	0.41	0.49	0.43	0.4	0.37	0.45	1	0.64	0.65
MRC2	0.35	0.56	0.52	0.56	0.48	0.25	0.58	0.64	1	0.66
c-KIT	0.17	0.43	0.5	0.5	0.18	0.32	0.47	0.65	0.66	1

Supplementary Table S4. Evaluation and Testing the Performance Between Gene-Combination Models

Gene combination	*AUC*	*p-value compared with All 10 genes model*	*p-value compared with* HMGA2 + MRC2 + Stratifin *model [DeLong's method (21)]*
HMGA2 + MRC2	0.9	0.007	0.06
HMGA2 + Stratifin	0.95	0.05	0.48
MRC2 + Stratifin	0.87	0.005	0.02
HMGA2 + MRC2 + Stratifin	0.95	0.08	–
HMGA2 + MRC2 + Stratifin + CyP1B1	0.97	0.14	0.15
All 10 genes	0.99	–	0.08

AUC, area under the curve.