

Supplementary Material

1 SUPPLEMENTARY TABLES AND FIGURES

1.1 Figures

Comparison of DEC performance on different k

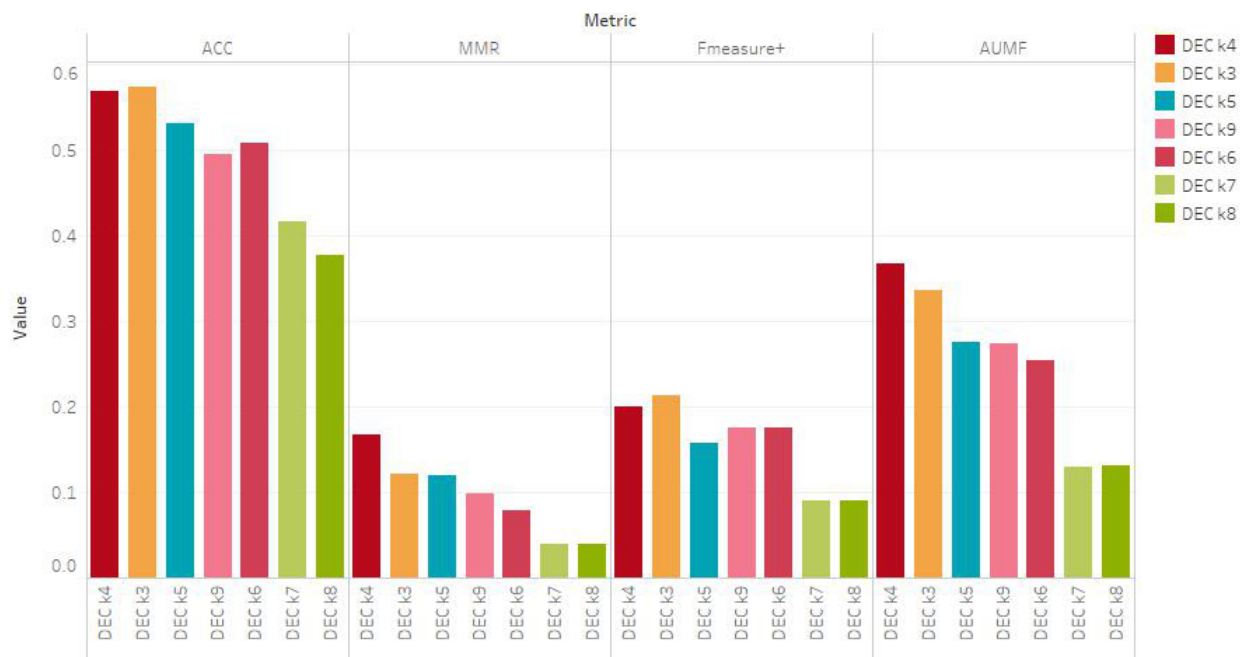


Figure S1: DEC performance on different k s.

Comparison of Methods

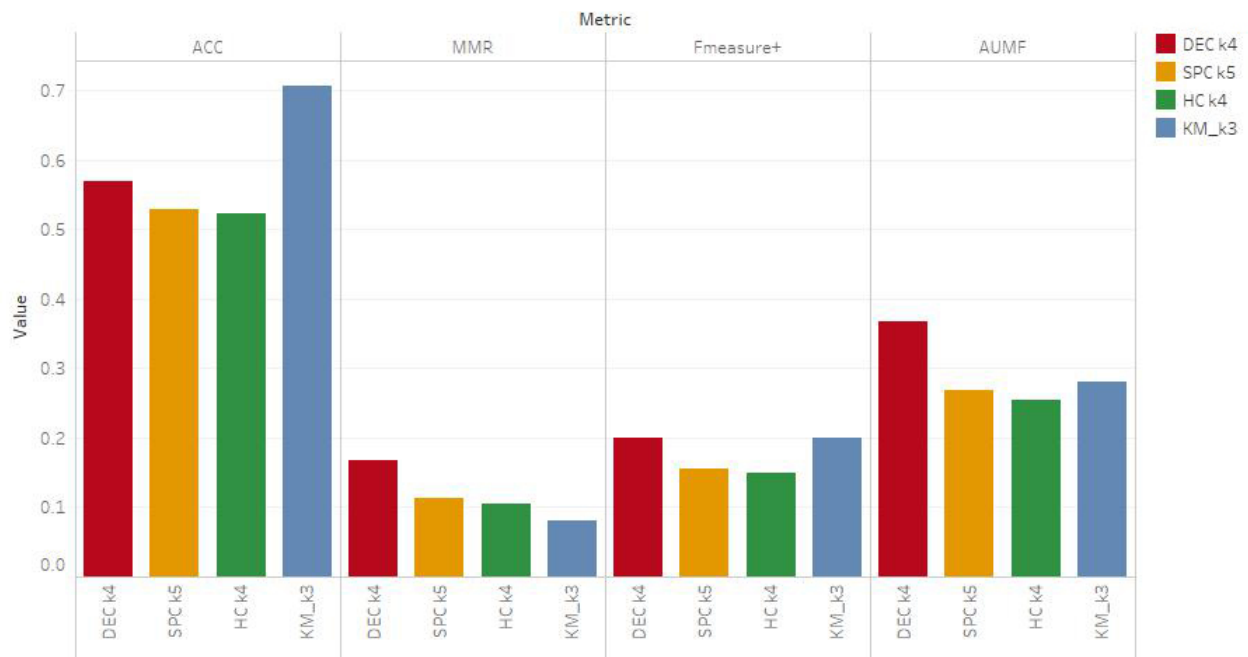


Figure S2: Comparison of the DEC results with other clustering methods.

Primary

GENE	Pvalue
CDH1	7.37e-05
TBX3	1.98e-04
MAP3K1	2.78e-04
CTCF	4.42e-04
CBFB	1.32e-03
PIK3CA	1.77e-03
MED23	2.05e-03
FBXO11	7.70e-03
GATA3	9.91e-03
COL4A1	1.17e-02
GPS2	1.36e-02
MAP4K1	1.76e-02
RASGRF2	1.76e-02
TAOK1	1.76e-02
TSC1	2.09e-02
CDKN1B	2.35e-02
TBL1XR1	2.35e-02
RUNX1	2.72e-02
EFTUD2	3.39e-02
EXT2	3.39e-02
ITGA11	3.39e-02
PRKCQ	3.39e-02
PTPRJ	3.39e-02
PIK3C2G	3.94e-02
TNC	3.94e-02
CD22	4.42e-02
DKK4	4.42e-02
PRLR	4.42e-02
TRIP10	4.42e-02
COL5A3	4.82e-02

Figure S3: Top gene signature of *Primary* subtype.

Progressive

GENE	Pvalue
TP53	3.30e-13
MYC	6.12e-08
GNAS	1.03e-06
ERBB2	1.58e-06
CACNA1F	2.34e-06
FGFR3	4.23e-06
AKT3	8.51e-06
BRAF	1.71e-05
RB1	2.33e-05
MPL	2.52e-05
AKT1	5.43e-05
CACNA1S	6.99e-05
FANCA	6.99e-05
CCNE1	9.07e-05
CCR7	1.13e-04
CDKN2A	1.23e-04
MACF1	1.47e-04
EGFR	1.65e-04
JAK1	2.38e-04
SRSF2	3.12e-04
LAMB4	3.59e-04
STK11	3.59e-04
BRCA1	4.10e-04
ANAPC5	4.70e-04
SPOP	5.79e-04
CARD11	6.55e-04
TLN1	6.55e-04
ARID4B	9.39e-04
COL4A6	9.39e-04
SETDB1	9.39e-04
KIT	1.07e-03
BRCA2	1.15e-03
NCOA3	1.23e-03
CLASP2	1.31e-03
RARB	1.31e-03
BCL2	1.69e-03
EZH2	1.84e-03
LAMB1	1.84e-03
MYB	1.84e-03
IDH2	2.20e-03
SKP2	2.52e-03
DDX3X	2.84e-03
GATA2	2.84e-03
NCF2	2.84e-03
ROCK2	2.84e-03
STAT4	2.84e-03
TNFAIP3	2.84e-03
TNKS1BP1	2.84e-03
IGF1R	3.22e-03
ARID1B	3.41e-03

Figure S4: Top gene signature of *Progressive* subtype.

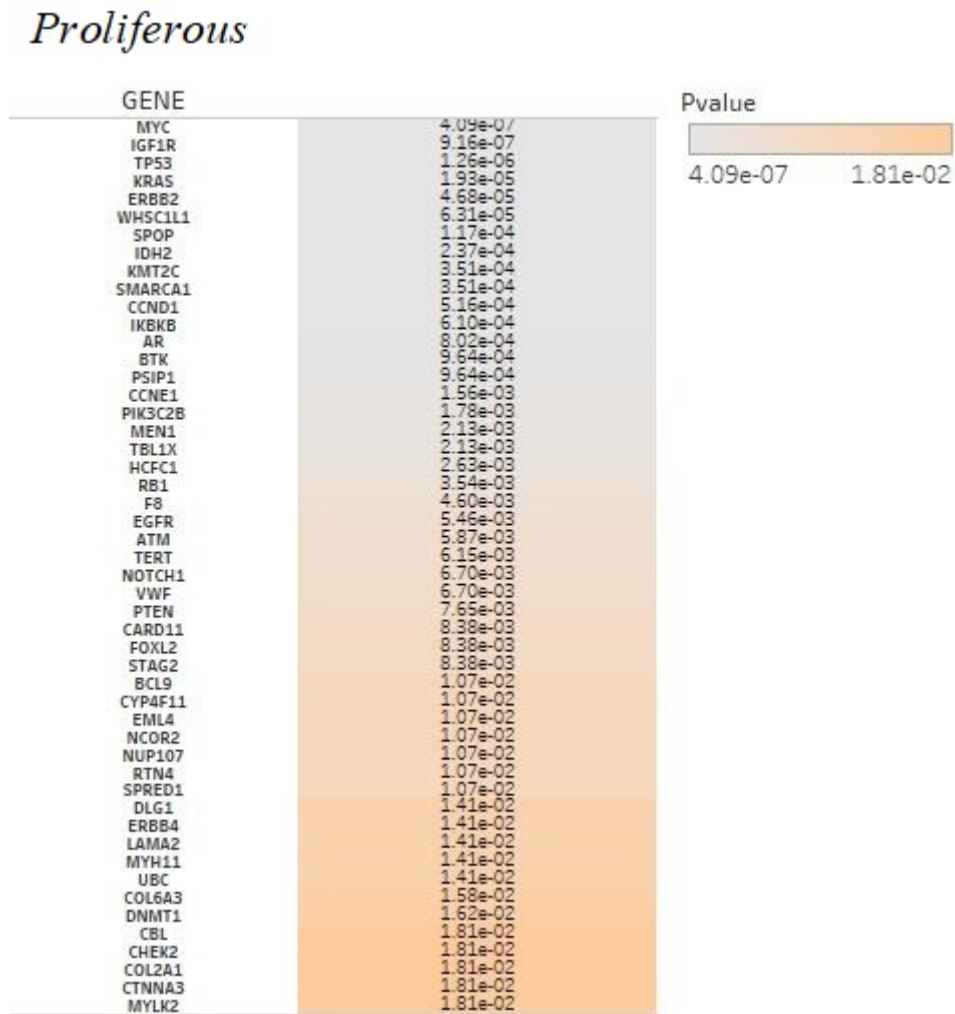
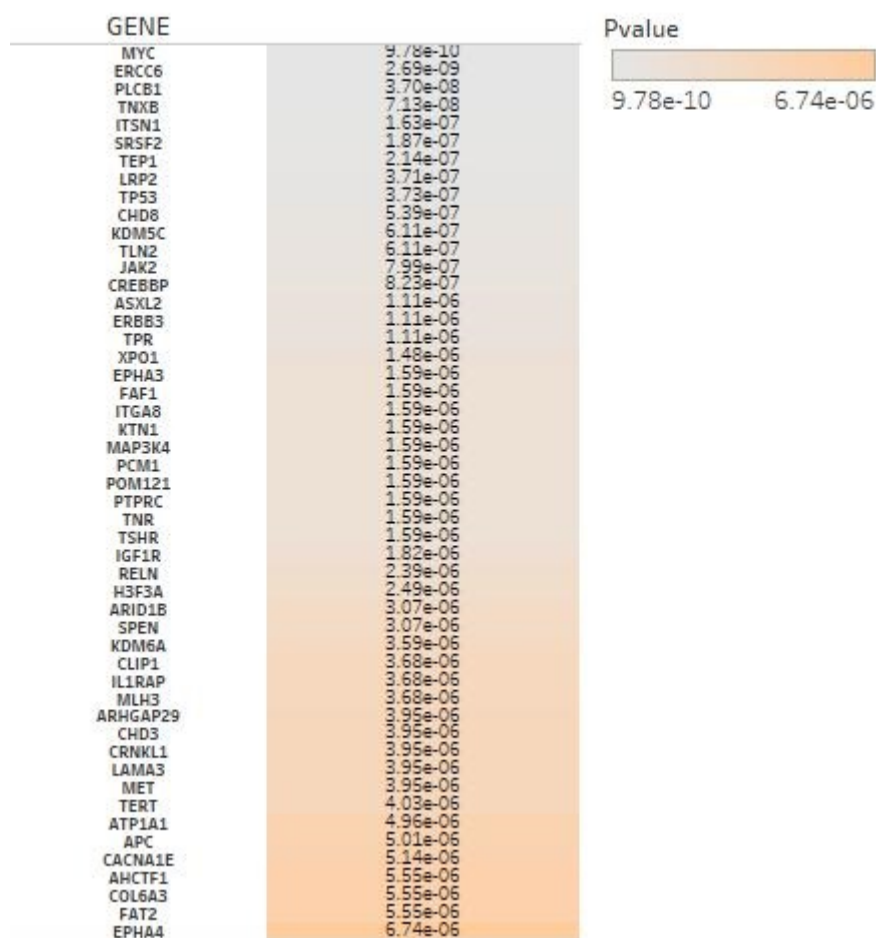


Figure S5: Top gene signature of *Proliferous* subtype.

PerilousFigure S6: Top gene signature of *Perilous* subtype.

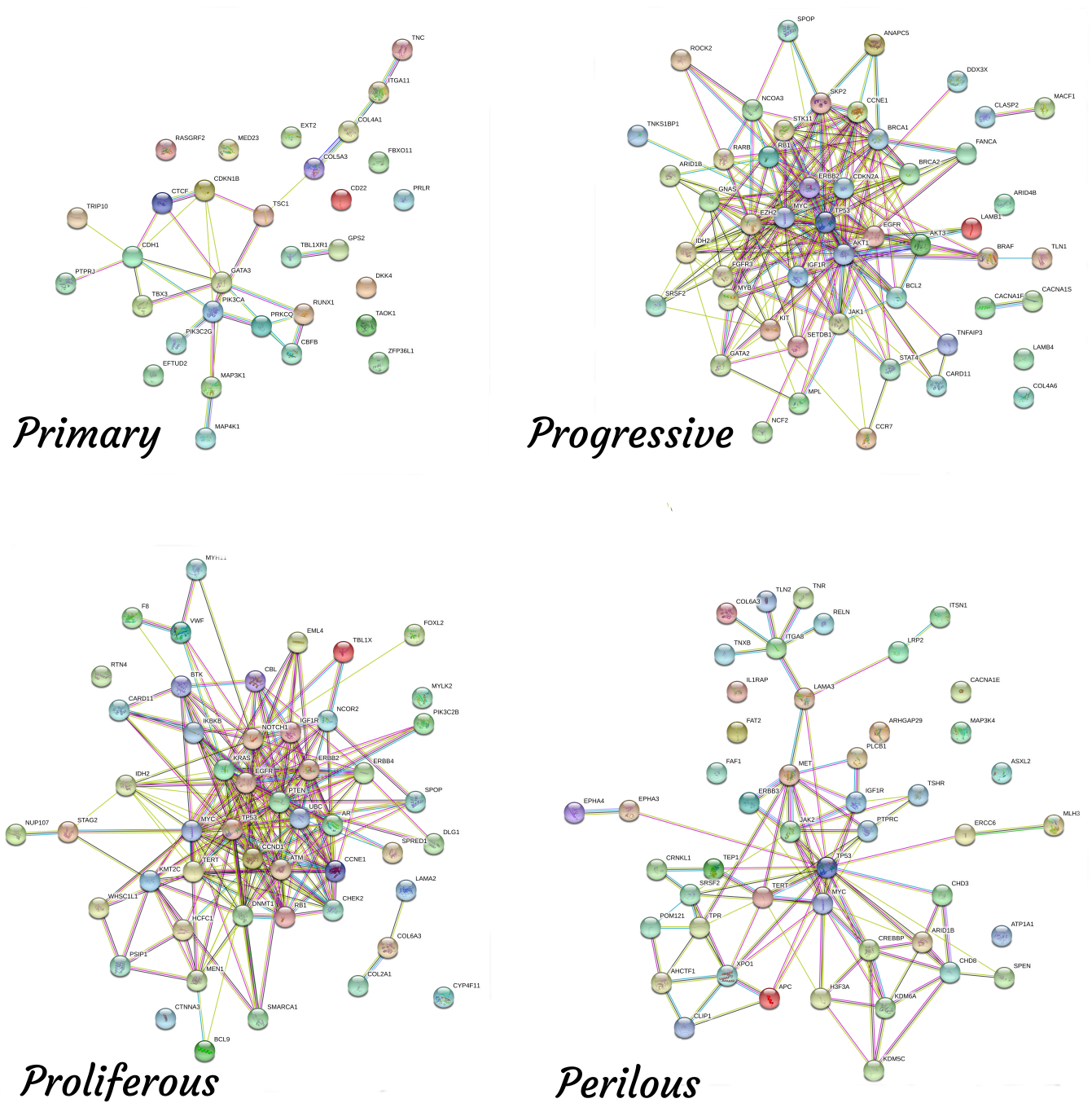


Figure S7: PPI network for each subtype.

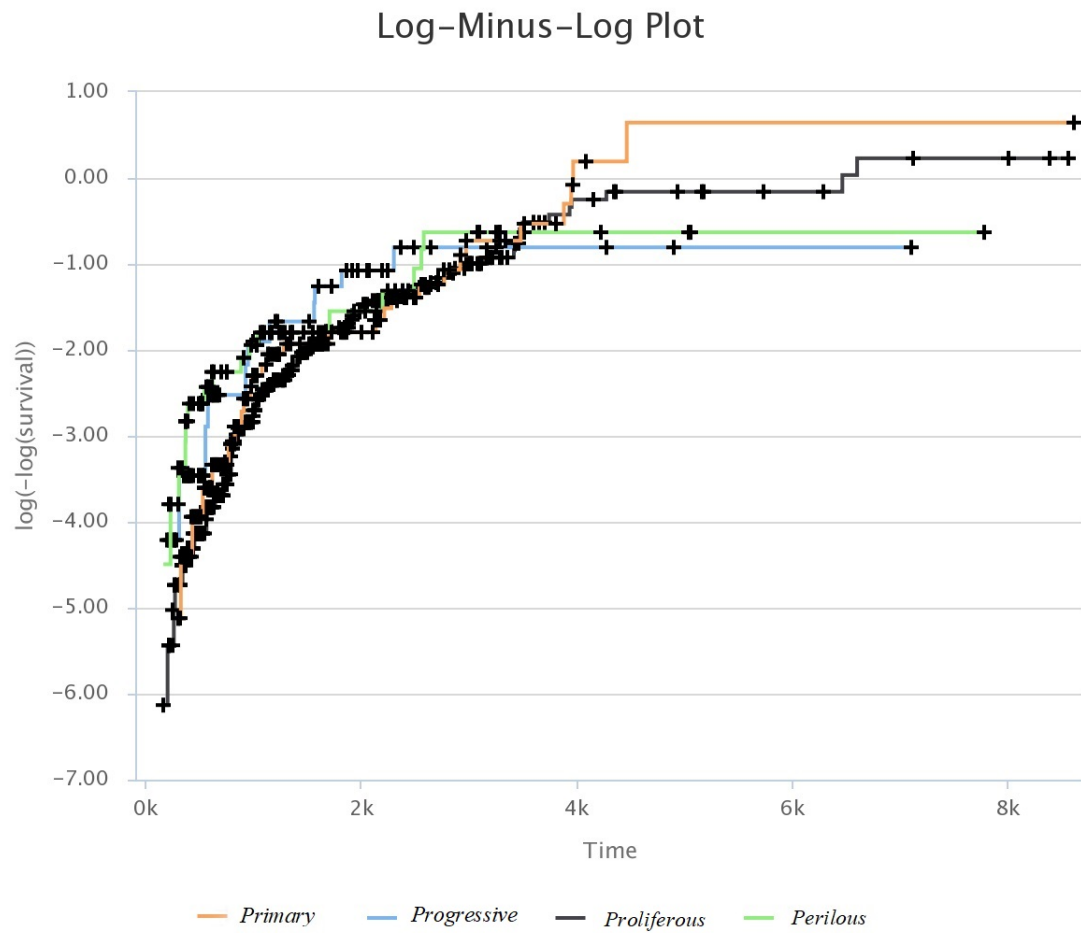


Figure S8: Cox hazard survival diagram.

Breast Carcinoma Estrogen Receptor Status		
	Negative	Positive
<i>Primary</i>	49	125
<i>Progressive</i>	35	40
<i>Proliferous</i>	53	412
<i>Perilous</i>	44	49

Figure S9: Contingency table of ER status relation with discovered subtypes ($p\text{-value} < 2.2e - 16$ by Chi-square test and $p\text{-value} = 1e - 06$ by Fisher's exact test).

Breast Carcinoma Progesterone Receptor Status		
	Negative	Positive
<i>Primary</i>	64	109
<i>Progressive</i>	47	29
<i>Proliferous</i>	93	369
<i>Perilous</i>	54	39

Figure S10: Contingency table of PR status relation with discovered subtypes ($p\text{-value} < 2.2e - 16$ by the Chi-square test and $p\text{-value} = 1e - 06$ by the Fisher's exact test.)

Lab Proc Her2 Neu Immunohistochemistry Receptor Status			
	Equivocal	Negative	Positive
<i>Primary</i>	24	88	44
<i>Progressive</i>	5	40	22
<i>Proliferous</i>	91	265	44
<i>Perilous</i>	15	44	23

Figure S11: Contingency table of HER2 status relation with discovered subtypes ($p - value = 1.445e - 07$ by the chi-square test and $p - value = 1e - 06$ by the Fisher's exact test).

TP53 Mutation Status		
	p53mut	p53wt
<i>Primary</i>	88	94
<i>Progressive</i>	43	34
<i>Proliferous</i>	92	396
<i>Perilous</i>	61	37

Figure S12: Contingency table of TP53 status relation with discovered subtypes ($p - value < 2.2e -$ in both the Chi-squared test and the Fisher's exact test).

Histological Type							
	Infiltrating Ductal Carcinoma	Infiltrating Lobular Carcinoma	Medullary Carcinoma	Metaplastic Carcinoma	Mixed Histology	Mucinous Carcinoma	Other, specify
<i>Primary</i>	147	23	3	1	1	2	5
<i>Progressive</i>	69	7					1
<i>Proliferous</i>	331	103		2	20	10	22
<i>Perilous</i>	80	8	2		2		4

Figure S13: Contingency table of histopathological subtypes relation with discovered subtypes ($p - value = 0.0001615$ by the chi-square test and $p - value = 5.4e - 05$ by the Fisher's exact test).

	MSDEC Subtypes			
	Primary	Progressive	Proliferous	Perilous
Luminal A	72	18	338	15
Luminal B	52	16	99	28
Basal-like	35	30	46	38
Her2	23	18	16	17

Figure S14: Contingency table of discovered subtypes relation with PAM50 subtypes (p – value $< 2.2E - 16$ by the Chi-square test and p – value $= 1E - 06$ by the Fisher's exact test).

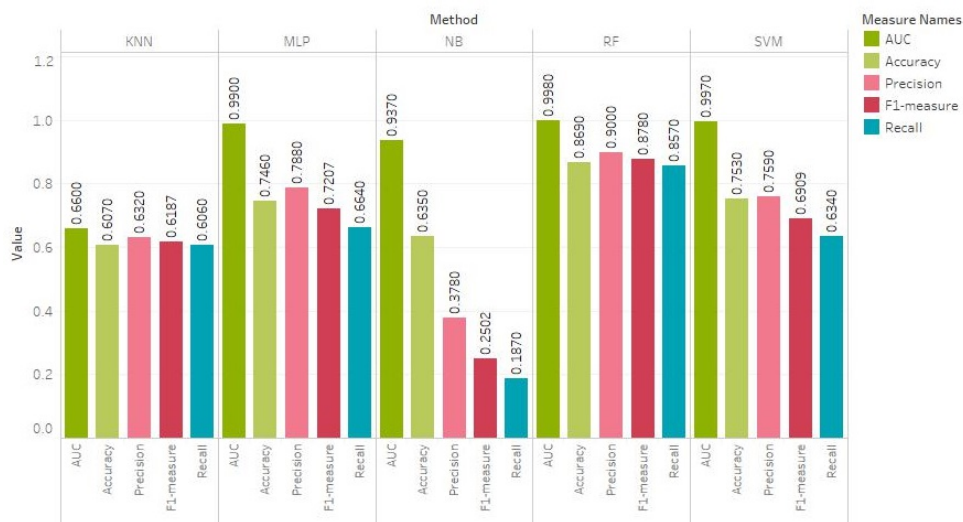


Figure S15: Comparison of random forest method with other classification methods.

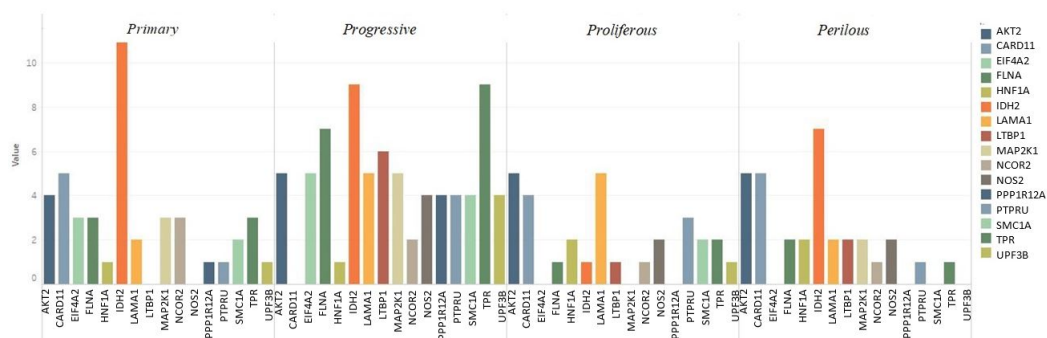


Figure S16: Mutational frequency of 16 important genes for classification of tumors in each subtype.

	cytokines and growth factors	transcription factors	homeodomain proteins	cell differentiation markers	protein kinases	translocated cancer genes	oncogenes	tumor suppressors
tumor suppressors	0	1	0	1	0	0	0	4
oncogenes	0	1	0	0	0	2	3	
translocated cancer genes	0	1	0	0	0	2		
protein kinases	0	0	0	0	4			
cell differentiation markers	0	0	0	3				
homeodomain proteins	0	0	0					
transcription factors	0	6						
cytokines and growth factors	1							

Figure S17: Family of gene signature in *Primary* subtype.

	cytokines and growth factors	transcription factors	homeodomain proteins	cell differentiation markers	protein kinases	translocated cancer genes	oncogenes	tumor suppressors
tumor suppressors	0	2	0	0	1	0	0	8
oncogenes	0	3	0	4	6	5	15	
translocated cancer genes	0	2	0	1	2	5		
protein kinases	0	0	0	4	10			
cell differentiation markers	0	0	0	6				
homeodomain proteins	0	0	0					
transcription factors	0	11						
cytokines and growth factors	0							

Figure S18: Family of gene signature in *Progressive* subtype.

	cytokines and growth factors	transcription factors	homeodomain proteins	cell differentiation markers	protein kinases	translocated cancer genes	oncogenes	tumor suppressors
tumor suppressors	0	2	0	0	2	0	0	6
oncogenes	0	3	0	1	2	8	14	
translocated cancer genes	0	2	0	0	0	9		
protein kinases	0	0	0	2	9			
cell differentiation markers	0	0	0	2				
homeodomain proteins	0	0	0					
transcription factors	0	8						
cytokines and growth factors	0							

Figure S19: Family of gene signature in *Proliferous* subtype.

	cytokines and growth factors	transcription factors	homeodomain proteins	cell differentiation markers	protein kinases	translocated cancer genes	oncogenes	tumor suppressors
tumor suppressors	0	2	0	0	0	0	0	4
oncogenes	0	2	0	0	1	6	8	
translocated cancer genes	0	2	0	0	0	6		
protein kinases	0	0	0	1	6			
cell differentiation markers	0	0	0	2				
homeodomain proteins	0	0	0					
transcription factors	0	8						
cytokines and growth factors	0							

Figure S20: Family of gene signature in *Perilous* subtype.

1.2 Tables

Complex Name	Proteins
<i>NCOR complex</i>	NCOR1, HDAC3, GPS2, TBL1XR1, TBL1X, CORO2A
<i>SMRT complex</i>	TBL1X, TBL1XR1, GPS2, HDAC3, NCOR2
<i>Kaiso – NCOR complex</i>	CORO2A, TRIM33, TBL1X, TBL1XR1, GPS2, HDAC3, KIF11, KDM4A, NCOR1, ZBTB33
<i>NCOR – HDAC3 complex</i>	TBL1X, TBL1XR1, GPS2, HDAC3, NCOR1
<i>Polycystin – 1 – E – cadherin – beta – catenin complex</i>	CDH1, CTNNB1, PKD1
<i>p27 – cyclinE – CDK2 complex</i>	CDK2, CDKN1B, CCNE1
<i>p27 – cyclinE – Cdk2 – Ubiquitin E3 ligase complex</i>	SKP1A, SKP2, CUL1, CKS1B, RBX1, CCNE1, CDKN1B, CDK2
<i>Polycystin – 1 – E – cadherin – beta – catenin – Flotillin – 2 complex</i>	CDH1, CTNNB1, FLOT2, PKD1
<i>Polycystin – 1 multiprotein complex</i>	ACTN1, CDH1, SRC, JUP, VCL, CTNNB1, PXN, BCAR1, PKD1, PTK2, TLN1

Table S1. Protein complexes of *Primary* subtype.

Complex Name	Proteins
<i>BRCC complex</i>	BRE, BRCA1, BRCA2, RAD51, BRCC3
<i>MSH26 – BLM – p53 – RAD51 complex</i>	BLM, TP53, MSH2, MSH6, RAD51
<i>BRCA1 – BARD1 – BACH1 – DNA damage complex II</i>	BARD1, BRCA1, RBBP8, RAD50, TOPBP1, MRE11A, NBN, BACH1
<i>BRCA1 – BARD1 – BACH1 – DNA damage complex I</i>	BARD1, BRCA1, MLH1, MSH6, TOPBP1, BACH1
<i>BRCA1 – BARD1 – BACH1 – DNA damage complex III</i>	BARD1, BRCA1, RBBP8, RAD50, TOPBP1, MRE11A, NBN, BACH1
<i>BRCA1A complex</i>	FAM175A, UIMC1, BARD1, BRCA1
<i>BRCA1C complex</i>	UIMC1, BARD1, BRCA1, RBBP8
<i>BRCA1B complex</i>	BARD1, BRCA1, BACH1
<i>BRCA1 – CTIP – ZBRK1 repressor complex</i>	BRCA1, RBBP8, ZNF350
<i>BARD1 – BRCA1 – CSTF64 complex</i>	BARD1, BRCA1, CSTF2
<i>BRAF53 – BRCA2 complex</i>	BRCA2, HDAC1, HDAC2, KDM1A, PHF21A, RCOR1, HMG20B
<i>BRCA1 – BARD1 – POLR2A complex</i>	BARD1, BRCA1, POLR2A
<i>BRCA1 – CtIP – CtBP complex</i>	BRCA1, CTBP1, RBBP8
<i>BRCA1 – LMO4 – CTIP complex</i>	BRCA1, RBBP8, LMO4
<i>BRCA1 – BARD1 – UbcH7c complex</i>	BARD1, BRCA1, UBE2L3

Table S2. Protein complexes of *Progressive* subtype.

Complex Name	Proteins
<i>Paf complex</i>	PAF1, CTR9, LEO1, WDR61
<i>p27 – cyclinE – CDK2 complex</i>	CDK2, CDKN1B, CCNE1
<i>MDC1 – MRN – ATM – FANCD2 complex</i>	MRE11A, FANCD2, MDC1, NBN, ATM, RAD50
<i>MDC1 – H2AFX – TP53BP1 complex</i>	H2AFX, MDC1, TP53BP1
<i>MDC1 – p53BP1 – SMC1 complex</i>	MDC1, SMC1A, TP53BP1
<i>SH3KBP1 – CBLB – EGFR complex</i>	CBLB, EGFR, SH3KBP1
<i>E2F4 – p130 complex</i>	RBL2, E2F4
<i>CIN85 – CBL complex</i>	CBL, SH3KBP1
<i>MLL – HCF complex</i>	SERPIND1, KMT2A, HCFC1, MEN1, MEN1, KDM6B, ASH2L, WDR5
<i>RAB5 – EEA1 complex</i>	EEA1, RAB5A
<i>EGFR – CBL – GRB2 complex</i>	CBL, EGFR, GRB2
<i>NCOR complex</i>	NCOR1, HDAC3, GPS2, TBL1XR1, TBL1X, CORO2A
<i>SMRT complex</i>	TBL1X, TBL1XR1, GPS2, HDAC3, NCOR2
<i>SMRT core complex</i>	TBL1X, HDAC3, NCOR2

Table S3. Protein complexes of *Proliferous* subtype.

<i>Complex Name</i>	<i>Proteins</i>
<i>Anaphase – promoting complex</i>	ANAPC1, ANAPC10, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC27
<i>SHARP – CtBP complex</i>	CTBP1, CTBP2, SPEN
<i>MDC1 – H2AFX – TP53BP1 complex</i>	H2AFX, MDC1, TP53BP1
<i>MDC1 – p53BP1 – SMC1 complex</i>	MDC1, SMC1A, TP53BP1
<i>JAK2 – PAFR – TYK2 complex</i>	PTAFR, JAK2, TYK2
<i>Hes1 – Tle1 complex</i>	Hes1, Tle1
<i>PLCB1 – PARD3 – PARD6A complex</i>	PLCB1, PARD3, PARD6A
<i>RSmad complex</i>	ARID1B, CREBBP, TRIM33, SMAD2, SMAD3, NCOA3, SMAD4, SMARCC1, SMARCC2, SMARCA4 (BAF190A)
<i>PTIP – HMT complex</i>	KMT2B, KMT2C, KDM6A, NCOA6, PAXIP1, PAGR1, DPY30, KDM6B, ASH2L, WDR5
<i>UTX – MLL2/3 complex</i>	KMT2C, KMT2D, KDM6A, N4BP2, NCOA6, PAXIP1, PROSER1, KDM6B, PPP6R3, ASH2L, WDR5, ZNF281
<i>ITGAV – ITGB3 – LAMA4 complex</i>	ITGAV, ITGB3, LAMA4
<i>HES1 promoter corepressor complex</i>	CREBBP, CDK7, POLR2A, EP300, RBPJ, SUPT6H
<i>ESR1 – MDM4 complex</i>	ESR1, MDM4

Table S4. Protein complexes of *Perilous* subtype.