Application of Genetic Programming in Analysis of Quantitative Gene Expression Profiles for Identification of Nodal Status in Bladder Cancer

Anirban P. Mitra, Arpit A. Almal, Ben George, David W. Fry, Peter F. Lenehan, Vincenzo Pagliarulo, Richard J. Cote, Ram H. Datar, William P. Worzel

University of Southern California, Los Angeles, CA; Genetics Squared, Inc., Ann Arbor, MI

BACKGROUND

➢ Urinary bladder cancer is the seventh most common cancer worldwide (3.2% of all cancers), with an estimated annual incidence of 330,000 new cases and 179,000 deaths each year (World Health Report 2004, WHO). Approximately 63,210 new cases of bladder cancer were expected in the United States in 2005 alone, with almost 13,190 deaths (Jemal et al, *CA Cancer J Clin* 2005).

➢ Nodal involvement is considered to be an independent risk factor for recurrence and survival after cystectomy for organ-confined bladder cancer (NCCN Practice Guidelines in Oncology - Bladder Cancer, Version 1.2006).

➢ Molecular changes in bladder cancer have been shown to precede morphologic changes that can be identified visually (Bonassi et al, *Mutat Res* 2001). Further, some tumors have specific molecular patterns that predispose them to be more morphologically aggressive, with a greater propensity to metastasize and recur, regardless of their clinical stage at diagnosis (Kawamukai et al, *Rays* 2004).

➢ Extensive prognostic studies on single markers have been performed in bladder cancer. However, our group has previously shown that combined analyses of multiple markers can be a better prognostic indicator than individual determinants (Chatterjee et al, *J Clin Oncol* 2004). Bladder cancer has a multifactorial etiology with distinct pathways contributing to its pathogenesis (Wu et al, *Nat Rev Cancer* 2005) which led to the genesis of this study in quantitatively investigating multiple markers and generating mathematical algorithms to determine nodal status.

STUDY COHORT

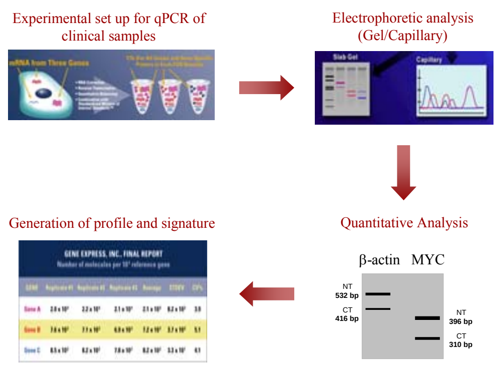
➢ Training set

TUMOR STAGE / NODAL STATUS	Normal Controls	Ta	T1	T2	T3	T4	Total
Node Positive		0	2	0	7	2	11
Node Negative	3	3	6	4	5	2	23
Total	3	3	8	4	12	4	34

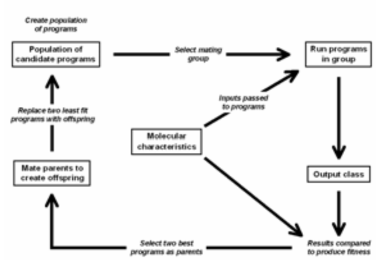
➢ Validation set

TUMOR STAGE / NODAL STATUS	Normal Controls	Ta	T1	T2	T3	T4	Total
Node Positive		0	1	0	7	2	10
Node Negative	2	7	4	4	3	1	21
Total	2	7	5	4	10	3	31

StaRT-PCR



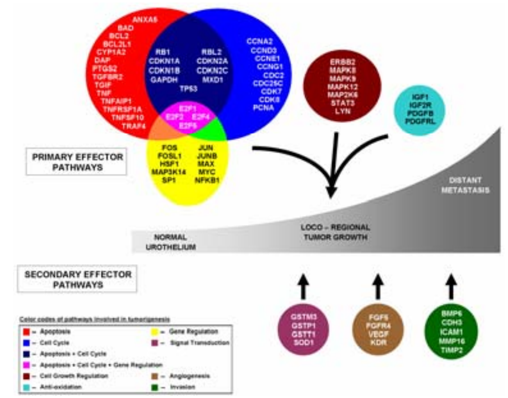
GENETIC PROGRAMMING PROCESS



FINAL META-RULE FOR NODE POSITIVE PATIENTS

Rule number	Classifier Rule
1	exp(exp(*HSF1*)) - exp(*MXD1*)/(*KDR* - *MAP2K6*) > 2.718
2	(*MAP2K6*/*KDR*) x (exp(*TGIF*) - *MAP2K6*/*ICAM1*) > .709
3	(*ICAM1* - *CDK8*)/(exp(*JUNB*) x (*JUNB* - exp(*TGFBR2*))) > 1.32
4	*ANXA5* x *MAP2K6*/(*KDR* x (*ICAM1* - *CDK8*)) > 1.701
5	(*ICAM1* - *MAP2K6*) x exp(*MAP2K6* - *KDR*) > 3653.813
6	(*ICAM1* - *CDK8*) x *TP53*/(exp(*TGFBR2*) x *PTGS2*) > 21941.453
7	(*CCND3*/*MAP2K6*) x (exp(*BMP6*) - (*KDR*/*MAP2K6*)) > .201
8	*MAP2K6*/(*CDKN1A* x exp(*MAPK12*) x (*CDC25C* - *KDR*)) > 7.703
9	(*ANXA5* - exp(*PDGFRL*))/(*CDKN1A* x (*KDR* - exp(*TGFBR2*))) > .044
10	*ANXA5*/(*CDKN1A* x (exp(*PTGS2*) - (*CDK8*/*ICAM1*))) > 79.002
11	*MAP2K6*/(*KDR* x (*ICAM1* - (*TNFAIP1*/exp(*PDGFB*)))) > 1.182

GENE PANEL

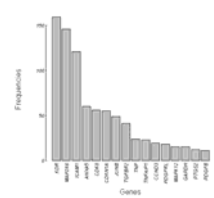


PERFORMANCE OF SELECTED META-RULE ON VALIDATION SET

	Pathologically Node Positive	Pathologically Node Negative
Predicted Node Positive by GP	6	2
Predicted Node Negative by GP	4	19

Accuracy: 81%
Sensitivity: 60%
Specificity: 90%
Positive Predictive Value: 75%
Negative Predictive Value: 83%

GENE USAGE (220 RULES)



GENE USAGE PROBABILITY DUE TO RANDOM CHANCE

Gene	Binomial probability
KDR	9.69E-130
MAP2K6	1.13E-110
ICAM1	4.10E-78
ANXA5	7.04E-20
CDK8	3.38E-17
CDKN1A	1.49E-16
JUNB	6.56E-13
TGFBR2	1.08E-08
TNF	1.11E-02
TNFAIP1	1.76E-02
CCND3	6.73E-02
PDGFRL	8.23E-02
MAPK12	1.04E-01
GAPDH	1.04E-01
PTGS2	7.05E-02
PDGFB	5.26E-02

GENE EXPRESSION MOTIFS

Rules	Common Motif	Implication for NP cases
2, 4, 7, 11	*MAP2K6* / *KDR* → **NP**	*MAP2K6>KDR*
5, 1	*MAP2K6* – *KDR* → **NP**	
2	*MAP2K6* / *ICAM1* → **NN**	*ICAM1>MAP2K6*
5	*ICAM1* – *MAP2K6* → **NP**	

Cumulative implication → *ICAM1>MAP2K6>KDR*

SUMMARY

➢ We present an objective and reproducible method for detection of nodal metastasis from the quantitative molecular profiles of primary bladder cancer tissues. A genetic programming system was used to generate classifier rules based on transcript profiles obtained by StaRT-PCR analysis that can provide a standardized output of quantitative gene expression relative to a housekeeping gene like β-actin.

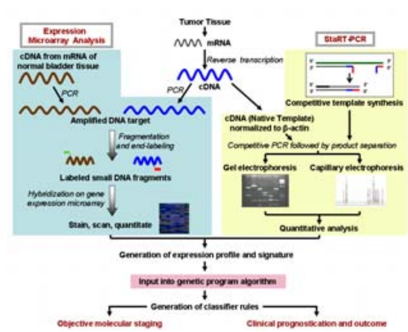
➢ The gene usage frequencies suggest the key involvement of *ICAM1*, *MAP2K6* and *KDR* genes in the development of nodal metastasis. These genes and their corresponding proteins have been separately shown to influence bladder cancer progression. Further studies are needed to clarify their precise biological role and examine them as new targets for therapeutic intervention.

➢ Of particular interest are the gene expression motifs involving the most frequently used genes. Combined analyses of the unique mathematical combinations in which these genes are organized in the classifier rules suggest novel relationships between specific genes and pathways. These also suggest class-specific signatures where a small number of genes can characterize tumors as node positive or node negative, and more importantly, provide an early indication of their progression towards node positive status.

➢ Genetic programming thus has the advantage of producing human-readable rules that define tangible relationships between the most influential genes. These rules can also express non-linear relationships that are more representative of biological systems. At the same time, genetic programming can limit the complexity of the rules while maintaining their robustness which can limit the cost of the procedure.

➢ Our group is currently considering several questions including an approach for multi-class problems, automated methods for selecting key transcripts and automated identification of significant motifs. Further studies will be aimed at correlating molecular markers and motifs with clinical outcome in an effort to employ them as reliable, reproducible and objective indicators of prognosis.

FUTURE WORK



Mitra AP et al. *BJU Int* 96:7-12; 2005.

Correspondence: Anirban P. Mitra (amitra@usc.edu)

Keck School of Medicine of the University of Southern California – 9th Annual Retreat of the Department of Pathology

January 20-21, 2006; Embassy Suites Hotel, Mandalay Bay Beach Resort, Oxnard, CA