Supplementary Material

1 SUPPLEMENTARY TABLES AND FIGURES

1.1 Figures

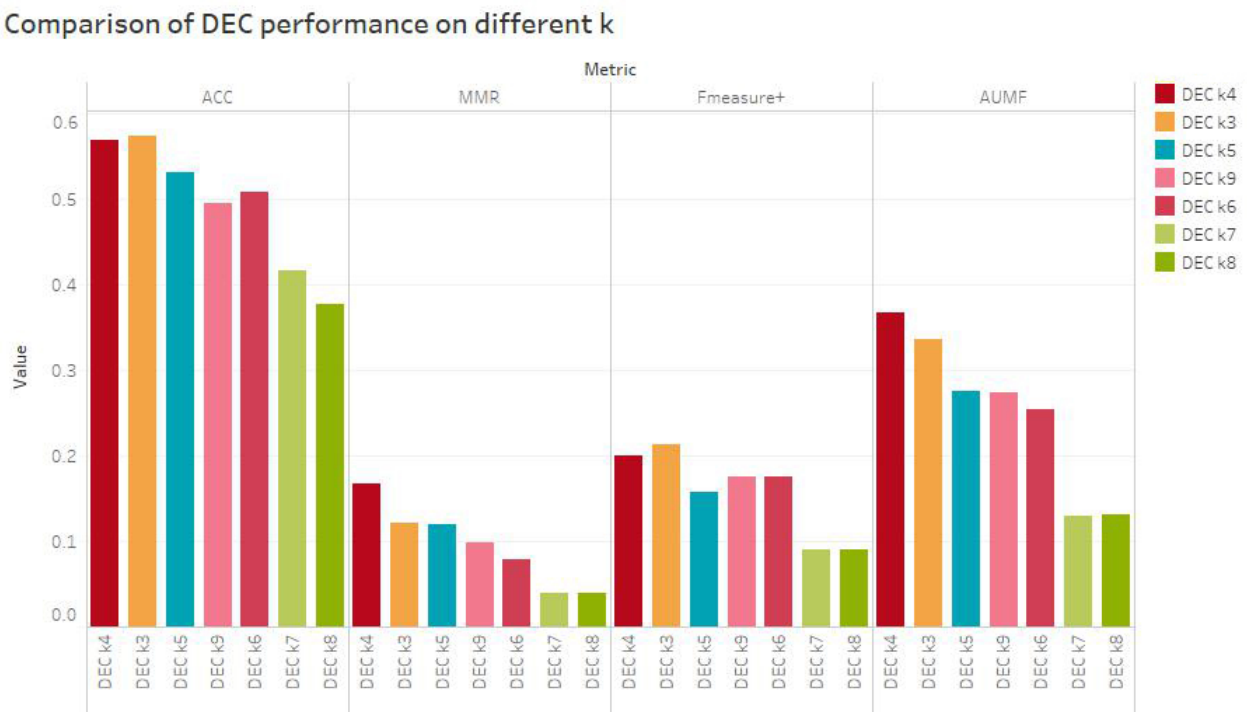

Figure S1: DEC performance on different ks.

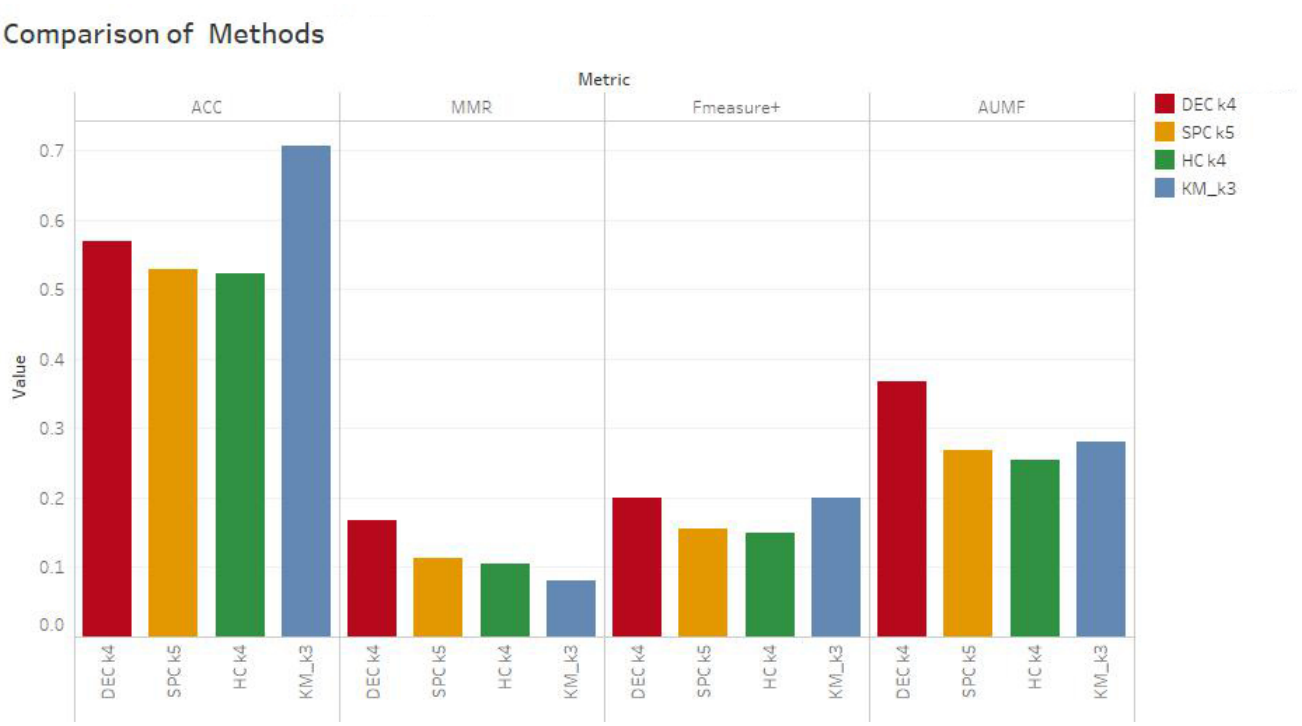

Figure S2: Comparison of the DEC results with other clustering methods.

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

Pvalue: 7.37e-05 — 4.82e-02

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

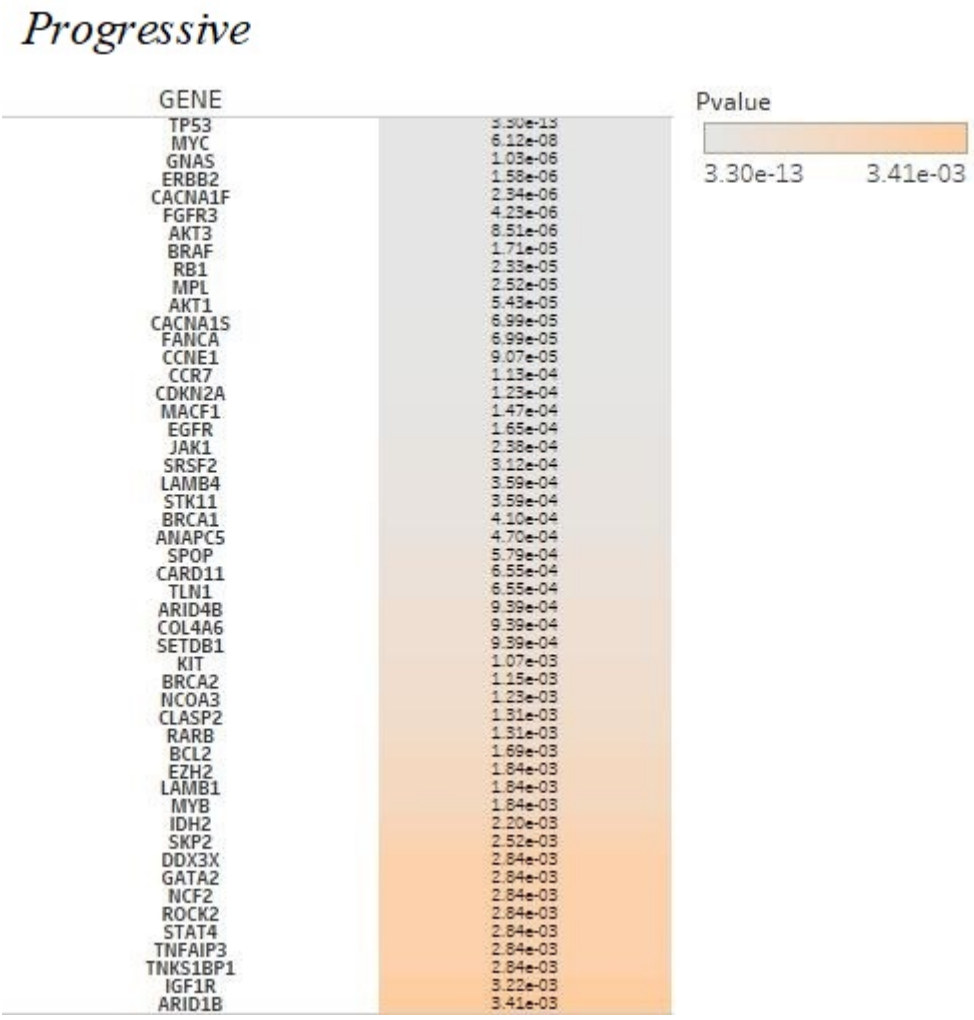

Figure S4: Top gene signature of $Progressive$ subtype.

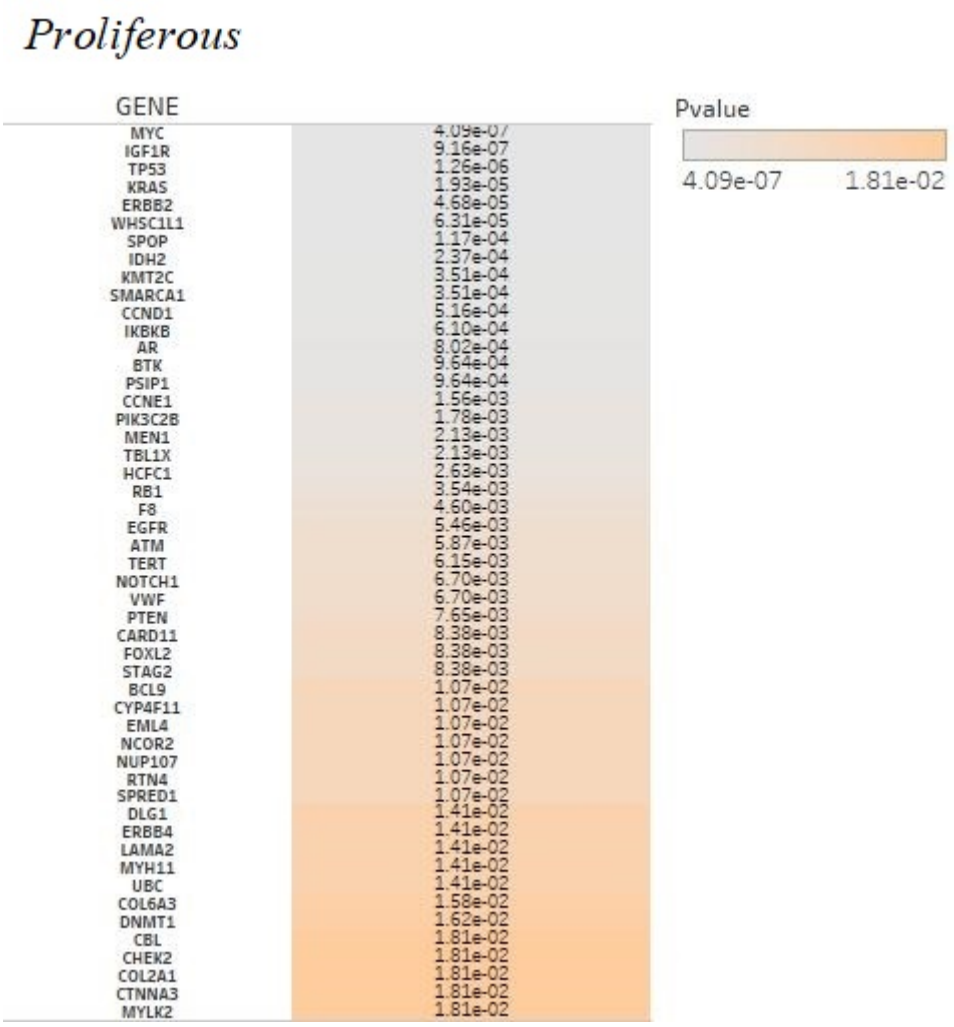

Figure S5: Top gene signature of $Proliferous$ subtype.

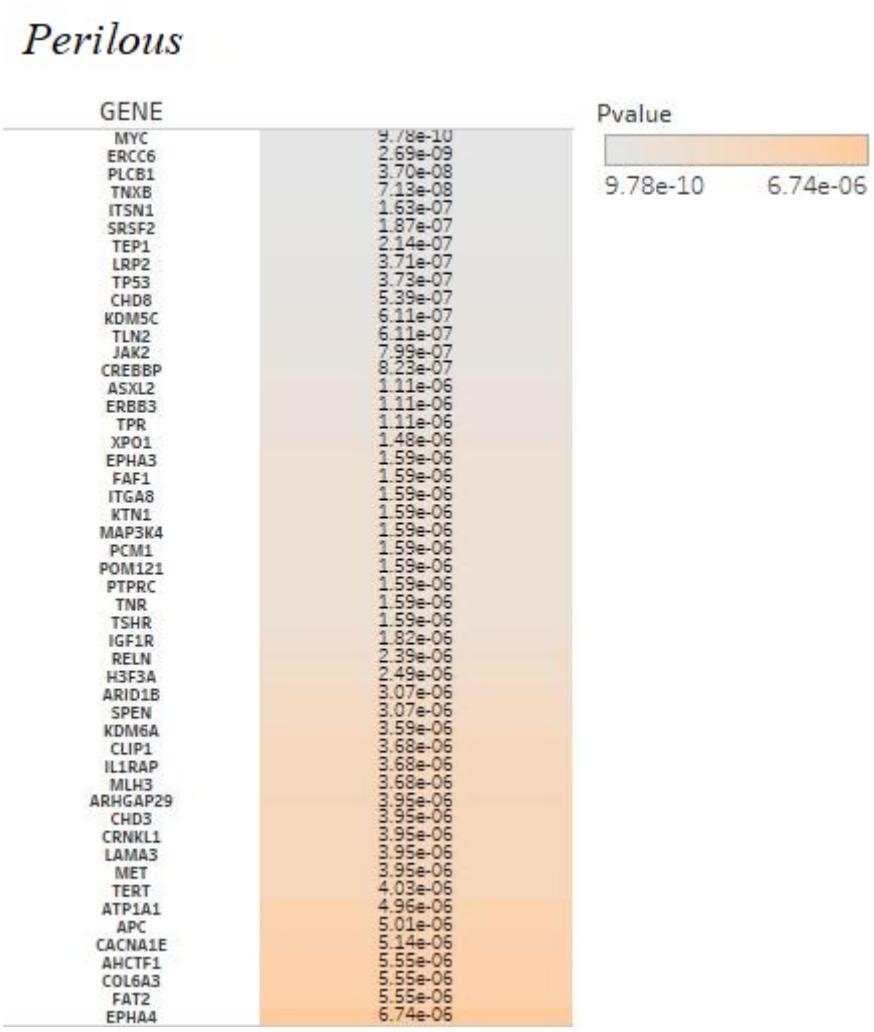

Figure S6: Top gene signature of $Perilous$ subtype.

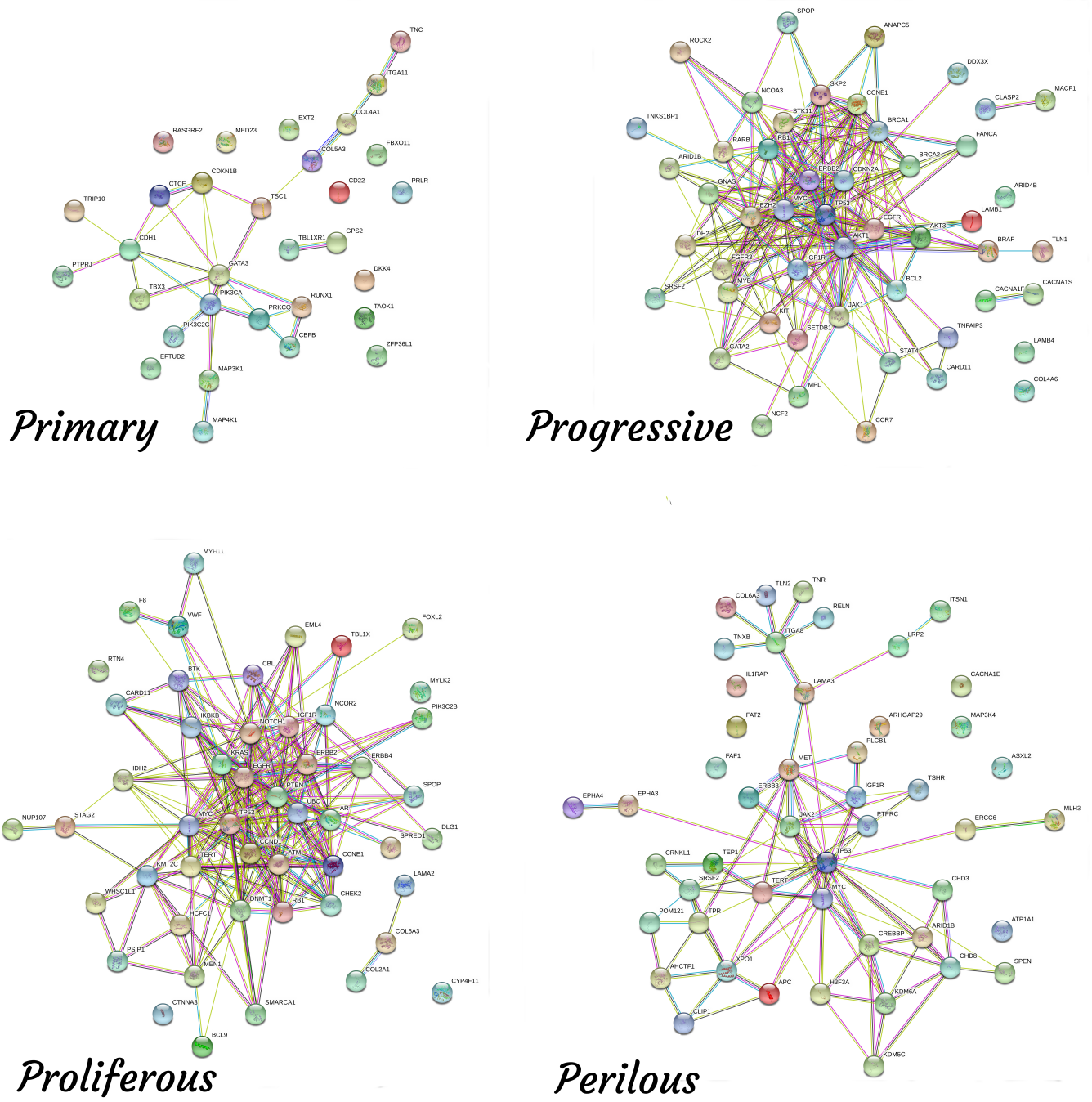

Figure S7: PPI network for each subtype.

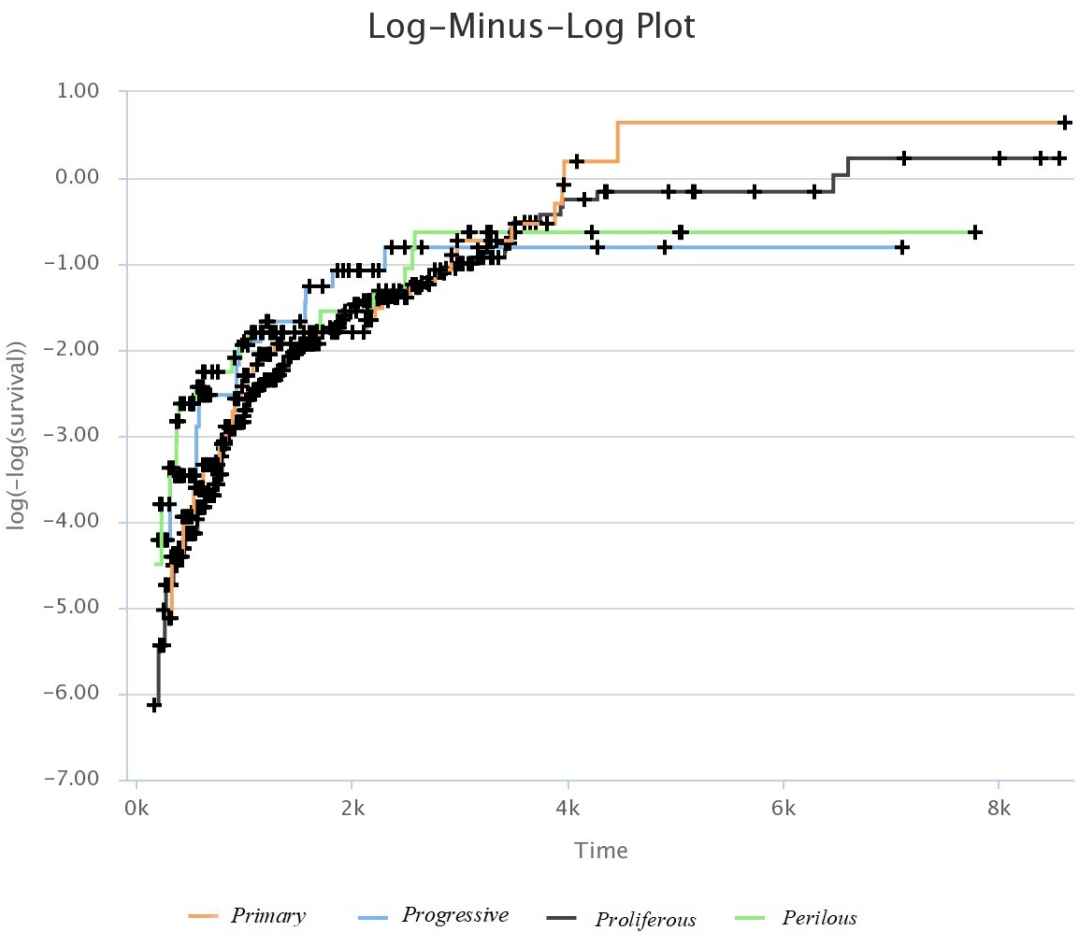

Figure S8: Cox hazard survival diagram.

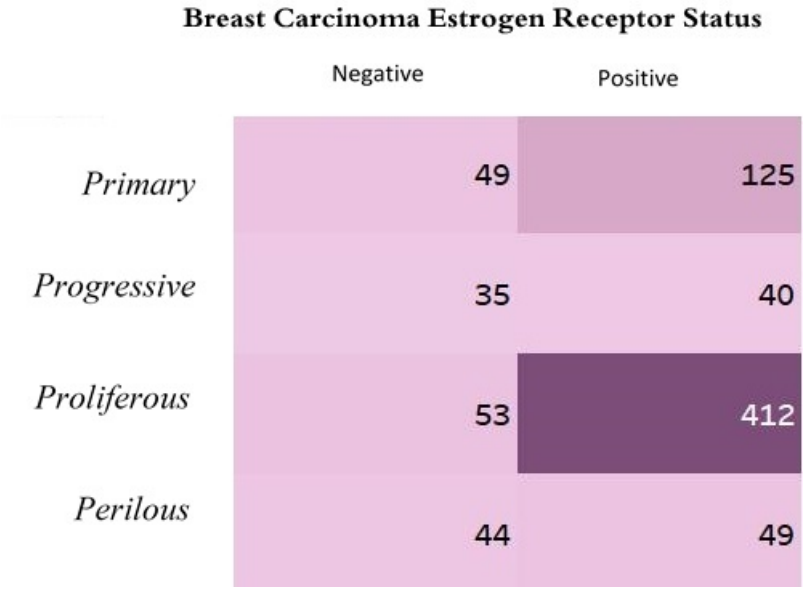

Figure S9: Contingency table of ER status relation with discovered subtypes ($p-value < 2.2e-16$ by Chi-square test and $p-value = 1e-06$ by Fisher's exact test).

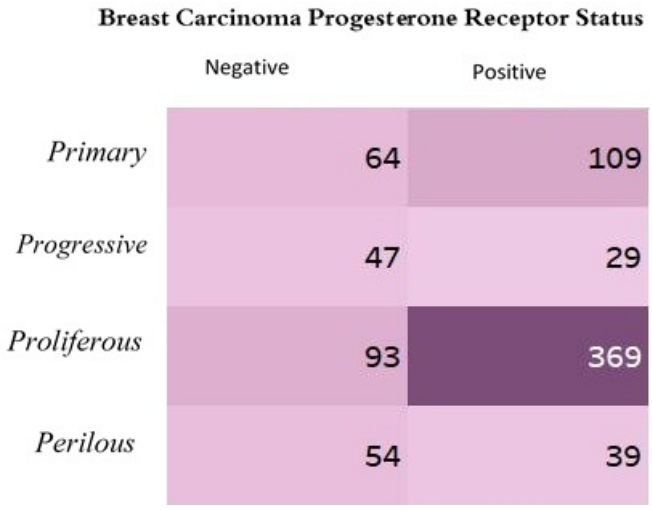

Figure S10: Contingency table of PR status relation with discovered subtypes ($p-value < 2.2e-16$ by the Chi-square test and $p-value = 1e-06$ by the Fisher's exact test.)

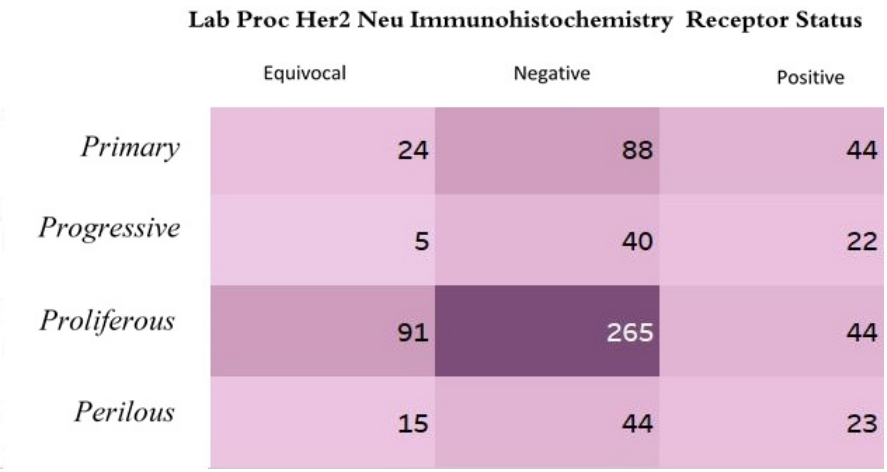

Figure S11: Contingency table of HER2 status relation with discovered subtypes ($p - value = 1.445e - 07$ by the chi-square test and $p - value = 1e - 06$ by the Fisher's exact test).

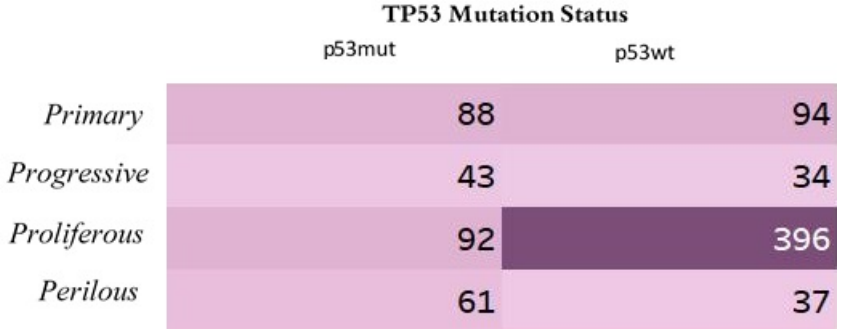

Figure S12: Contingency table of TP53 status relation with discovered subtypes ($p - value < 2.2e-$ in both the Chi-squared test and the Fisher's exact test).

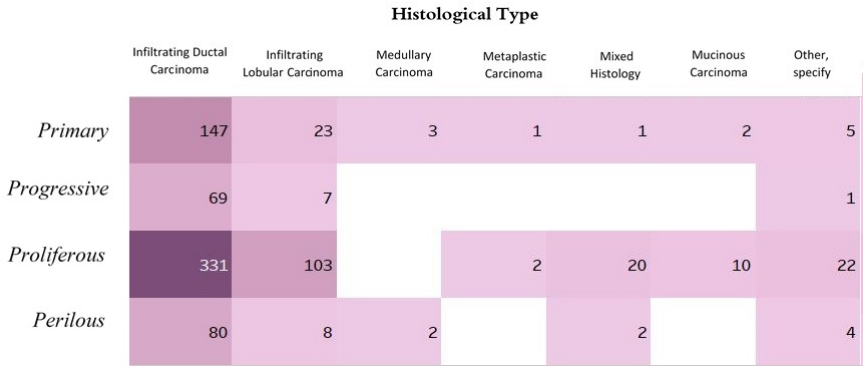

Figure S13: Contingency table of histopathological subtypes relation with discovered subtypes ($p - value = 0.0001615$ by the chi-square test and $p - value = 5.4e - 05$ by the Fisher's exact test).

	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).

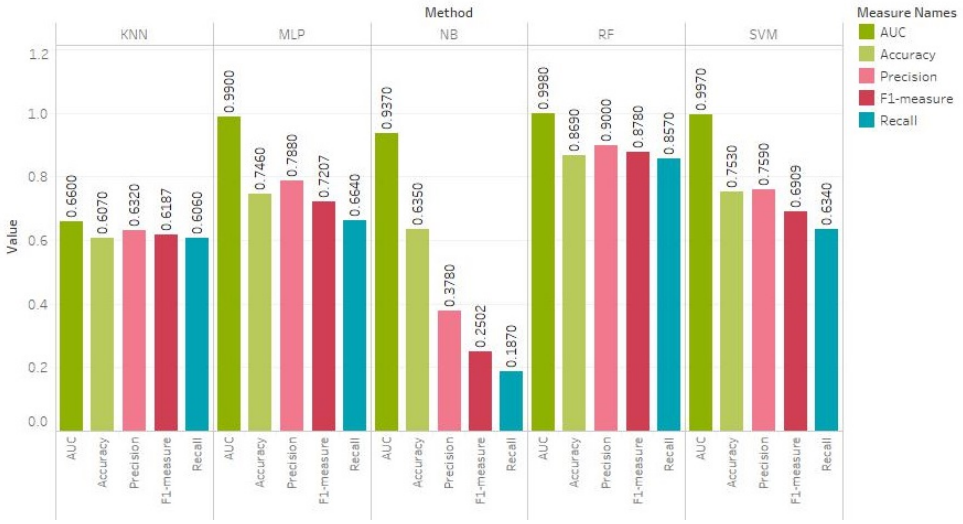

Figure S15: Comparison of random forest method with other classification methods.

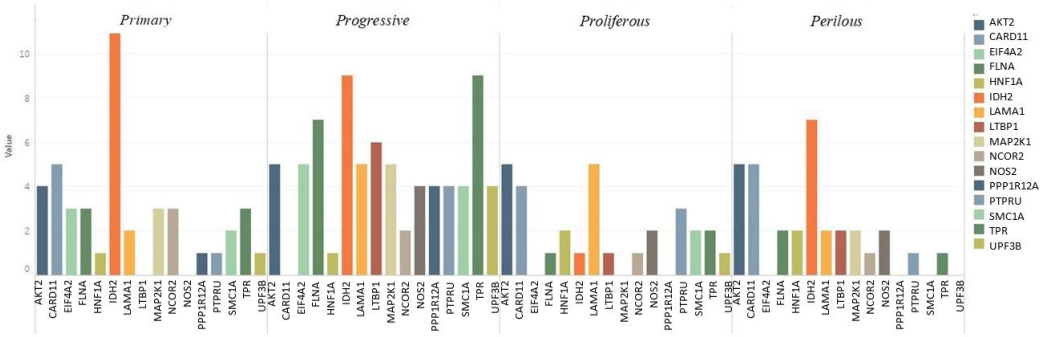

Figure S16: Mutational frequency of 16 important genes for classification of tumors in each subtype.

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

Table S1. Protein complexes of $Primary$ subtype.

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

Table S2. Protein complexes of $Progressive$ subtype.

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

Table S3. Protein complexes of $Proliferous$ subtype.

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

Table S4. Protein complexes of $Perilous$ subtype.