Bioinformatics in Medical Research

Bioinformatics Institute
A*STAR

NANYANG TECHNOLOGICAL UNIVERSITY

Genome and transcriptome delineation of two major oncogenic pathways governing invasive ductal breast cancer development.

Vladimir A. Kuznetsov

13/04/2016

vladimirk@bii.a-star.edu.sg

Bioinformatics Institute/A*STAR

A*STAR, BII, GGEDA

Agency for Science, Technology and Research
SINGAPORE

A*STAR BII

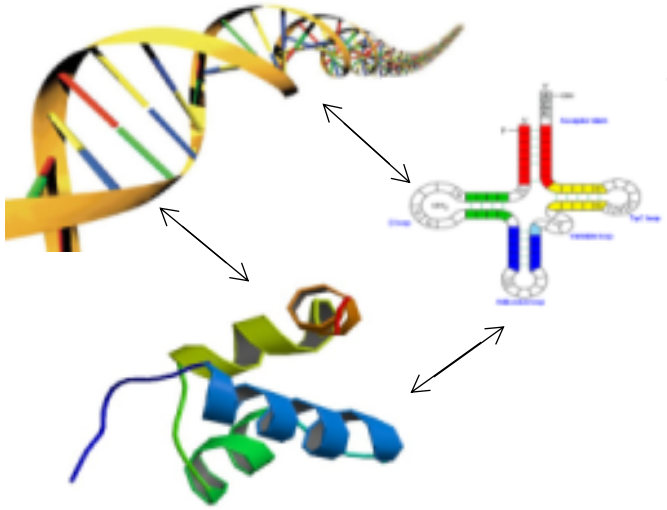
A*STAR Mission and Vision Statements

Mission Statement

The Agency for Science, Technology and Research (A*STAR) drives mission-oriented research that advances scientific discovery and technological innovation.

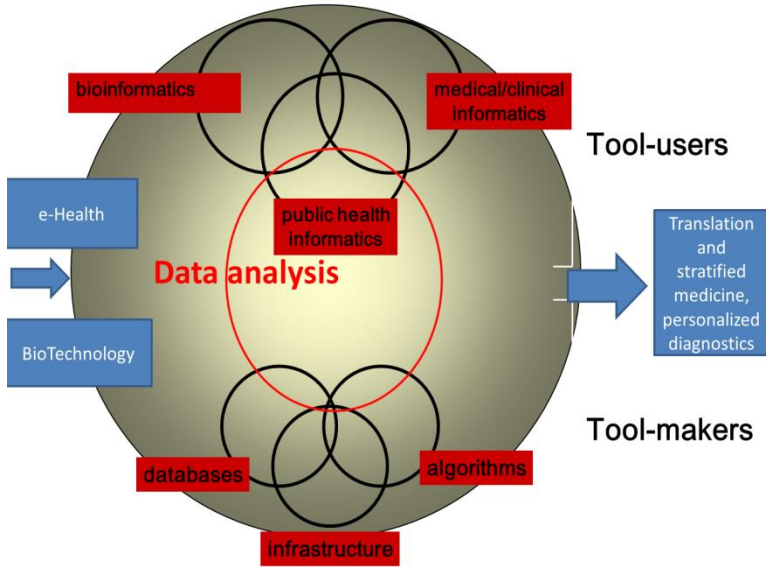
N9 (by Thompson Reuters, 2015) among government Research Organization

Genome and Gene Data Expression Analysis



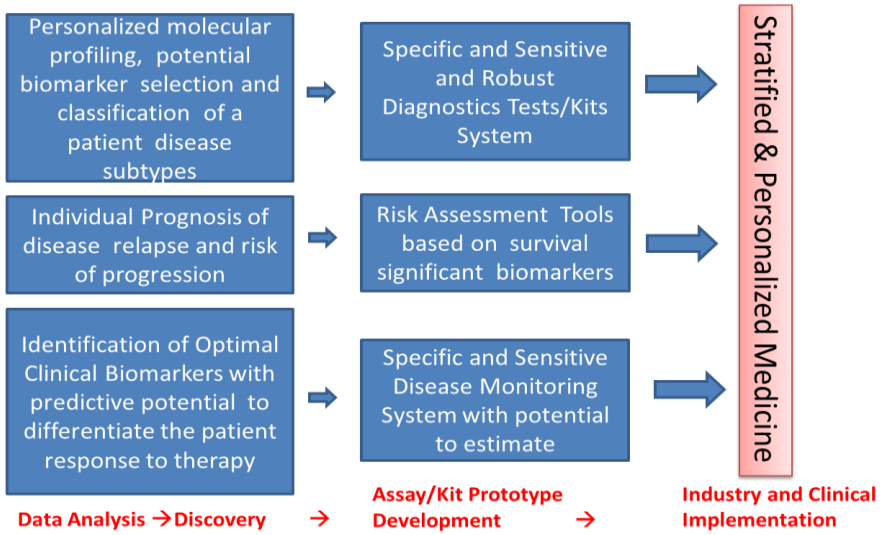
BII Cancer Biomarker Program

Integrative Bioinformatics, Computational Genomics, Systems Biology Medicine for Clinical Cancer Research



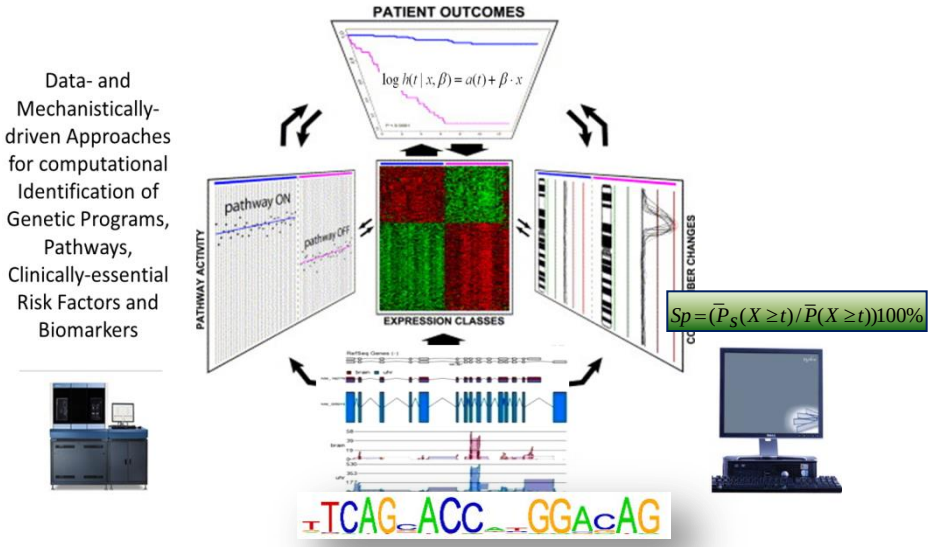
Kuznetsov VA, Hwee Kuan Lee, Maurer-Stroh S., Molnár MJ, Pongor S, Eisenhaber B, Eisenhaber F. How bioinformatics influences health informatics: Usage of biomolecular sequences, expression profiles and automated microscopic image analyses for clinical needs and public health. *Health Information Science and Systems* Publication date 2012/10/17, 1:1

Integrative Genomics and Clinical Bioinformatics Strategy for Novel Biomarker Discovery and Stratified Medicine Acceleration



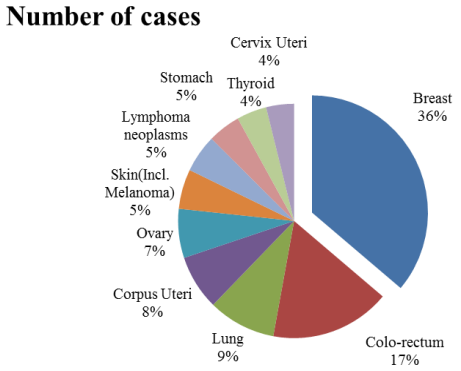
Computational Integration and Systems Analysis of Onco-Genomics and Clinical Data for Discovery of Biomarkers

Data- and Mechanistically-driven Approaches for computational Identification of Genetic Programs, Pathways, Clinically-essential Risk Factors and Biomarkers



Breast Cancer &Genetic Grades

Breast cancer (BC) is the most common cancer in women worldwide, with nearly 1.7 million new cases diagnosed per year.
Invasive ductal carcinoma is a major histologic type of BC.
70-80% of diagnosed BC are **Invasive ductal carcinoma** (IDC).

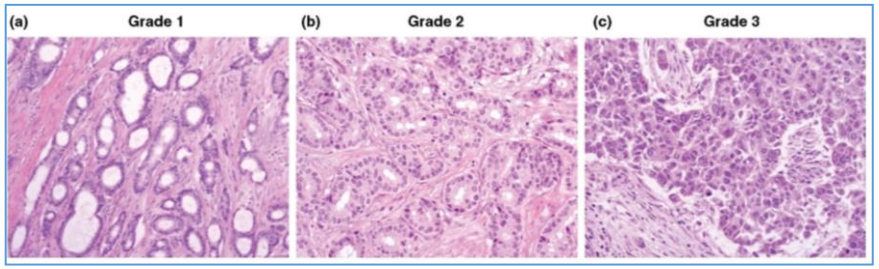


Top 10 cancers affecting Singapore women

Breast Cancer Grading Classification

Tumor histologic grading (HG) system reflects the growth rate (by mitotic index) and tissue structural disorder and invasiveness. HG is used for prognosis of BC and treatment assignment.

This classification assumes a linear cancer progression model: HG1 →HG2→HG3



Limitations: 55-85% inter-observer agreement; Grade 2 lacks prognostic power; Genetic basis of H2 tumors is poorly understood. Thus, a large proportion of tumors are characterized as intermediate-grade(HG2), making determination of optimal treatments difficult.

Histologic Grade: G=T+N+M

	Tubule Formation (% of Carcinoma Composed of Tubular Structures)	Score
1. T	> 75%	1
	10-75%	2
	less than 10%	3
	Nuclear Pleomorphism (Change in Cells)	**Score**
2. N	Small, uniform cells	1
	Moderate increase in size and variation 2	2
	Marked variation	3
	Mitosis Count (Cell Division)	**Score**
3. M	Up to 7	1
	8 to 14	2
	15 or more	3

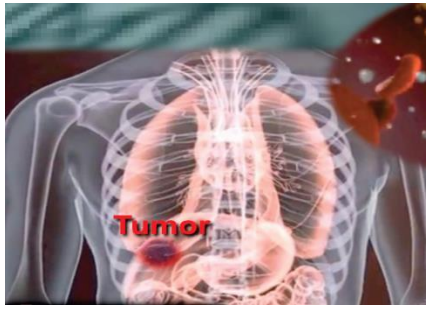
Classification	Immunoprofile	Other characteristics
Luminal A	ER^+, $PGR^{+/-}$, $HER2^-$	Ki67 low, endocrine responsive, often chemotherapy responsive
Luminal B	ER^+, $PGR^{+/-}$, $HER2^+$	Ki67 high, usually endocrine responsive, variable to chemotherapy. HER2+ are trastusumab responsive
Basal	ER^-, PGR^-, $HER2^-$	EGFR+ and/or cytokeratin 5/6+, Ki67 high, endocrine nonresponsive, often chemotherapy responsive
Claudin-low	ER^-, PGR^-, $HER2^-$	Ki67, E-cadherin, claudin-3, claudinin-4 and claudinin-7 low. Intermediate response to chemotherapy
HER2	ER^-, PGR^-, $HER2^+$	Ki67 high, trastusumab responsive, chemotherapy responsive

Goals:

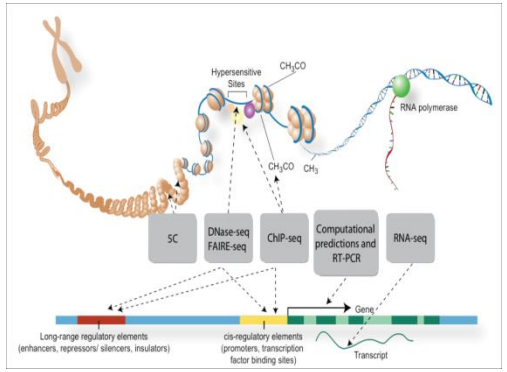
- develop prognostic and risk of cancer recurrence stratification tools

- discover clinical biomarkers

Clinical data



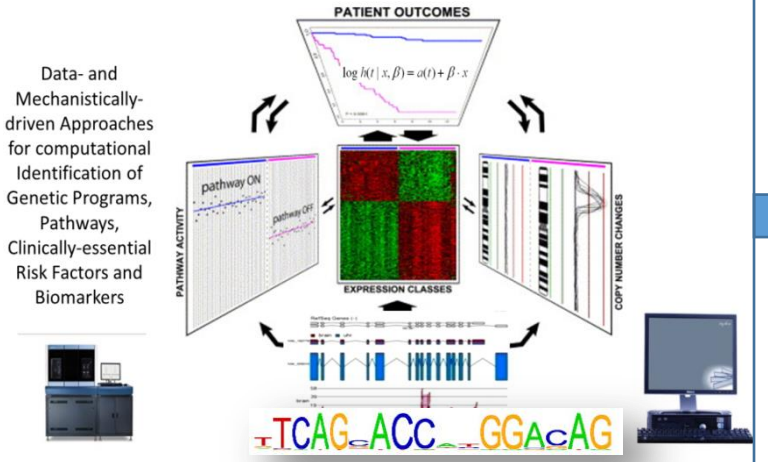
Clinical Omics Data



Data analysis : feature & pathway selection, disease modeling and prediction

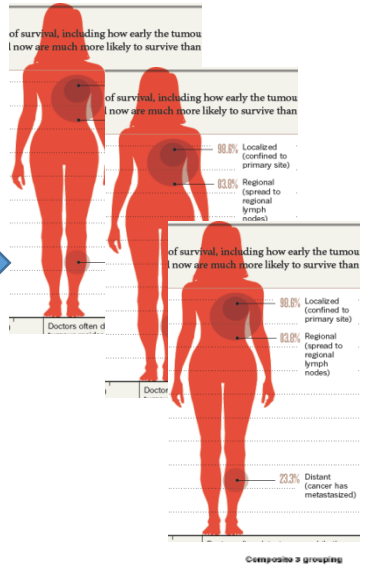
Computational Integration and Systems Analysis of Onco-Genomics and Clinical Data for Discovery of Biomarkers

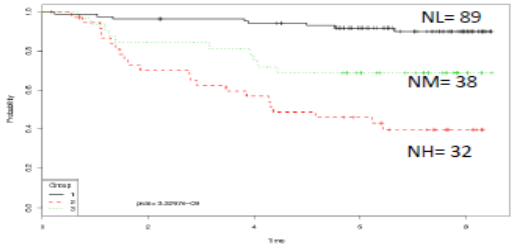
Data- and Mechanistically-driven Approaches for computational Identification of Genetic Programs, Pathways, Clinically-essential Risk Factors and Biomarkers



$$Sp = (\bar{P}_S(X \geq t) / \bar{P}(X \geq t))100\%$$

Patient 's stratification & prognosis





Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J, Lindahl T, Pawitan Y, Hall P, Nordgren H, Wong JE, Liu ET, Bergh J, Kuznetsov VA, Miller LD. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res. 2006 Nov 1;66(21):10292-301; Motakis et al, (2009); IEEE Biomed . Engineering ; Three Patent Applications, 2006-2012.

Could microarray gene expression analysis answer?

Statistically Weighted Syndromes (SWS) Analysis*

Training phase

Step 1:
Conversation of the continue variables in to discrete variables:

Step 2: Feature combination selection via control of the informatively measure and partition boundary stabilization

Step 3:
Weighted voting function calculating, feature selection, contingency table

→ Validation, prediction

* V.A. Kuznetsov, Ivshina et al. (1996) *Math. & Comput. Modeling* , 23(6), 95-119; A.M. Jackson,…,and V.A. Kuznetsov (1998) *J. Urology* 159, 1054-1063; Kuznetsov et al, *Intern J. Comput. Sci.& Network Security*, 2006 6(12).

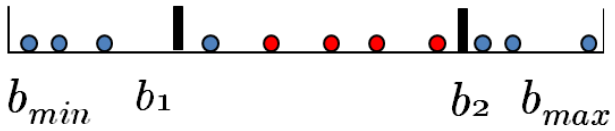
Major steps of feature selection by SWS method

Step 1:

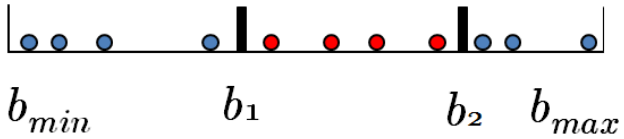
Construction of discrete variables from continue ones. Optimal partition of 1-D and 2-D permitted domains of values (e.g. expression signals) minimizing the number of erroneous classification of the objects in training set based on moving boundaries and maximization of Ch-2 function (or minimization of average entropy function).

Partition of 1-D variable

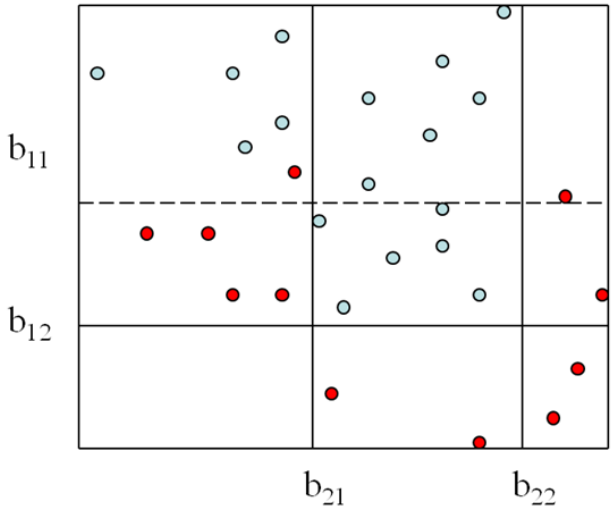
Correct number of boundaries, but incorrect localization of boundary b_1



Correct number of boundaries and correct boundary locations



Partition of 2-D variable



Step 2:

Due to permutation analysis, removing "low-informative" and "instable" discrete variables and selecting "highly-informative" and "robust" 1-D and 2-D discrete variables (called predictors or syndromes).

Step 3:

Calculating the "voting" function of each selected predictors and providing statistical weights for each "syndrome" based on its statistical significance in used training set .

kuznetsov@copyright

Research Article

Genetic Reclassification of Histologic Grade Delineates New Clinical Subtypes of Breast Cancer

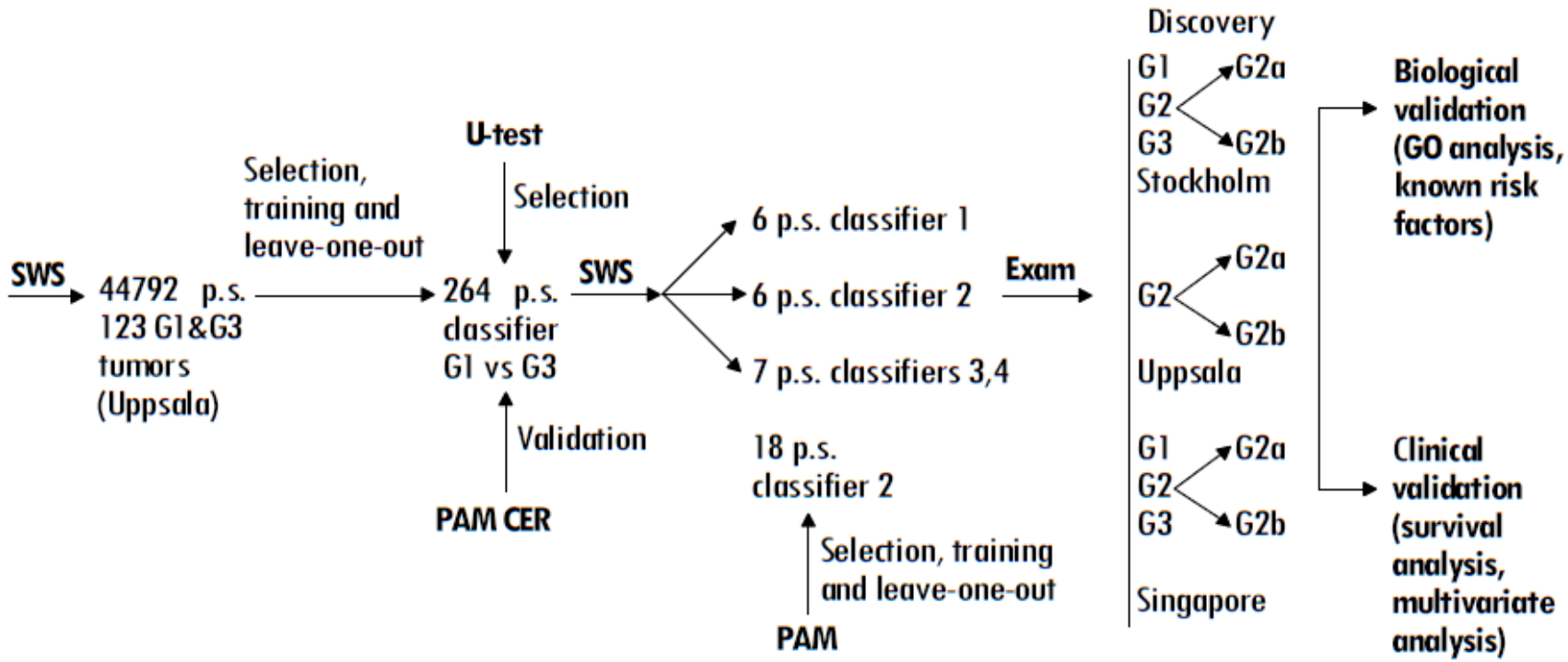
Anna V. Ivshina,[1] Joshy George,[1] Oleg Senko,[5] Benjamin Mow,[2] Thomas C. Putti,[3] Johanna Smeds,[6] Thomas Lindahl,[6] Yudi Pawitan,[7] Per Hall,[7] Hans Nordgren,[8] John E.L. Wong,[2,4] Edison T. Liu,[1] Jonas Bergh,[6] Vladimir A. Kuznetsov,[1] and Lance D. Miller[1]

Table 1. Affymetrix probe sets comprising the SWS and PAM genetic grade signatures

Unigene name (build 186)	Gene symbol	Genbank accession no.	Affy ID
Barren homologue (*Drosophila*)*	***BRRN1***	**D38553**	**A.212949_at**
Cell division cycle associated 8	*CDCA8*	BC001651	A.221520_s_at
V-myb myeloblastosis viral oncogene homologue (avian)-like 2	*MYBL2*	NM_002466	A.201710_at
Hypothetical protein FLJ11029	***FLJ11029***	**BG165011**	**B.228273_at**
FBJ murine osteosarcoma viral oncogene homologue B †	*FOSB*	NM_006732	A.202768_at
Chromosome 6 open reading frame 173	***C6orf173***	**BG492359**	**B.226936_at**
Serine/threonine-protein kinase 6	***STK6***	**BC027464**	**A.208079_s_at**
Anillin, actin binding protein (scraps homologue, *Drosophila*)	*ANLN*	AK023208	B.222608_s_at
Centromere protein E, 312 kDa	*CENPE*	NM_001813	A.205046_at
TTK protein kinase	*TTK*	NM_003318	A.204822_at
Signal peptide, CUB domain, EGF-like 2 †	*SCUBE2*	AI424243	A.219197_s_at
V-fos FBJ murine osteosarcoma viral oncogene homologue †	*FOS*	BC004490	A.209189_at
TPX2, microtubule-associated protein homologue (*Xenopus laevis*)	*TPX2*	AF098158	A.210052_s_at
Kinetochore protein Spc24	*Spc24*	AI469788	B.235572_at
Forkhead box M1	*FOXM1*	NM_021953	A.202580_x_at
Maternal embryonic leucine zipper kinase	***MELK***	**NM_014791**	**A.204825_at**
Cell division cycle associated 5	*CDCA5*	BE614410	B.224753_at
Cell division cycle associated 3	*CDCA3*	NM_031299	A.221436_s_at
Serine/threonine-protein kinase 6 ‡	***STK6***	**NM_003600**	**A.204092_s_at**

*Probe sets in boldface comprise the SWS signature.
† Higher expression in GI tumors.
‡ All probe sets comprise the PAM signature, except STK6 (A.204092_s_at).

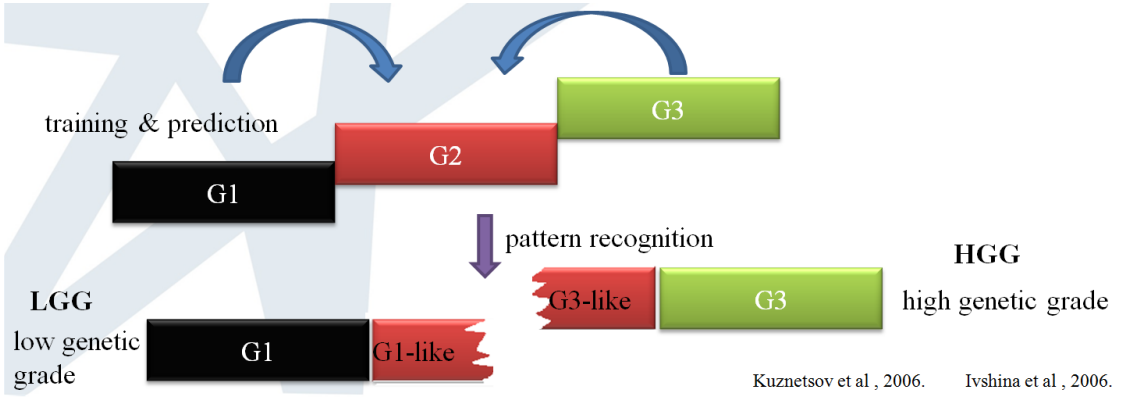
Schema of discovery and validation of the genetic G2a and G2b breast cancer groups



SWS: Statistically Weighted Syndromes method; PAM: Prediction Analysis for microarray method; CER: Class Error Rate plot; p.s. probe set: G1: grade 1; G3: grade 2; G3: grade 3; G2a: grade G2a; G2b: grade 2b; GO: gene ontology

Parallel cancer progression model

Concept of independent Low Genetic Grade(LGG) and High Genetic Grade (HGG) Tumors (Ivshina et al, 2005,2006)



Using unsupervised grouping of HG2 we demonstrated

→ $G1 \approx G1\text{-like} \neq G3\text{-like} \approx G3$

Low Genetic Grades | High Genetic Grades

LGG=G1+G1-like | HGG=G3+G3-like

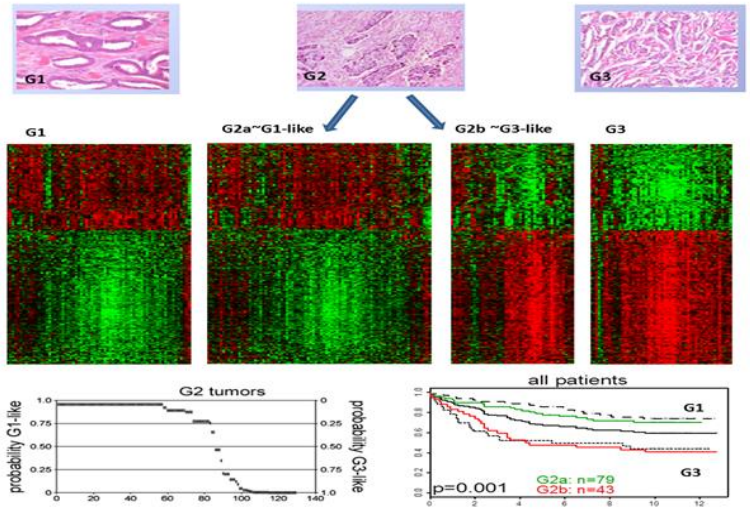
Our hypotheses : HG2 does not exist as a molecular entity; Low grade (LGG) and high grade (HGG) are major independent molecular and clinical IDC classes. LGG and HGG represent distinct tumor clinical classes (Ivshina et al, 2006)

TAG as a basis system for molecular classification and prognosis

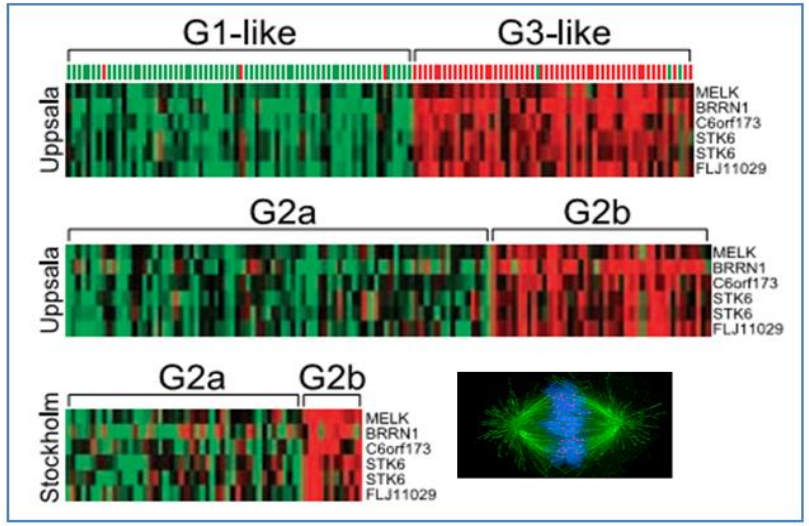
264 Affymetrix probesets (232 gene) basic TAG signature

264 Affymetrix probesets TAG signature

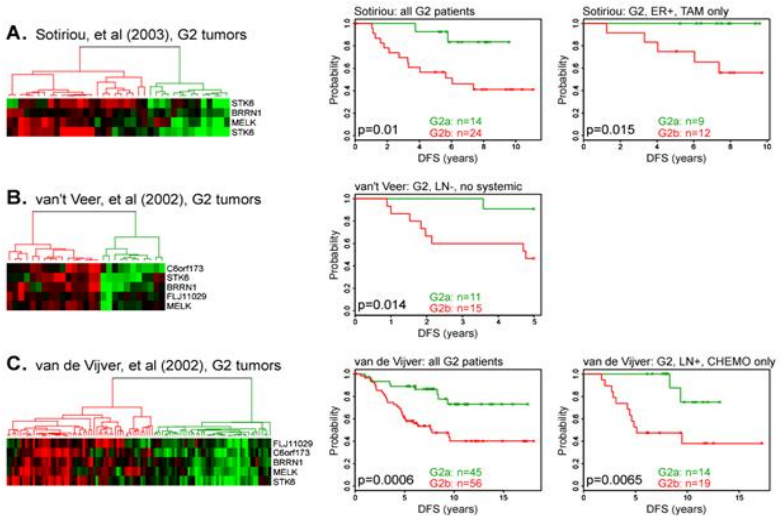
Discovery of Breast Tumor Aggressiveness Grading (BTAG)



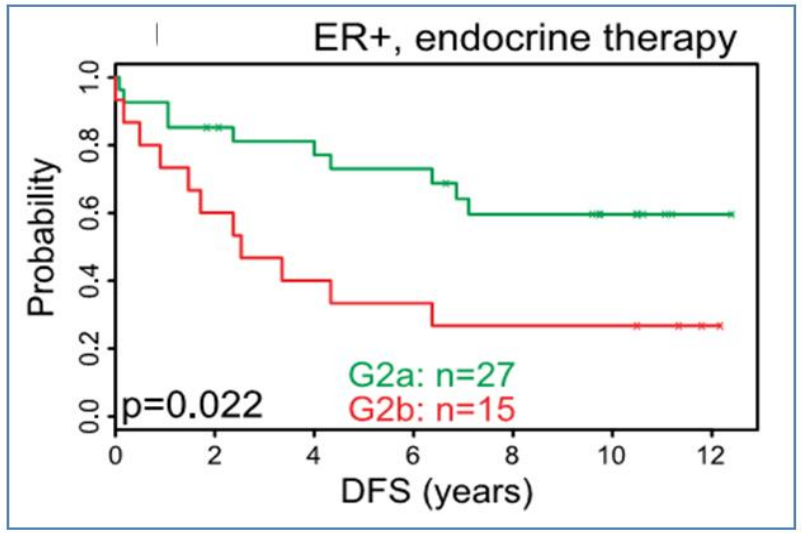
5-gene TAG signature Identification



5-gene TAG signature cross-cohort validation



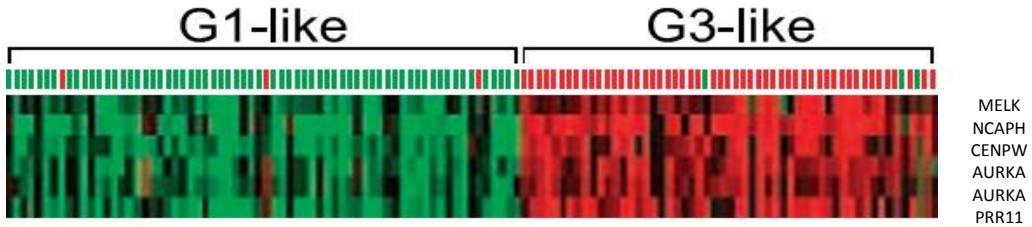
G2 Patient Stratification and Treatment Prediction



Technical and clinical distinction of grades 2 subclasses.

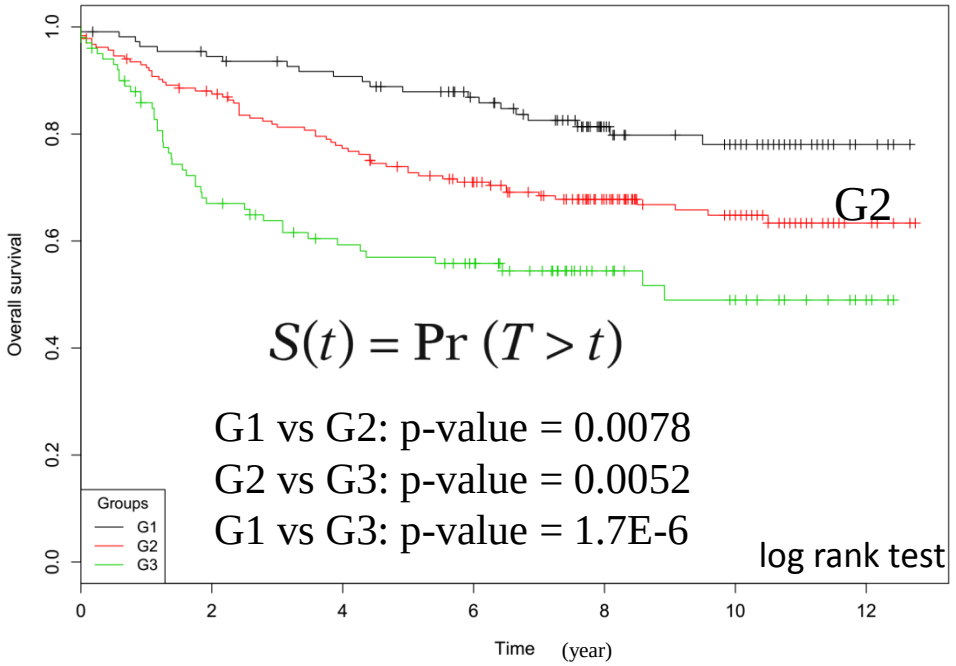
Grade	G1-like	G2	G3-like	Total
No. of patients	38(30.4%)	4 (3.2 %)	83 (66.4 %)	125 patients
Pro.	Pr > 0.75	0.75<Pr>0.25	Pr > 0.75	---

>96% of G2 samples assigned to the low or the high grade with >0.75 probability

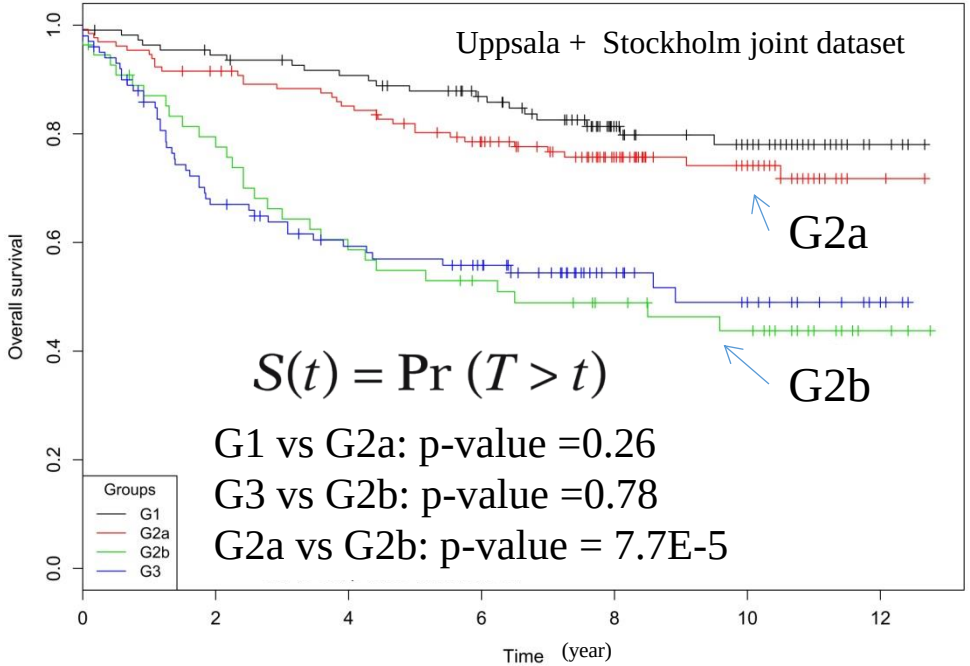


This distinction is supported further by survival analysis

Kaplan–Meier survival function

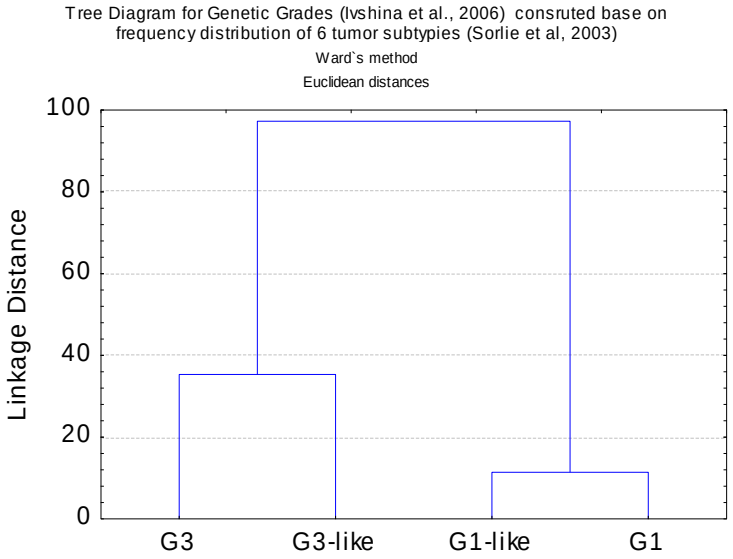


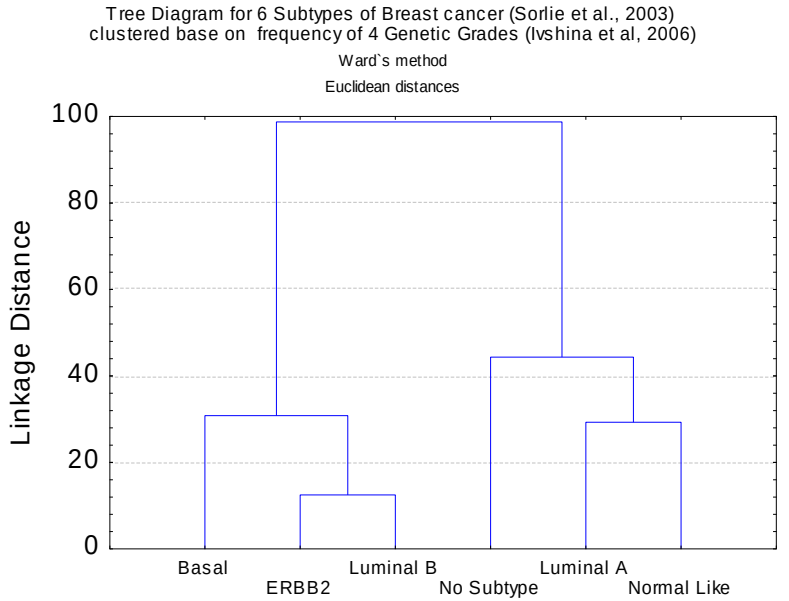
Kaplan–Meier survival function



Ivshina et al , 2006.

Dual Classification: Genetic Grade (SWS 5-genes signature) vs 6 Intrinsic Subtypes of Breast Cancer Classifier (465 UniGene Clusters) [Uppsala cohort]



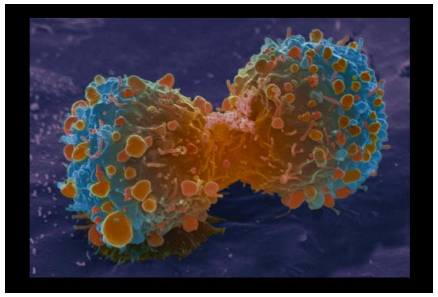


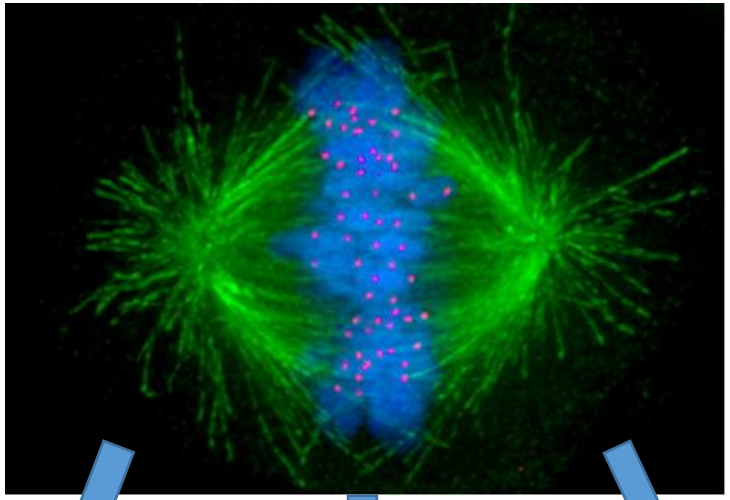
Subtypes (Uppsala)	G1	G1-like	G3-like	G3	Total
Normal Like (HRT)	26	24	1	1	*52*
No Subtype	7	13	1	1	*22*
Luminal A	32	33	12	6	*83*
Luminal B (AT-)	1	5	15	10	*31*
ERBB2 (AT-)	2	3	10	13	*28*
Basal (AT+)	2	1	7	24	*34*
Total	*70*	*79*	*46*	*55*	*250*

Subtypes (Uppsala)	G1	G1-like	G3-like	G3
Normal Like (HRT)	37.1	30.4	2.2	1.8
No Subtype	10.0	16.5	2.2	1.8
Luminal A	45.7	41.8	26.1	10.9
Luminal B (AT-)	1.4	6.3	32.6	18.2
ERBB2 (AT-)	2.9	3.8	21.7	23.6
Basal (AT+)	2.9	1.3	15.2	43.6
Total	*100.0*	*100*	*100.0*	*100.0*

AT: adjuvant therapy (CMF); HRT: Hormone Replace Therapy

Biological characteristics of Tumor Grading Signature(TAG)





centrosome duplication
chromosome segregation

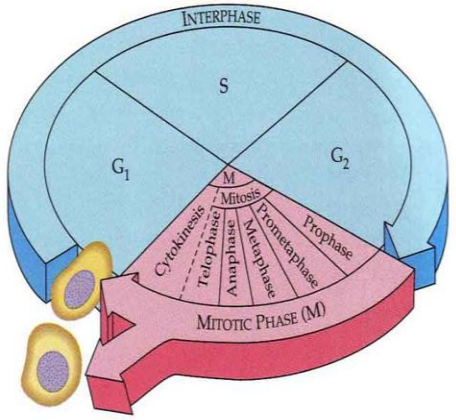
spindle assembly checkpoint

chromosomal alignment

Cell Cycle/Mitosis

Kinetohore cytokinesis

microtubule cytoskeleton



The mitotic phase is divided into two parts, **Mitosis** and **Cytokinesis**. Mitosis is the how the duplicated chromosomes are seperated. Cytokinesis is the division ofthe cytoplasm.

Patho-biological Interpretation of small TAG signature:

TAG signature shows a signature of actively transcribed chromatin and cell division

Goal

➢ **Generalization of independent progression hypothesis of HG1 and HG3 tumors to involve subclasses of HG2 tumors using integrative genomics datasets.**

Resources

The Cancer Genome Atlas (TCGA) Data

- biggest database for molecular profiles of different cancers

For breast cancer this DB includes:

- Comprehensive clinical data (include hist. grade data)
- Gene expression for IDC (n= 430 patients)
- Somatic DNA copy number (n= 430 patients)
- Somatic DNA point mutations (n= 418 patients)

Other datasets:

Experimental validation resource: **BII-OriGene** data. (n=84)

Luay Aswad

Ow Ghim Siong

Surya Pavan Yenamandra

Oleg Grinchuk

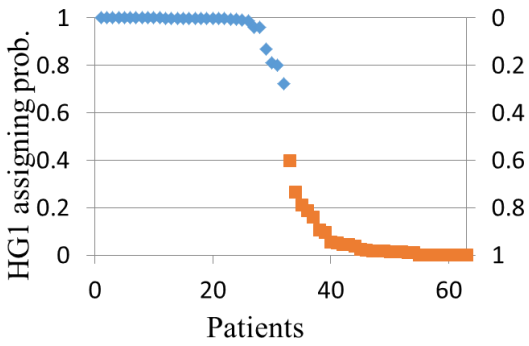
Anna IVSHINA

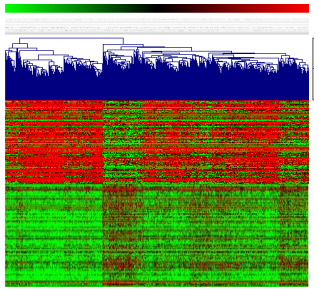
Aswad L, Yenamandra SP, Ow GS, Grinchuk O, Ivshina AV, Kuznetsov VA. Genome and transcriptome delineation of two major oncogenic pathways governing invasive ductal breast cancer development. **Oncotarget**. 2015 Oct 10. doi:10.18632/oncotarget.5543

Tools &Methods:

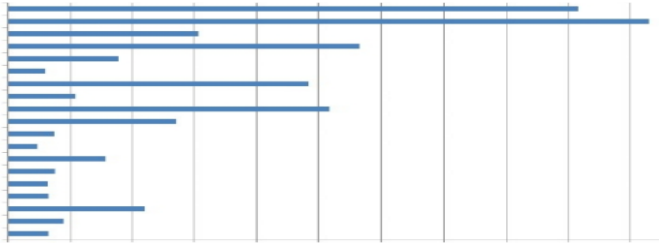
- Pattern recognition and feature selection:
  - Prediction analysis of Microarray (PAM) Modification of nearest-centroid method (Tibshirani R et. al. 2002)
  - Statistically Weighted Syndrome (SWS, Kuznetsov et al, 1996; Mathematical and computer modelling;
  - Ivshina et al, Cancer Res. 2006, Kuznetsov et al, 2006))
- Hierarchical clustering (HC)
- Functional enrichment analysis (GO analysis; David, MetaCore)
- DDg : Survival prediction analysis (E. Motakis, Ivshina, VA Kuznetsov. IEEE Engineering in Biology and Med, 2009).
- Biomarkers novelty analysis (Ghim Siong Ow, VA Kuznetsov. BMC Genomics VA, 2014)
- Multivariate analysis, Nonparametric statistical tests. Data subsampling method (Aswad L, Kuznetsov, 2015).
- Circular binary segmentation analysis were achieved to infer the amplified and deleted regions (diploid =1);

* Quantitative PCR validation.





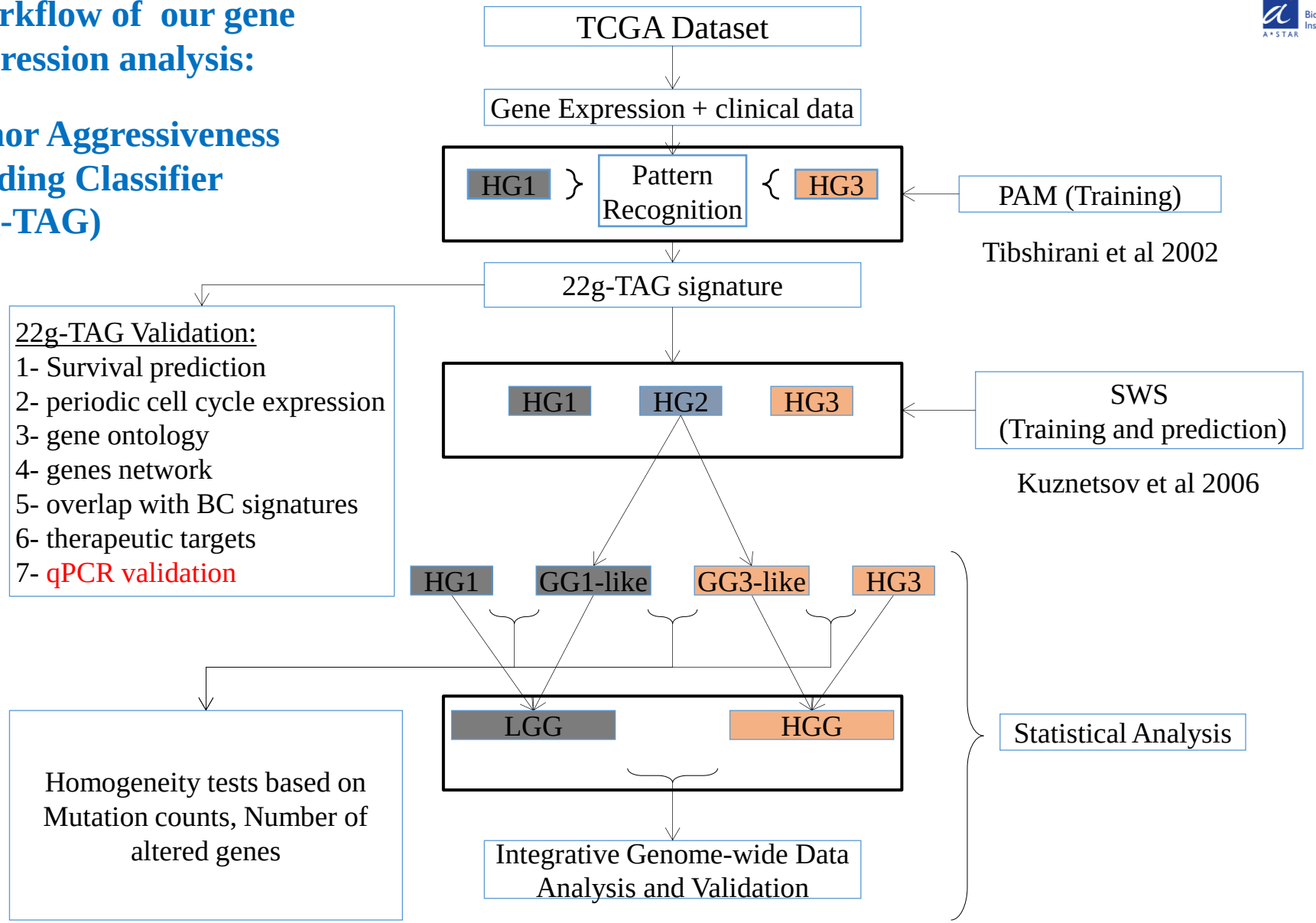
-log p-value



Bioinformatics Institute
A*STAR

Workflow of our gene expression analysis:

Tumor Aggressiveness Grading Classifier (22g-TAG)

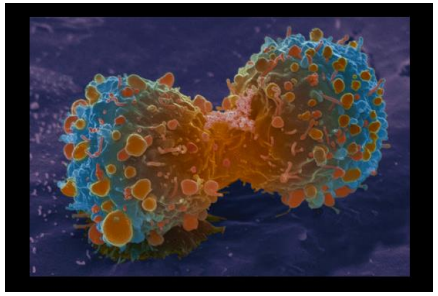


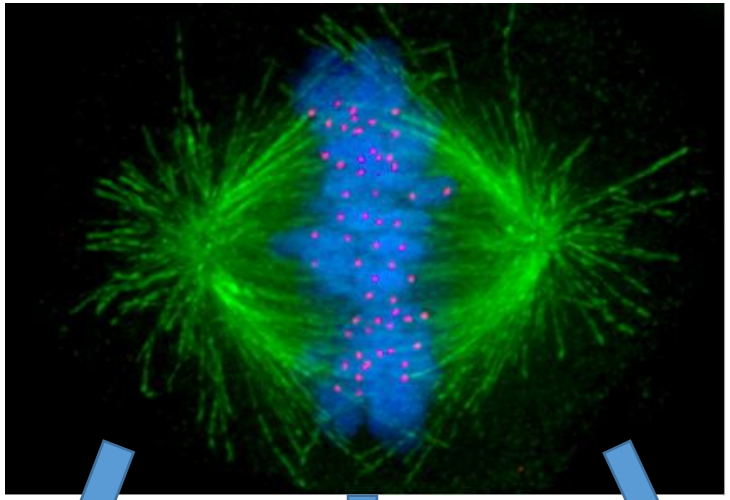
22g-TAG classifier of low grade (G1) and high grade (G3) tumors

# Probesets	Gene Symbol	Gene Name	FC	P-value	Freq. in 72 BC Sig.
1	***BUB1***	budding uninhibited by benzimidazoles 1 homolog (yeast)	3.43	3.82E-15	13
1	***CDC45***	CDC45 cell division cycle 45-like (S. cerevisiae)	3.7	1.16E-14	5
2	***CDCA5***	cell division cycle associated 5	>3.32	6.62E-15	10
2	***CDCA8***	cell division cycle associated 8	>3.12	6.62E-15	16
3	***CENPA***	centromere protein A	>3.57	2.87E-15	12
2	***CENPN***	centromere protein N	>3.71	9.75E-15	9
1	***FAM72A***	family with sequence similarity 72, member A,B	3.31	1.09E-14	6
1	***KIF14***	kinesin family member 14	3.51	1.66E-14	7
1	***KIF2C***	kinesin family member 2C	3.89	5.53E-15	13
3	***MCM10***	minichromosome maintenance complex component 10	>4.29	3.3E-15	8
1	***MELK***	maternal embryonic leucine zipper kinase	3.68	5.31E-15	20
1	***MTFR2***	family with sequence similarity 54, member A	3.47	8.82E-16	7
1	***MYBL2***	v-myb myeloblastosis viral oncogene homolog (avian)-like 2	4.56	2.57E-16	18
3	***ORC6***	origin recognition complex, subunit 6 like (yeast)	>3.8	2.75E-15	**1**
1	***PIF1***	PIF1 5'-to-3' DNA helicase homolog (S. cerevisiae)	3.79	8.86E-15	**1**
2	***SHCBP1***	SHC SH2-domain binding protein 1	>3.32	4.55E-16	4
4	***TICRR***	TOPBP1-Interacting Checkpoint And Replication Regulator	>3.12	1.8E-16	3
1	***UBE2C***	ubiquitin-conjugating enzyme E2C	4.7	2.25E-13	13
1	***NOSTRIN***	nitric oxide synthase trafficker	0.18	4.3E-14	6
3	***CAPN8***	calpain 8	<0.12	9.48E-13	**0**
2	***KIF13B***	kinesin family member 13B	<0.27	4.21E-15	7
2	***NAT1***	N-acetyltransferase 1 (arylamine N-acetyltransferase)	<0.04	1.75E-12	11

A*STAR Bioinformatics Institute

Biological characteristics of Tumor Grading Signature(TAG)





spindle assembly checkpoint

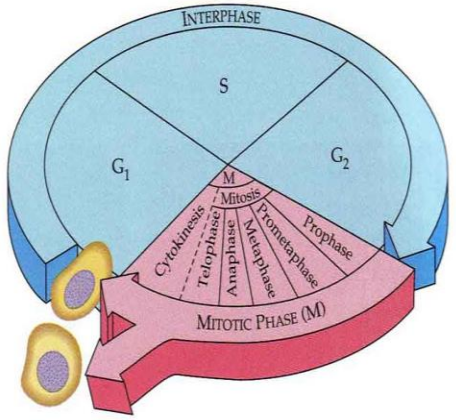
centrosome duplication
chromosome segregation

chromosomal alignment

Cell Cycle/Mitosis

Kinetohore cytokinesis

microtubule cytoskeleton

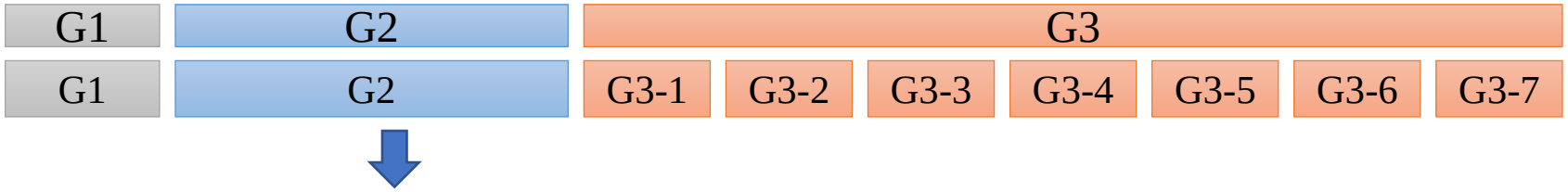


The mitotic phase is divided into two parts, **Mitosis** and **Cytokinesis**. Mitosis is the how the duplicated chromosomes are seperated. Cytokinesis is the division ofthe cytoplasm.

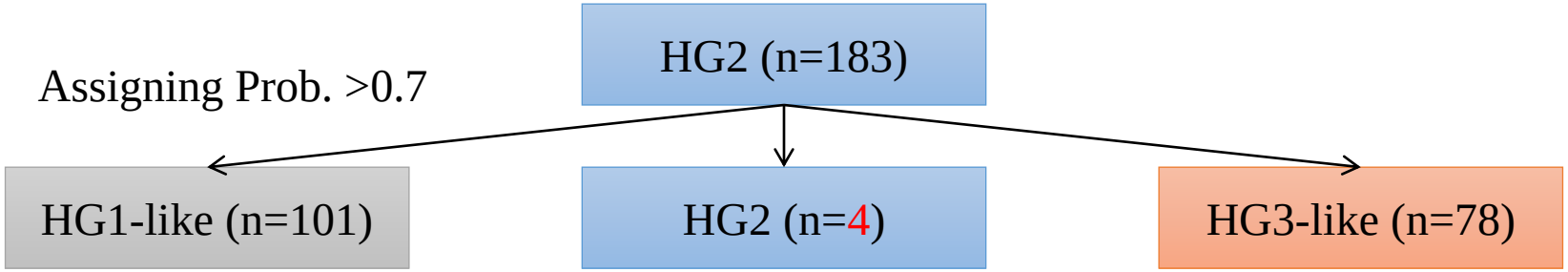
Patho-biological Interpretation of small TAG signature:

TAG signature shows a signature of actively transcribed chromatin and cell division

22g-TAG based sub-classification of HG2 using SWS algorithm:

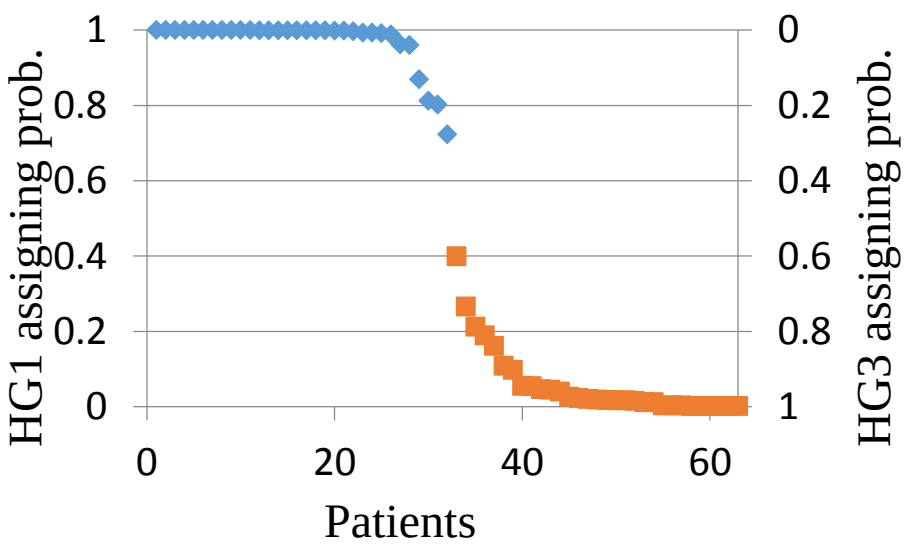


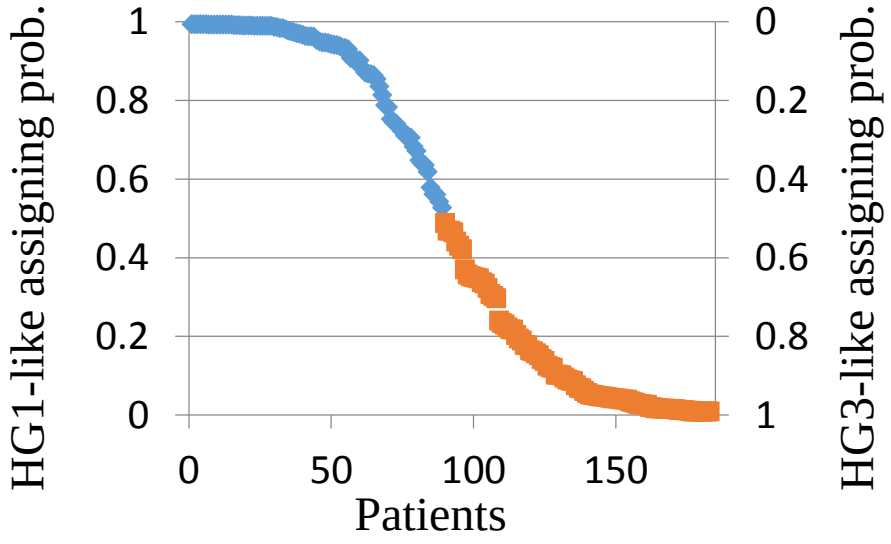
HG2 samples as class prediction set using 39 probes (22 genes) using results of Statistical weighted Syndrome (SWS) classification algorithm



55.2% **2.2%** 42.6%

The small number of samples assigned as HG2 demonstrates the lack of intermediate grades





Multivariate analysis:
22g-TAG signature outperforms other clinical parameters for patients survival in univariate analysis (independent datasets)

Uppsala cohort n=249

	Hazard Ratio	p-value
Tumor Size	1.016	0.002
Age	0.997	0.722
ER	0.858	0.623
PgR	0.685	0.172
LN	2.109	0.00047
22g-TAG	**2.997**	**7.E-06**
Hist. Grades	1.773	0.00024
HER2	1.313	0.302

Stockholm cohort n=147

	Hazard Ratio	p-value
Tumor Size	1.013	0.25
Age	0.994	0.667
ER	0.594	0.172
PgR	0.484	0.027
LN	1.028	0.934
22g-TAG	**5.883**	**0.00038**
Hist. Grades	2.114	0.003

Singapore cohort n=88

	Hazard Ratio	p-value
Tumor Size	1.022	0.167
Age	1	0.994
ER	0.396	0.046
PgR	0.314	0.019
LN	2.647	0.049
22g-TAG	**4.683**	**0.008**
Hist. Grades	2.723	0.008
HER2	6.063	0.004

Marseille cohort n=248

	Hazard Ratio	p-value
Tumor Size	1.373	0.057
Age	1.002	0.796
ER	0.663	0.068
PgR	0.807	0.338
LN	1.567	0.051
22g-TAG	**2.575**	**0.001**
Hist. Grades	1.626	0.003
HER2	1.514	0.225
P53 mut.	1.721	0.031
ki67	1.507	0.145

Differentially Expressed Genes (DEG) between the 4 sub-classes

n=90797 probesets, FC<0.75 or FC<1.25 , Wilcoxon test FDR <0.01

HG1 (n=32) | HG1-like (n=101) | HG3-like (n=78) | HG3 (n=215)

0 probe sets

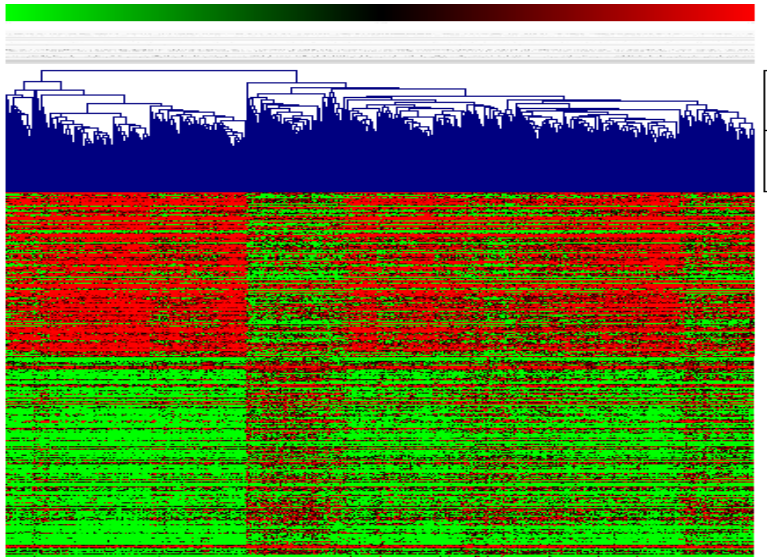
4933 probe sets
871 Down reg. genes in HG3-like
1250 Up reg. in genes HG3-like
enriched significantly in
18 + 145 GO biological process

1837 probe sets
416 Down reg. genes in HG3
389 Up reg. in genes HG3
there is no significant enrichment
in any GO biological process

Supportive evidence: Hierarchical Clustering:

Using the 4933 probes sets that were differentially expressed between HG1-like and HG3-like and using Euclidean distance and average link agglomerative method.

LGG and HGG grading classes are correlated with SWS based 22g-TAG classifier



4933 differentially expressed probe sequences

	Cluster 1	Cluster 2
HG1	24	8
HG1-like	80	21
HG3-like	7	71
HG3	25	190

Cohen's Kappa k= 0.67 ; p= 6.31×10^{-43}

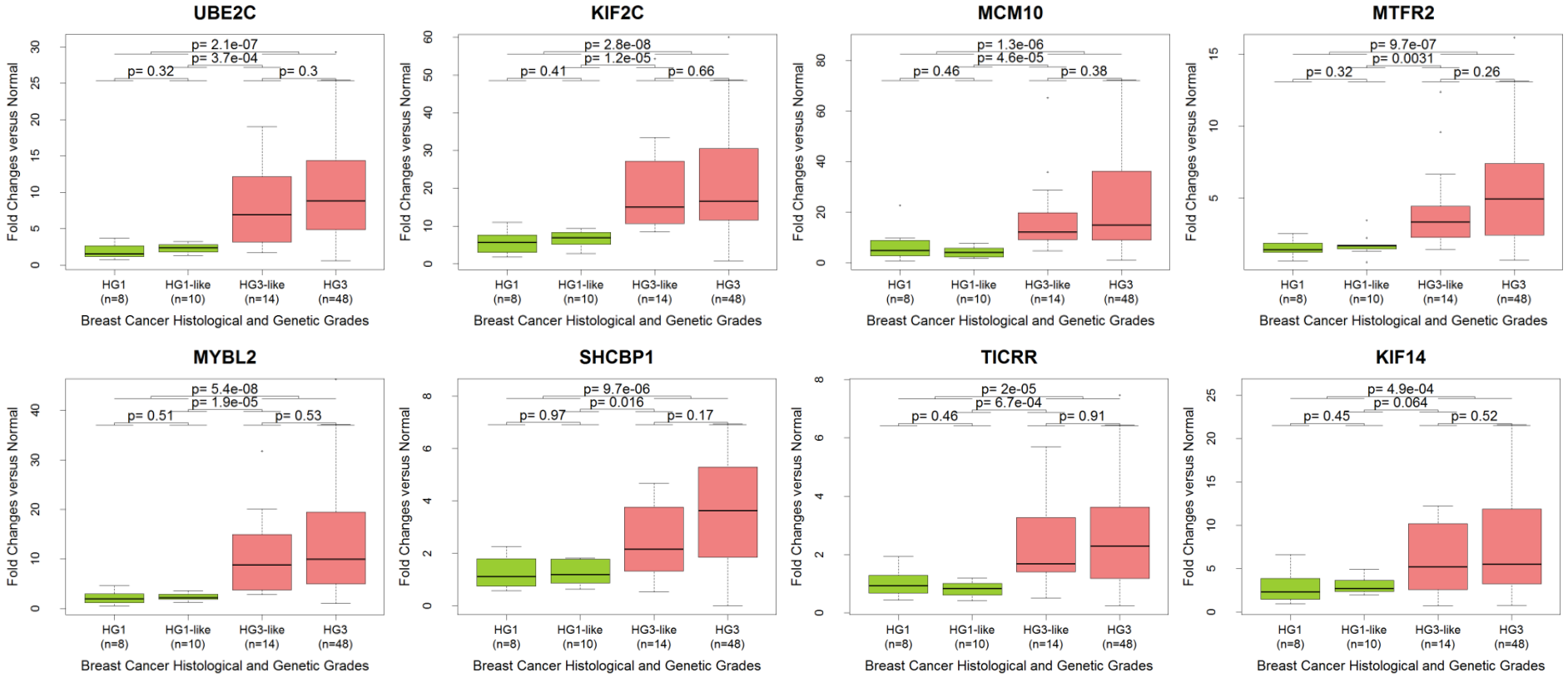
Using unsupervised clustering
→HG1 ≈ HG1-like ≠ HG3-like ≈ HG3

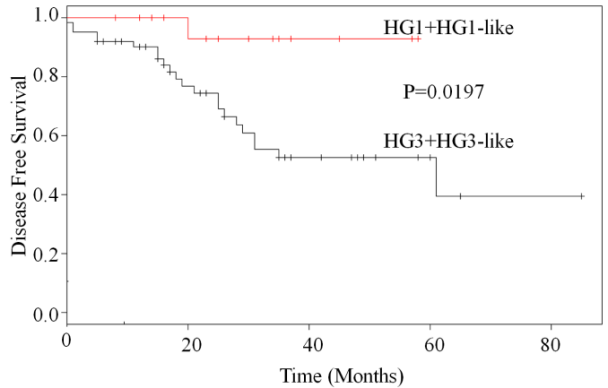
Low Genetic Grades
LGG

High Genetic Grades
HGG

Experimental Validation: qPCR-based validation of the 22g-TAG on 80 patients samples from OriGene

HG1=8, HG2=24, HG3=48 → Normalized Ct values were used for sub-classification of HG2 tumors using SWS. Average training accuracy is 83.3% (sensitivity: 66.6 ± 7.2%, specificity: 91.7± 3.6%).

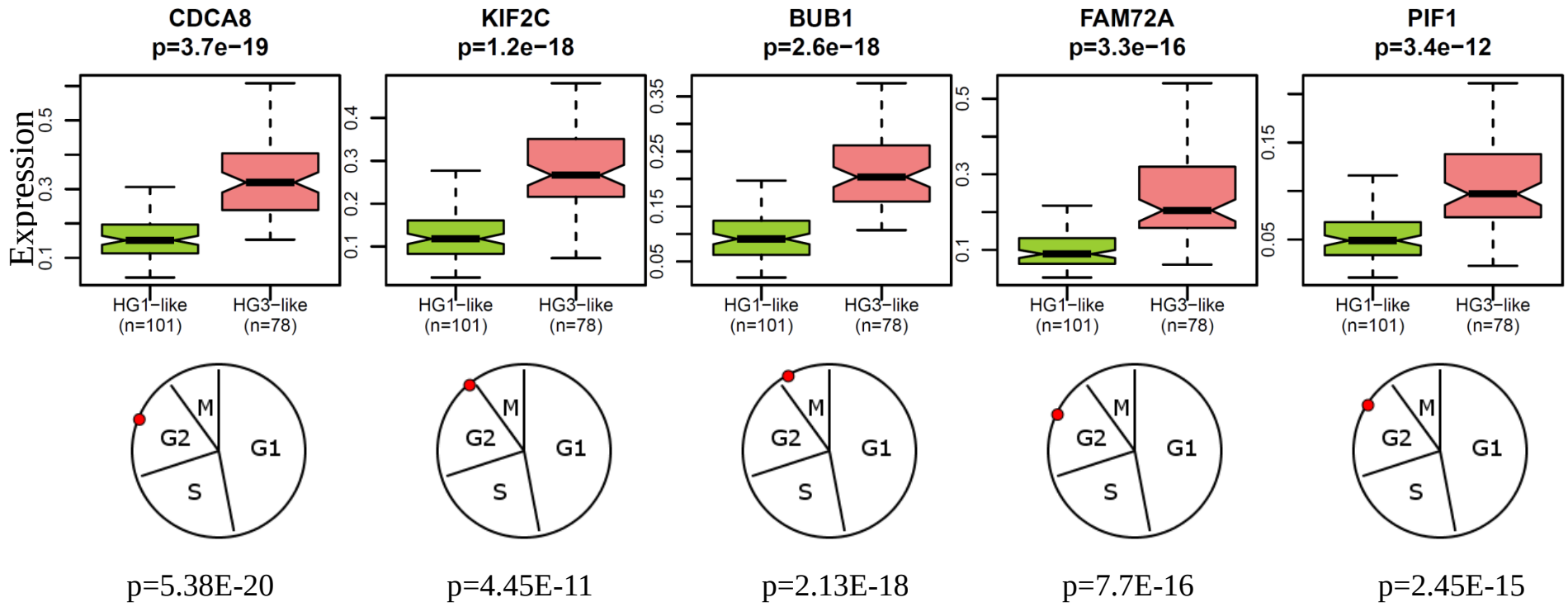




Conclusion: The 22g-TAG qPCR assay provides strong stratification of the BC patients onto low- and high- risk of disease classes. All microarray DIG data were fully confirmed.

18/22 genes of 22g-TAG have periodic expression through cell cycle

Maximum expression mostly at G/M Phase



p=5.38E-20 p=4.45E-11 p=2.13E-18 p=7.7E-16 p=2.45E-15

3 out of the 4 non-periodic genes are tumor suppresser-like genes.

22g-TAG:oncogenic (cell cycle; periodic) and tumor suppressor gene signature (non-peridoic)

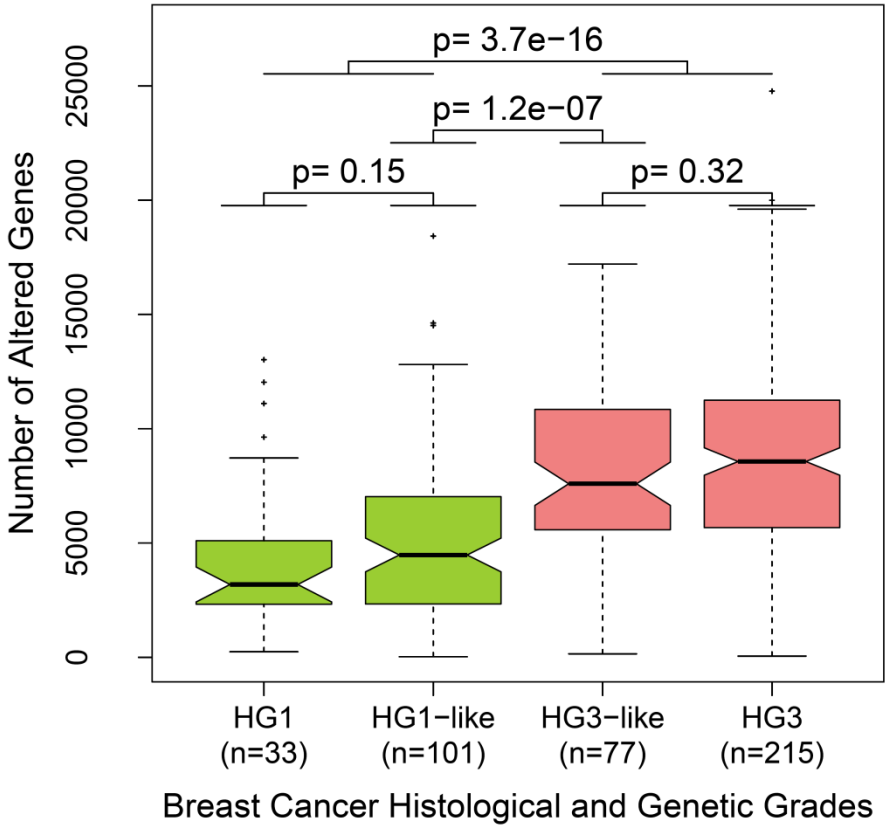
Dichotomization of genomic DNA CNV and SNP based on our 22g-TAG classifier

Bioinformatics Institute
A*STAR

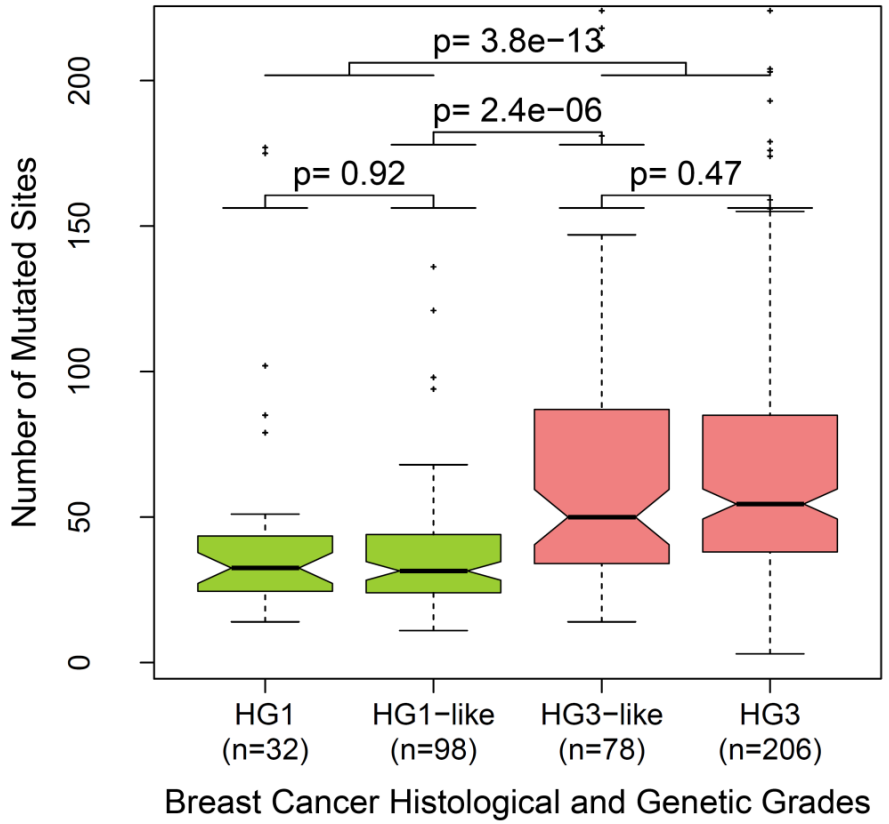
We analyzed DNA CNV data from 426 TCGA samples, and point mutation data from 414 TCGA samples that also contained gene expression data.

Counting the number of events per sample

DNA Copy Number Variation



Genome wide point mutations

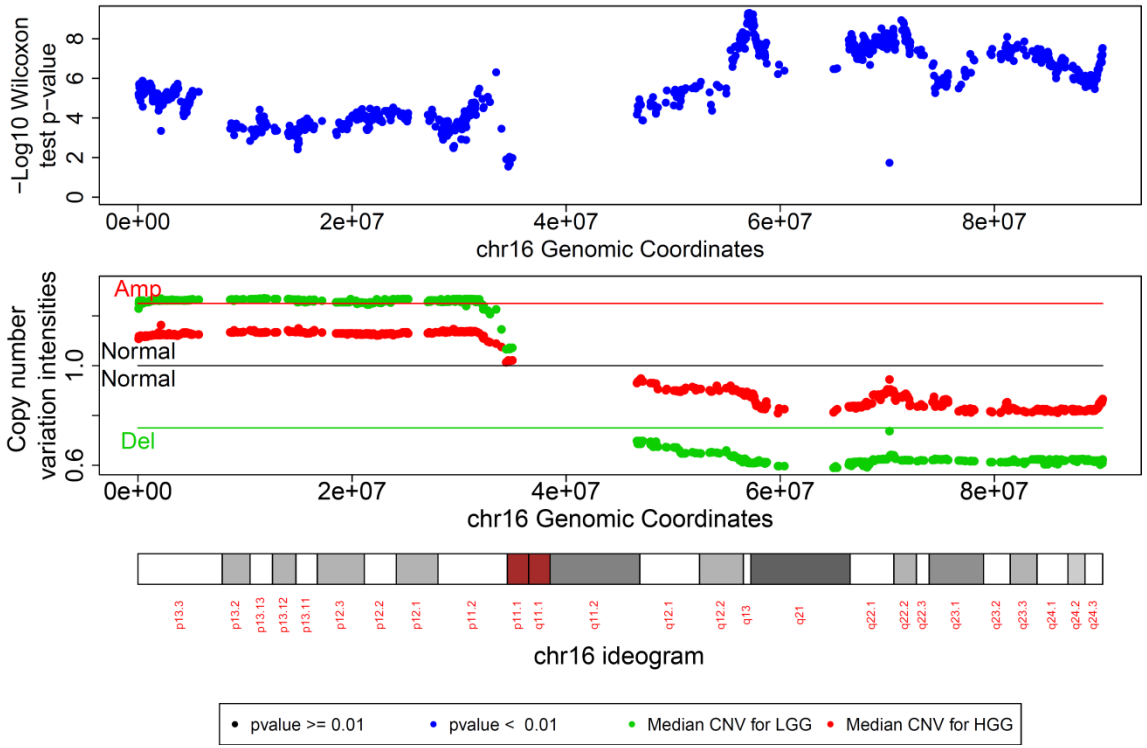


Conclusion: Two major classes of IDC tumors are reproduced at genomic level.

Why genome variation sub-classification of HG2 is important?

16q loss is the major event that differentiates between LGG and HGG tumors.

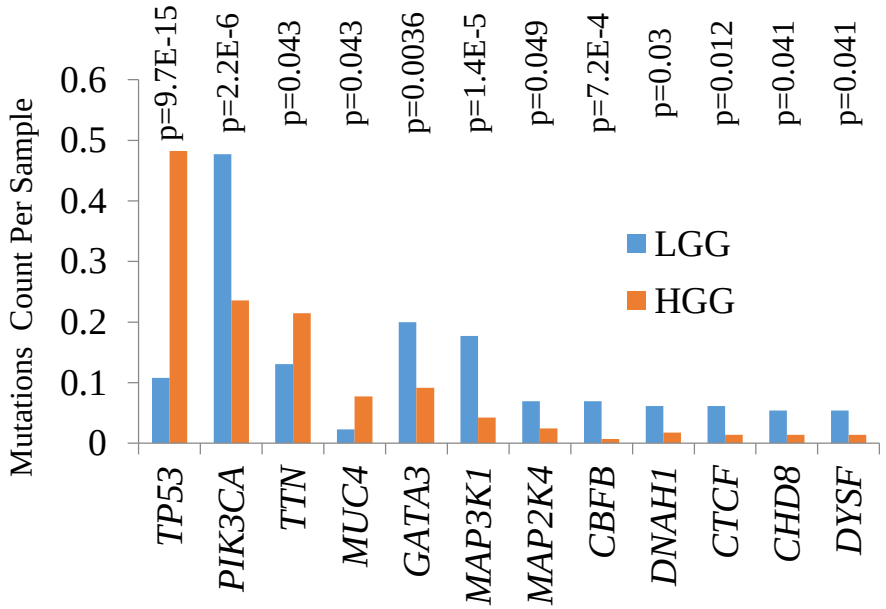
chromosome	Number of genes	chromosome bands
chr16	971	16p13.3-16p11.2, 16q11.216q24.3
chr17	192	17p13.2-17p11.2
chr8	694	8p23.3,8p23.1-8p12,8q12.2-8q24.3
Total	1858	



Conclusion:
Independent progression pathway can be generalized to LGG and HGG tumors

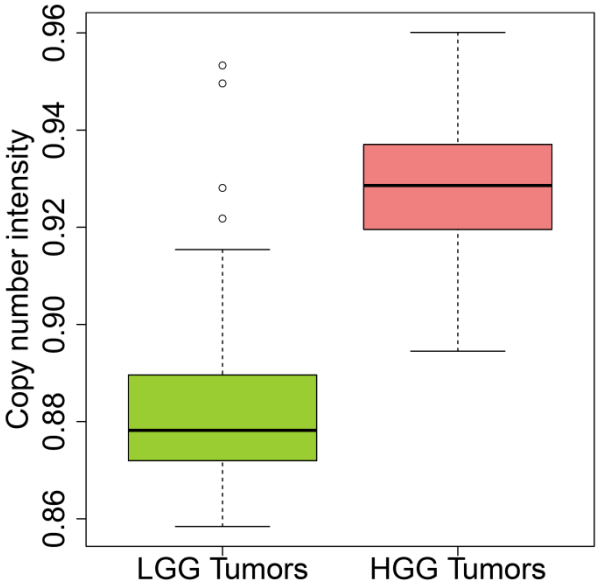
pvalue >= 0.01 pvalue < 0.01 Median CNV for LGG Median CNV for HGG

Mutation profiles



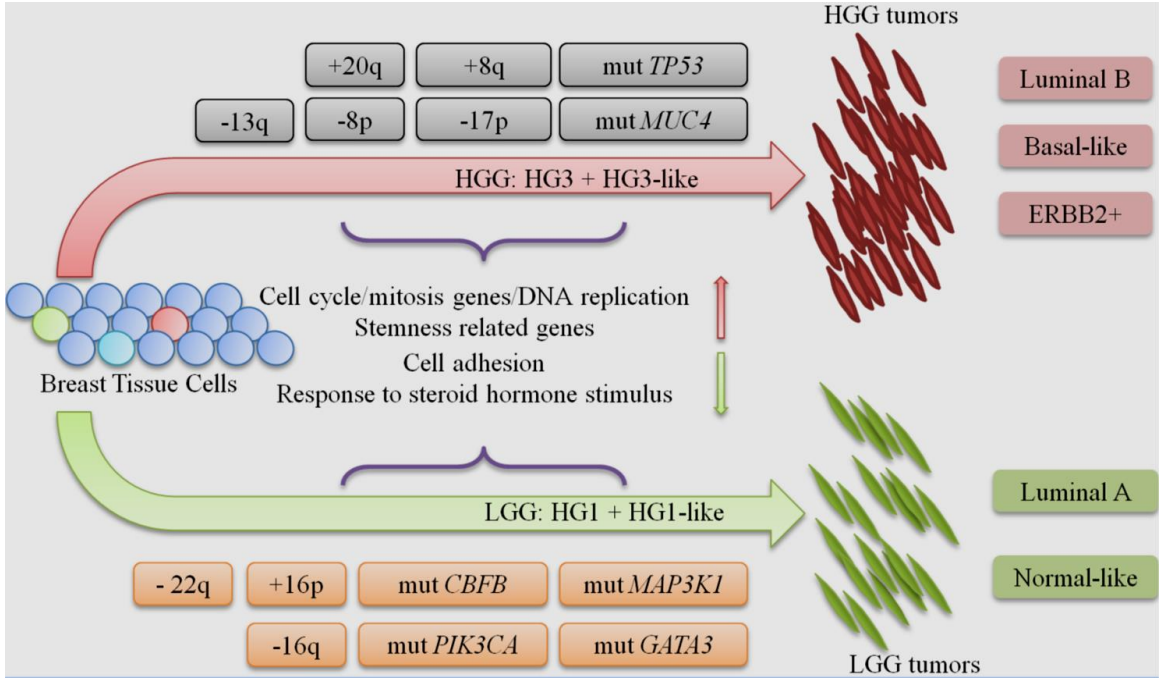
9/12 genes are highly mutated in **LGG tumor**
→ supports our model of the two independent progression pathways

Copy number intensity values



$P < 1.10^{-16}$

A genetic model of the two major oncogenic pathways of IDC



- Genetic characterization of two IDC oncogenic pathways associated with LGG and HGG tumor classes with improbable inter-grades tumor progression. Our genetic results strongly suggest that low-grade and high-grade tumors develop in patient populations independently.
- qPCR based panel of 22-gene TAG signature: Dichotomization of intermediate grade (HG2) IDC patients based on 22 oncogenic cell cycle genes and tumor suppressor factors.
- Narrowing clinically relevant BC biomarker space: → impact in diagnostics, patient prognosis, improvement of IDC therapeutics.
- A different expression patterns of precursors/cancer stem cells was observed in HGG vs LGG. They propose different origin for HGG and LGG tumors.
- For patients with breast conservation surgery(BCS), our grading classifier and associated genetic characteristics could be useful for assignment of the patients' eligibility to neo-adjuvant and immunotherapy.

Summary

TAG Signature is a highly sensitive, reproducible and robust Breast Cancer Classification and Prognostic System which is actually ready to translate it to real clinical needs

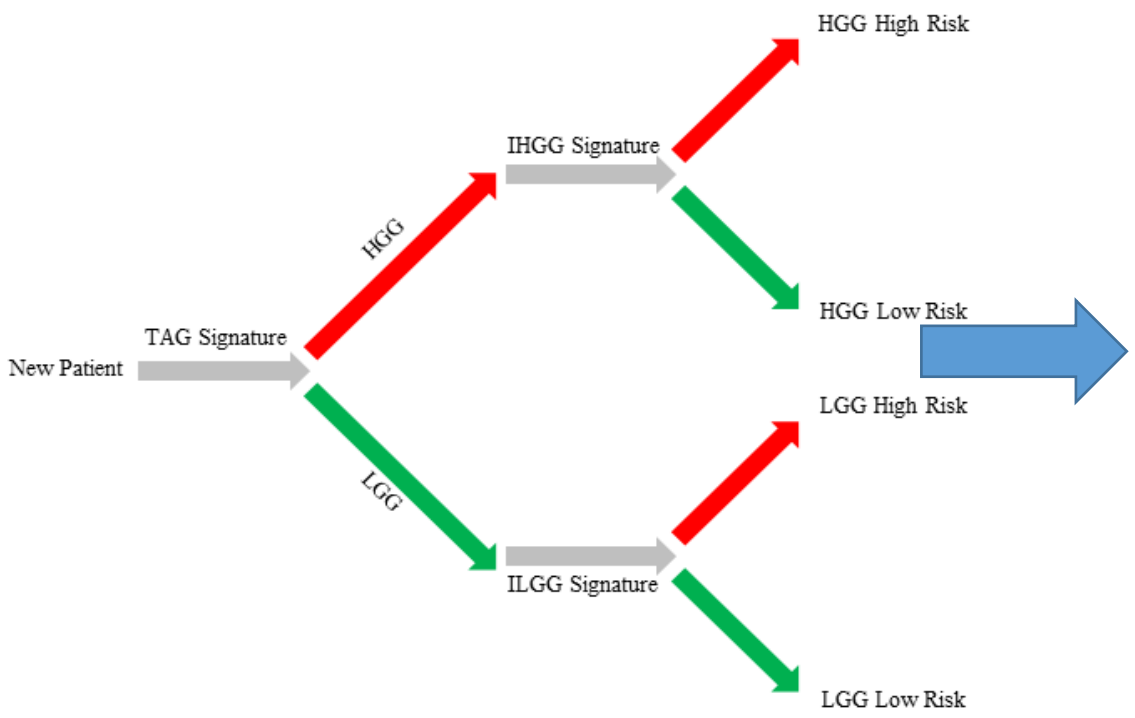
In detail experimentally studied, big clinical and genomics data-supported prognostic system with

- **high sensitivity**
- **high accuracy,**
- **Accurate disregarding tumor mass, LN or ER/Pr status**
- **Robust across cohort and ethnic groups**
- **Cross-platform reproducible**
- **Automatically supported by statistically-based personalized prognosis scoring**

New opportunities for personalized diagnostics, prognosis and treatment

Bioinformatics Institute
A*STAR

New tumor classes and subclasses



oncogenic pathways, clinical biomarkers, therapeutic targets, systemic therapeutic strategies

BC genetic grading system considers cancer cell origin, mutations, chromosome, gene expression alterations. It could be used for diagnostics/classification of oncogenic pathways governing BC classes, subclasses development and progression.

Publications-2015:

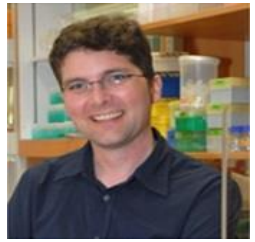
- **1**: Aswad L, Yenamandra SP, Ow GS, Grinchuk O, Ivshina AV, Kuznetsov VA#. Genome and transcriptome delineation of two major oncogenic pathways governing invas ductal breast cancer development. **Oncotarget**. 2015 Oct 10. doi:10.18632/oncotarget.5543.
- **2**: Jenjaroenpun P, Chew CS, Yong TP, Choowongkomon K, Thammasorn W, Kuznetsov VA.# The TTSMI database: a catalog of triplex target DNA sites associated with genes and regulatory elements in the human genome. **Nucleic Acids Res**. 2015; Jan; 43(Database issue):D110-6. doi: 10.1093/nar/gku970.
- **3**: Jenjaroenpun P, Wongsurawat T, Yenamandra SP, Kuznetsov VA# . QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. **Nucleic Acids Res**. 2015 Apr 16. pii: gkv344. [Epub ahead of print]
- **4**: Lim KW, Jenjaroenpun P, Low ZJ, Khong ZJ, Ng YS, Kuznetsov VA#, Phan AT#. Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. **Nucleic Acids Res**. 2015 May 9. pii: gkv355.
- **5**: Grinchuk O, Efthymios Motakis, Yenamandra SP, Ow GS, Jenjaroenpun P, Tang Zhiqun, A. Yarmishyn1, Ivshina AV, Kuznetsov VA#. Sense-antisense gene-pairs in breast cancer and associated pathological pathways. **Oncotarget** 2015.Oct 25; DOI: 10.18632/oncotarget.6255.
- **6:** Giannakakis A, Zhang J, Jenjaroenpun P, Nama S, Zainolabidin N, Aau MY, Yarmishyn AA, Vaz C, Ivshina AV, Grinchuk OV, Voorhoeve M, Vardy LA, Sampath P, Kuznetsov VA#, Kurochkin IV#, Guccione E#. Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. **Sci Reports**. 2015 May 29;5:9737. doi: 10.1038/srep09737.
- **7**: Ow GS and Kuznetsov VA.# Multiple signatures of a disease in potential biomarker space: Getting the signatures consensus and identification of novel biomarkers. **BMC Genomics** 2015, 16(Suppl 7):S2 doi:10.1186/1471-2164-16-S7-S2
- **8:** Goh WQJ, Ow GS, Kuznetsov VA, Chong S, Lim YP. DLAT subunit of the pyruvate dehydrogenase complex is up-regulated in gastric cancer-implications in cancer therapy. **Amer J Translational Res**. 2015;7(6):1140-1151.
- **9**:Nymoen DA, Hetland Falkenthal TE, Holth A, Ow GS, Ivshina AV, Tropé CG, Kuznetsov VA, Staff AC, Davidson B. Expression and clinical role of chemoresponse-associated genes in ovarian serous carcinoma. **Gynecol Oncol**. 2015 Jul 29. pii: S0090-8258(15)30091-3. doi:10.1016/j.ygyno.2015.07.107.
- **10:** Batagov OV, Yenamandra SP, Jenjaroenpun P, Ivshina AV, Kuznetsov VA.#2015. The key role of MECOM complex locus and its products in high-grade ovarian carcinoma pathogenesis and clinical outcomes. 2015. **Cancer Research** 75 (15 Supplement), 1990
- **11**: Ghim Siong OW, Zhiqun TANG and Vladimir A KUZNETSOV# Big data and computational biology strategy for personalized prognosis. **Oncotarget,** (accepted)

Acknowledgements:

My team

CHEW Chee Siang

PHAN Anh Tuân



Stephan Gasser

YONG Tai Pang

LIM Yoon Pin

Ernesto Guccione

Collaborators:

Martin Lavin

Abrey J. Yo

Larry Stanton

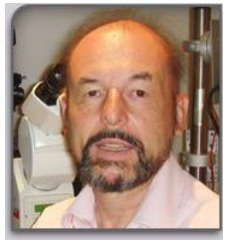
Igor Kurochkin.

Prabha Sampath

Ben Davidson

Wenyi Feng
SUNY Upstate Medical University

Pierce Chow
KH; NCCS



J.P. Thiery

This work was supported by funding from the Biomedical Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

Thank you!

*vladimirk@BII.a-star.edu.sg