

## Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Potti A, Mukherjee S, Petersen R, et al. A genomic strategy to refine prognosis in early-stage non–small-cell lung cancer. *N Engl J Med* 2006;355:570-580.

## **Methods:**

### **Patients and samples**

The initial analysis used 91 tumor samples of patients with early stage (Ia/Ib, IIa/IIb and IIIa) NSCLC, who also had clearly defined clinical outcome data, identified from the Duke Lung Cancer Prognostic Laboratory. We determined the percentage tumor content and histologic type of each tumor before RNA extraction. Of the 91 RNA samples, 89 were of sufficient quality for gene expression analysis. Our initial goal was to identify gene expression patterns characteristic of certain patient cohorts within the group. The cohort of patients with early-stage NSCLC was selected to have an equal mix of the two major histologic subtypes: squamous cell carcinoma and adenocarcinoma. In addition, each histologic subset had approximately equal number of patients who survived over 5 years and those who died within 2.5 years of initial diagnosis of a documented disease recurrence.

The ACOSOG Z0030 study is a completed prospective, multi-institutional phase III trial of 1100 patients with stage I NSCLC randomized to complete resection with mediastinal lymph node dissection or sampling. A subset of 416 patients had fresh-frozen tumor collected and banked at ACOSOG Central Specimen Bank at Washington University. Forty samples from patients with at least 28 months of follow-up were obtained for RNA extraction and microarray analysis. Of these, 25 cases were found to have both acceptable tumor cell content and adequate RNA quality for analysis. Approximately half ( $n = 13$ ) of these patients had died of cancer recurrence.

The CALGB 9761 study is a completed multi-institutional prospective phase II trial of approximately 500 patients with clinical stage I and II NSCLC, and was designed to assess the prognostic significance of micrometastatic disease using RT-PCR assay of expression of mucin-1 and carcinoembryonic antigen. Patients had fresh-frozen tumor and lymph nodes collected according to a rigorous, quality-controlled protocol such that high quality RNA was extracted from over 90% of tumors. The RNA samples derived from tumors of 84 patients were analyzed by microarray analysis (using Affymetrix U133A GeneChip). This was a blinded external validation step: the gene expression-based predictions of recurrence were made without a priori knowledge of the outcome and were independently validated with clinical outcome (survival) by a CALGB statistician. The mean follow-up for patients in this group was 5.3 years. There were 34 patients with recurrence, and 50 patients who were disease free at the time of follow-up. None of the patients in the Duke, ACOSOG, and CALGB cohorts received adjuvant chemotherapy or external beam radiation. Table 1 provides a summary of the clinical and demographic characteristics of the patients enrolled in the training (Duke), and validation (ACOSOG and CALGB) cohorts. Full details are provided in Supplementary Material (Supplementary Table-1).

## Methods:

**Tumor analyses.** Approximately 30 mg of lung cancer tissue was added to a chilled BioPulverizer H tube [Bio101 Systems, Carlsbad, CA]. Lysis buffer from the Qiagen RNeasy Mini kit was added and the tissue homogenized for 20 seconds in a Mini-Beadbeater [Biospec Products, Bartlesville, OK]. Tubes were spun briefly to pellet the garnet mixture and reduce foam. The lysate was transferred to a new 1.5 ml tube using a syringe and 21 gauge needle, followed by passage through the needle 10 times to shear genomic DNA. Total RNA was extracted from tumors using the Qiagen RNeasy Mini kit. The samples from the Duke Cohort and ACOSOG Z0030 were prepared and arrayed using Affymetrix U133 plus 2.0 GeneChips at the Duke Microarray Facility, and the samples from CALGB 9761 were prepared and arrayed using Affymetrix U133A GeneChips at the University of Michigan.

**Gene expression arrays.** For the Duke and ACOSOG samples, total RNA extracted from the tumor tissue with RNeasy kits (Qiagen, Nalencia, CA, USA) was assessed for quality with an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA). Hybridization targets (probes for hybridization) were prepared from total RNA according to standard Affymetrix protocols. The amount of starting total RNA for each reaction was 10 µg. Briefly, first-strand cDNA was generated using a T7- linked oligo-dT primer, followed by second-strand synthesis. An in vitro transcription reaction was performed to generate cRNA containing biotinylated UTP and CTP, which was then chemically fragmented at 95°C for 35 min. The fragmented, biotinylated cRNA was incubated in MES buffer (2-[N-morpholino]ethansulfonic acid) containing 0.5 mg/ml acetylated bovine serum albumin to the Affymetrix GeneChip Human U133 plus 2.0 arrays at 45°C for 16 hr, according to the directions of the manufacturer. The arrays contained over 54,000 probes, representing genes. Arrays were washed and stained with streptavidin-phycoerythrin (SAPE, Molecular Probes). Signal amplification was performed using a biotinylated anti-streptavidin antibody (Vector Laboratories, Burlingame, CA) at 3µg/ml. This was followed by a second staining with SAPE. Normal goat IgG (2 mg/ml) was used as a blocking agent. Scans were performed with an Affymetrix GeneChip scanner and the expression value for each gene was calculated using the Affymetrix Microarray Analysis Suite (v5.0), computing the expression intensities in 'signal' units defined by software. Scaling factors were determined for each hybridization based on an arbitrary target intensity of 500. Scans were rejected if the scaling factor exceeded a factor of 30. Expression was calculated using the robust multi-array average (RMA) algorithm implemented in the Bioconductor (<http://www.bioconductor.org>) extensions to the R statistical programming environment. RMA generates log-2 scaled measures of expression using a linear model robustly fit to background-corrected and quantile-normalized probe-level expression data and has been shown to have a better ability to detect differential expression in spike-in experiments. The probe sets were screened to remove control genes and those with a small variance and those expressed at low levels.

**Statistical analysis.** We carried out statistical analysis using the metagene construction and binary prediction tree analysis as previously described<sup>26-30</sup>. The initial step filtered out genes whose maximum expression did not exceed the median value of expression or did not vary more than two-fold across the samples, to remove genes with extremely low levels of expression or little variance. The remaining genes (approximately 20,700) were then used to generate a model in which a restricted set of differentially expressed genes could distinguish the comparison groups and ultimately predict recurrence. This set of genes was then further screened by computing the simple correlation between expression of each gene and the binary recurrence outcome across samples, ranking genes by the strength of correlation and then restricting the focus to the top 10% (about 2070 genes). These genes were then clustered and used to compute metagene summaries as described below.

In the leave-one-out cross-validation analyses of the Duke training data, this process of gene screening and selection was reapplied for each sample. K-means clustering was used to create groupings of genes with between 15 and 50 genes per cluster, and a single metagene expression summary was computed for each group. The metagene for a cluster of genes is the dominant singular factor (principal component), computed using a singular value decomposition of expression levels of the genes in the metagene cluster on all samples. It represents the dominant average expression pattern of the cluster across tumor samples<sup>26</sup>. The set of metagenes and clinical factors are then used in binary classification tree analysis to recursively partition the samples into smaller subsets within which predictions of recurrence (0 = 5 year disease-free survival from diagnosis of recurrence, 1 = death within 2.5 years from diagnosis of recurrence) are made in terms of estimated relative probabilities<sup>27, 31, 32</sup>. The analysis computes and weighs many such trees, and integrates them to provide overall risk predictions for each individual patient. By identifying the subset of metagenes receiving the highest weight across the trees, we identified the corresponding clusters of genes that most heavily contribute to overall risk predictions<sup>26</sup>. The dominant metagenes that constitute the final model are described in the online Supplement.

To compare the prognostic efficacy of genomic and clinical strategies, clinical variables previously shown to be of prognostic value (age, gender, tumor size, stage of disease, histologic subtype and smoking history) were treated as factors or principal components (similar to metagenes in the genomic model) in a classification tree analysis to generate a 'clinical model', identical to the approach used to create the genomic model. The end result is a probability of recurrence which represents the conglomerate prognostic value of the individual clinical variables. Using Graphpad software, we computed a c-statistic (comparable to area under the curve in a receiver operated characteristic (ROC) curve when predicting binary outcomes) for the model including just the clinical variables, a c-statistic for a model that included the genomic prediction of recurrence, and a c-statistic for a model that included both clinical and genomic variables.

Accuracy of a model was defined using the 50% probability as the cut-off - if the model's estimate for probability of recurrence was >50%, the patient was classified as high risk (i.e., the model predicts recurrence). And if the model estimates a probability of recurrence <50%, the patient is classified as being at low risk for recurrence.

**Supplementary Table 1.** Summary of the individual clinical, demographic, and outcome characteristics of the patients enrolled in the training (Duke), and validation (ACOSOG and CALGB) cohorts.

**Supplementary Table 2. Genes in the metagenes selected for predicting recurrence.**  
The genes contained in the top 9 metagenes selected in the tree model analyses together with Gene Ontology (GO) annotations.

**Supplementary Figure 1. Prediction of recurrence based on the genomic model as a function of NSCLC stage.**

A. Predictions of recurrence as a function of clinical stage.  
B. Kaplan Meier estimates of survival by stage of NSCLC using the genomic model. The red curve represents patients predicted to be at high risk (>50% probability of recurrence) and the blue curve represents patients predicted to be at low risk (<50% probability of recurrence). Tick marks indicate patients whose data were censored by the time of last follow up OR owing to death.

**Supplementary Figure 2. Prediction of recurrence as a function of histological subtype.**

A. Same as in Supplementary Figure 2, red symbols indicate patients with recurrence and blue symbols indicate those without recurrence.  
B. Kaplan Meier estimates of survival as a function of histologic subtype. Tick marks indicate patients whose data were censored by the time of last follow up OR owing to death.

**Supplementary Figure 3. The performance of the metagene model to a previously published squamous NSCLC dataset (courtesy Dr. Zhifu Sun, Mayo Clinic).**

The predictive model generated with the entire Duke set of samples was used to estimate recurrence probabilities for the ACOSOG samples. Red symbols (▲) indicate patients with recurrence and blue symbols (■) indicate those without recurrence.

Supplementary Table 1.

Sample Set	Sex	Histology	Age	Stage	Status(A/D)
Duke	F	S	73	IV	1
Duke	M	S	66	2A	1
Duke	F	A	72	1B	0
Duke	F	S	66	1B	1
Duke	F	A	68	1A	0
Duke	M	S	69	1A	0
Duke	F	S	68	1A	0
Duke	F	A	69	1A	1
Duke	F	S	73	1B	1
Duke	F	A	63	1A	1
Duke	M	A	51	1B	1
Duke	M	S	62	1A	1
Duke	F	A	65	1A	1
Duke	F	A	69	1A	0
Duke	M	A	72	1B	1
Duke	F	S	68	1B	0
Duke	M	S	76	1B	1
Duke	M	A	69	2B	0
Duke	M	A	67	1A	0
Duke	M	S	75	1A	0
Duke	M	A	65	1B	1
Duke	M	S	57	2A	0
Duke	F	A	54	1A	0
Duke	M	A	53	1B	1
Duke	M	S	51	1B	0
Duke	M	S	66	1A	0
Duke	F	S	64	1B	1
Duke	M	S	74	1B	0
Duke	M	S	51	1B	1
Duke	M	S	69	1B	0
Duke	M	S	80	2B	1
Duke	F	A	73	1B	1
Duke	F	S	48	1B	0
Duke	M	S	65	1A	0
Duke	M	S	65	1A	0
Duke	M	S	70	2B	1
Duke	M	A	69	2B	0
Duke	M	A	54	1A	1
Duke	M	S	82	1A	0
Duke	M	S	54	1A	0
Duke	M	S	73	1B	1
Duke	M	A	57	1A	0
Duke	M	S	68	2B	1
Duke	M	S	32	2A	0
Duke	F	A	57	1A	0
Duke	F	S	64	2B	0
Duke	F	A	70	1A	0
Duke	F	A	68	1A	1
Duke	M	A	64	1A	0
Duke	M	A	73	1A	1
Duke	F	A	43	1B	0
Duke	F	A	63	1A	1

Duke	M	A	75	1A	0	
Duke	F	A	68	3A	0	
Duke	M	S	69	1B	1	
Duke	F	S	57	1A	0	
Duke	M	A	55	1A	0	
Duke	M	S	64	1B	1	
Duke	F	S	47	1B	1	
Duke	F	A	67	2B	1	
Duke	F	A	75	1A	0	
Duke	M	A	73	1B	1	
Duke	M	A	70	2B	1	
Duke	M	S	73	1B	1	
Duke	M	A	56	2B	1	
Duke	M	A	65	2A	1	
Duke	M	A	66	1A	0	
Duke	M	S	58	1B	1	
Duke	M	S	79	1A	0	
Duke	F	S	66	1B	0	
Duke	M	S	76	1B	1	
Duke	M	A	71	1B	0	
Duke	M	A	67	3A	1	
Duke	M	S	55	2B	1	
Duke	F	A	79	3A	1	
Duke	M	S	81	3A	1	
Duke	M	A	83	1A	0	
Duke	M	A	62	1A	0	
Duke	F	A	66	1A	1	
Duke	M	S	60	1A	1	
Duke	M	S	68	1B	0	
Duke	F	S	83	1A	0	
Duke	M	S	72	2A	1	
Duke	M	S	55	1B	0	
Duke	F	A	69	1B	0	
Duke	M	A	50	1B	1	
Duke	F	A	68	1A	0	
Duke	F	S	55	2B	1	
Duke	F	A	76	IB	0	Identifieur
CALGB 9761	M	A	70	IB	1	87290
CALGB 9761	M	A	58	IB	0	78918
CALGB 9761	F	A	48	IB	1	83787
CALGB 9761	M	A	82	IIB	1	85152
CALGB 9761	M	A	62	IA	1	86281
CALGB 9761	M	A	82	IIIB	1	79124
CALGB 9761	M	A	68	IA	0	79124
CALGB 9761	M	A	58	IB	0	83790
CALGB 9761	M	A	63	IA	1	87135
CALGB 9761	M	A	78	IIB	1	86011
CALGB 9761	F	A	57	IB	1	79525
CALGB 9761	F	A	77	IA	0	78503
CALGB 9761	F	A	86	IIIB	1	79189
CALGB 9761	F	A	78	IIIB	0	79176
CALGB 9761	M	A	73	IB	0	87255
CALGB 9761	F	A	70	IA	0	82247
CALGB 9761	F	A	75	IA	0	79629
CALGB 9761	F	A	49	IA	0	83505
CALGB 9761	F	A	49	IA	0	83057

CALGB 9761	F	A	51	IA	0	77996
CALGB 9761	M	A	66	IA	0	77960
CALGB 9761	F	A	76	IB	0	78290
CALGB 9761	F	A	78	IA	0	78328
CALGB 9761	M	A	76	IA	1	77946
CALGB 9761	F	A	60	IIIA	1	77738
CALGB 9761	M	A	73	IA	1	78119
CALGB 9761	F	A	70	IV	0	78109
CALGB 9761	F	A	72	IB	0	70592
CALGB 9761	F	A	55	IIIB	1	70888
CALGB 9761	F	A	64	IA	0	73916
CALGB 9761	F	A	53	IIB	0	71789
CALGB 9761	F	A	77	IA	0	71621
CALGB 9761	M	A	55	IA	0	77059
CALGB 9761	M	A	67	IB	1	69314
CALGB 9761	M	A	53	IIIB	0	77556
CALGB 9761	F	A	70	IA	0	77430
CALGB 9761	F	A	72	IB	1	68864
CALGB 9761	F	A	57	IB	0	76295
CALGB 9761	M	A	43	IA	1	71886
CALGB 9761	F	A	65	IA	0	70719
CALGB 9761	F	A	66	IB	0	69914
CALGB 9761	F	A	52	IIA	1	75704
CALGB 9761	F	A	69	IA	0	70709
CALGB 9761	F	A	33	IB	0	70160
CALGB 9761	M	A	59	IIIA	1	74083
CALGB 9761	M	A	62	IA	1	69526
CALGB 9761	F	A	46	IIA	0	86908
CALGB 9761	F	A	63	IIA	0	76021
CALGB 9761	M	A	60	IIA	1	82902
CALGB 9761	F	A	43	IIA	0	86908
CALGB 9761	F	A	67	IIA	0	76021
CALGB 9761	M	A	50	IIA	1	82902
CALGB pilot/VAMC	M	A	62	1b	0	
CALGB pilot/VAMC	M	A	72	1b	1	
CALGB pilot/VAMC	M	A	68	1a	1	
CALGB pilot/VAMC	M	A	63	1a	0	
CALGB pilot/VAMC	M	A	58	2b	1	
CALGB pilot/VAMC	M	A	77	3a	1	
CALGB pilot/VAMC	M	A	60	2a	0	
CALGB pilot/VAMC	M	A	77	1b	0	
CALGB pilot/VAMC	M	A	78	1a	1	
CALGB pilot/VAMC	M	A	49	2a	0	
CALGB pilot/VAMC	M	A	69	1b	1	
CALGB pilot/VAMC	M	A	67	2a	1	
CALGB pilot/VAMC	M	A	68	3a	0	
CALGB pilot/VAMC	M	A	46	2b	0	
CALGB pilot/VAMC	M	A	65	1a	0	
CALGB pilot/VAMC	M	A	46	2b	1	
CALGB pilot/VAMC	M	A	65	1a	1	
CALGB pilot/VAMC	M	A	75	2b	1	
CALGB pilot/VAMC	M	A	57	3b	1	
CALGB pilot/VAMC	M	A	72	1b	0	
CALGB pilot/VAMC	M	A	64	3a	1	
CALGB pilot/VAMC	M	A	59	1b	0	

CALGB pilot/VAMC	M	A	65	1b	0
CALGB pilot/VAMC	M	A	69	1a	0
CALGB pilot/VAMC	M	A	77	1b	1
CALGB pilot/VAMC	M	A	67	2b	0
CALGB pilot/VAMC	M	A	67	1b	0
CALGB pilot/VAMC	M	A	74	3a	1
CALGB pilot/VAMC	M	A	68	3b	0
CALGB pilot/VAMC	M	A	67	3a	1
CALGB pilot/VAMC	M	A	57	1a	1
CALGB pilot/VAMC	M	A		1b	1
ACOSOG	M	S	51	1B	1
ACOSOG	M	S	66	1A	0
ACOSOG	F	S	68	1B	1
ACOSOG	F	A	48	2B	0
ACOSOG	M	S	74	1B	0
ACOSOG	M	S	65	1A	0
ACOSOG	M	S	73	1B	0
ACOSOG	F	S	47	1B	1
ACOSOG	M	A	53	1B	0
ACOSOG	M	S	68	1B	0
ACOSOG	M	A	65	1B	0
ACOSOG	F	S	57	2A	1
ACOSOG	M	A	80	2B	1
ACOSOG	F	S	75	1B	1
ACOSOG	F	A	73	1B	0
ACOSOG	M	S	65	1A	1
ACOSOG	M	A	47	1B	1
ACOSOG	M	S	51	1B	0
ACOSOG	F	A	70	2B	1
ACOSOG	M	A	56	2B	1
ACOSOG	M	A	65	2A	1
ACOSOG	M	A	70	2B	0
ACOSOG	M	A	66	1A	1
ACOSOG	F	S	41	1A	1
ACOSOG	F	S	64	1B	0

**Supplementary Table 2.**

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 79</b>					
<a href="#">200983_x_at</a>	CD59 antigen p18-20 (antigen identified by monoclonal antibodies 16.3A5, EJ16, EJ30, EL32 and G344)	CD59	immune response cell surface receptor linked signal transduction blood coagulation	---	membrane fraction plasma membrane
<a href="#">201481_s_at</a>	phosphorylase, glycogen; brain	PYGB	carbohydrate metabolism glycogen catabolism	phosphorylase activity transferase activity, transferring glycosyl groups	---
<a href="#">201627_s_at</a>	insulin induced gene 1	INSIG1	metabolism cell proliferation	---	---
<a href="#">202812_at</a>	glucosidase, alpha; acid (Pompe disease, glycogen storage disease type II)	GAA	carbohydrate metabolism glycogen catabolism generation of precursor metabolites and energy	hydrolase activity, hydrolyzing O-glycosyl compounds alpha-glucosidase activity	lysosome
<a href="#">203796_s_at</a>	B-cell CLL/lymphoma 7A	BCL7A	---	actin binding	---
<a href="#">203856_at</a>	vaccinia related kinase 1	VRK1	protein amino acid phosphorylation	protein serine/threonine kinase activity ATP binding transferase activity	---
<a href="#">205118_at</a>	---	---	---	---	---
<a href="#">206022_at</a>	Norrie disease (pseudoglioma)	NDP	vacuole organization and biogenesis signal transduction cell-cell signaling neurogenesis visual perception perception of sound cell proliferation	growth factor activity	extracellular space
<a href="#">206986_at</a>	---	---	---	---	---
<a href="#">208341_x_at</a>	chorionic somatomammotropin hormone 2	CSH2	signal transduction cell-cell signaling pregnancy	growth hormone receptor binding hormone activity structural constituent of chorion (sensu Insecta)	extracellular region extracellular space
<a href="#">208871_at</a>	dentatorubral-pallidolusian atrophy (atrophin-1)	DRPLA	central nervous system development	protein binding	nucleus cytoplasm
<a href="#">209479_at</a>	chromosome 6 open reading frame 80	C6orf80	---	receptor activity	---
<a href="#">210220_at</a>	frizzled homolog 2 (Drosophila)	FZD2	establishment of tissue polarity G-protein coupled receptor protein signaling pathway development Wnt receptor signaling pathway	non-G-protein coupled 7TM receptor activity G-protein coupled receptor activity	integral to plasma membrane
<a href="#">210842_at</a>	neuropilin 2	NRP2	angiogenesis electron transport cell adhesion neurogenesis axon guidance	receptor activity vascular endothelial growth factor receptor activity electron transporter activity semaphorin receptor activity	membrane fraction integral to membrane
<a href="#">210890_x_at</a>	killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 1	KIR2DL1	immune response negative regulation of natural killer cell activity	receptor activity HLA-C specific inhibitory MHC class I receptor activity	integral to plasma membrane
<a href="#">211090_s_at</a>	PRP4 pre-mRNA processing factor 4 homolog B (yeast) PRP4 pre-mRNA processing factor 4 homolog B (yeast)	PRPF4B	nuclear mRNA splicing, via spliceosome protein amino acid phosphorylation RNA splicing	protein serine/threonine kinase activity ATP binding transferase activity	nucleus
<a href="#">211168_s_at</a>	regulator of nonsense transcripts 1	RENT1	mRNA catabolism, nonsense-mediated decay regulation of translational termination DNA restriction	RNA binding helicase activity protein binding ATP binding type III site-specific deoxyribonuclease activity hydrolase activity	cytoplasm
<a href="#">211207_s_at</a>	acyl-CoA synthetase long-chain family member 6	ACSL6	lipid metabolism fatty acid metabolism acyl-CoA metabolism metabolism	magnesium ion binding catalytic activity long-chain-fatty-acid-CoA ligase activity ligase activity	mitochondrial outer membrane peroxisomal membrane microsome plasma membrane
<a href="#">213457_at</a>	malignant fibrous histiocytoma amplified sequence 1 MFHAS1	MFHAS1	small GTPase mediated signal transduction	GTP binding	---

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 40</b>					
<u>200045_at</u>	ATP-binding cassette, sub-family F (GCN20), member 1 ATP-binding cassette, sub-family F (GCN20), member 1	ABCF1	---	nucleotide binding ATP binding ATPase activity	---
<u>200881_s_at</u>	DnaJ (Hsp40) homolog, subfamily A, member 1	DNAJA1	protein folding response to unfolded protein signal transduction G-protein coupled receptor protein signaling pathway G-protein signaling, adenylate cyclase activating pathway pregnancy protein secretion	heat shock protein binding unfolded protein binding	---
<u>200981_x_at</u>	GNAS complex locus	GNAS	protein-nucleus import, docking signal transduction protein transport	GTPase activity signal transducer activity GTP binding	extracellular region heterotrimeric G-protein complex plasma membrane
<u>200995_at</u>	Importin 7	IPO7	protein modification proteolysis and peptidolysis neuropeptide signaling pathway metabolism	small GTPase regulatory/interacting protein activity binding RAN protein binding protein transporter activity	soluble fraction nucleus nuclear pore cytoplasm
<u>201116_s_at</u>	carboxypeptidase E	CPE	proteolysis and peptidolysis	carboxypeptidase activity carboxypeptidase A activity carboxypeptidase E activity metallopeptidase activity hydrolase activity	plasma membrane
<u>201120_s_at</u>	progesterone receptor membrane component 1	PGRMC1	---	receptor activity steroid binding	microsome integral to plasma membrane
<u>201138_s_at</u>	Sjogren syndrome antigen B (autoantigen La)	SSB	transcription from Pol III promoter RNA processing tRNA modification RNA-nucleus export histone mRNA metabolism	tRNA binding mRNA binding	nucleus cytoplasm ribonucleoprotein complex
<u>201159_s_at</u>	N-myristoyltransferase 1	NMT1	N-terminal protein myristoylation protein-lipoylation	glycylpeptide N-tetradecanoyltransferase activity acyltransferase activity transferase activity	---
<u>201182_s_at</u>	chromodomain helicase DNA binding protein 4	CHD4	chromatin assembly or disassembly regulation of transcription from Pol II promoter chromosome organization and biogenesis (sensu Eukaryota) protein ubiquitination chromatin modification	chromatin binding ATP-dependent DNA helicase activity ubiquitin-protein ligase activity ATP binding zinc ion binding	ubiquitin ligase complex chromatin nucleus
<u>201455_s_at</u>	aminopeptidase puromycin sensitive	NPEPPS	proteolysis and peptidolysis	aminopeptidase activity membrane alanyl aminopeptidase activity metallopeptidase activity zinc ion binding hydrolase activity	nucleus cytosol
<u>202666_s_at</u>	actin-like 6A	ACTL6A	regulation of cell growth chromatin remodeling regulation of transcription, DNA-dependent signal transduction response to pest, pathogen or parasite chromatin modification	chromatin binding motor activity structural constituent of cytoskeleton	nucleus actin filament TIP60 histone acetyltransferase complex
<u>203018_s_at</u>	synovial sarcoma, X breakpoint 2 interacting protein	SSX2IP	cell adhesion	protein binding	nucleus
<u>205556_at</u>	msh homeo box homolog 2 (Drosophila)	MSX2	skeletal development regulation of transcription, DNA-dependent development	transcription factor activity	nucleus
<u>206302_s_at</u>	nudix (nucleoside diphosphate linked moiety X)-type motif 4	NUDT4	intracellular signaling cascade cyclic nucleotide metabolism calcium-mediated signaling cyclic-nucleotide-mediated signaling regulation of RNA-nucleus export intracellular transport	diphosphoinositol-polyphosphate diphosphatase activity hydrolase activity	intracellular
<u>209963_s_at</u>	erythropoietin receptor	EPOR	signal transduction small GTPase mediated signal transduction	erythropoietin receptor activity guanyl-nucleotide exchange factor activity	integral to plasma membrane
<u>210349_at</u>	calcium/calmodulin-dependent protein kinase IV	CAMK4	protein amino acid phosphorylation signal transduction	protein serine/threonine kinase activity calcium- and calmodulin-dependent protein kinase activity calmodulin binding ATP binding transferase activity	nucleus
<u>211442_x_at</u>	cytochrome P450, family 3, subfamily A, polypeptide 43	CYP3A43	electron transport xenobiotic metabolism	unspecific monooxygenase activity	membrane fraction endoplasmic reticulum microsome membrane
<u>211444_at</u>	---	---	---	---	---

<a href="#">211720_x_at</a>	ribosomal protein, large, P0 ribosomal protein, large, P0	RPLP0	protein biosynthesis translational elongation	RNA binding structural constituent of ribosome	intracellular ribosome cytosolic large ribosomal subunit (sensu Eukaryota)
<a href="#">211778_s_at</a>	zinc finger protein 339 zinc finger protein 339	ZNF339	regulation of transcription, DNA-dependent	DNA binding zinc ion binding	nucleus
<a href="#">212360_at</a>	adenosine monophosphate deaminase 2 (isoform L)	AMPD2	purine nucleotide metabolism purine ribonucleoside monophosphate biosynthesis	AMP deaminase activity hydrolase activity	---
<a href="#">212787_at</a>	YLP motif containing 1	YLPM1	---	---	nucleus
<a href="#">213244_at</a>	secretory carrier membrane protein 4	SCAMP4	protein transport	---	integral to membrane extracellular region cytoskeleton integral to plasma membrane
<a href="#">213693_s_at</a>	Mucin 1, transmembrane	MUC1	---	actin binding hormone activity	---
<a href="#">213935_at</a>	abhydrolase domain containing 5	ABHD5	proteolysis and peptidolysis aromatic compound metabolism	aminopeptidase activity peptidase activity	---
<a href="#">214421_x_at</a>	cytochrome P450, family 2, subfamily C, polypeptide 9	CYP2C9	electron transport	monooxygenase activity (S)-limonene 7-monooxygenase activity	endoplasmic reticulum microsome membrane

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 31</b>					
<u>200834_s_at</u>	ribosomal protein S21	RPS21	protein biosynthesis cell differentiation	structural constituent of ribosome	intracellular ribosome
<u>201037_at</u>	phosphofructokinase, platelet	PFKP	glycolysis	magnesium ion binding 6-phosphofructokinase activity ATP binding kinase activity transferase activity	cytoplasm 6-phosphofructokinase complex
<u>201637_s_at</u>	fragile X mental retardation, autosomal homolog 1	FXR1	apoptosis	nucleic acid binding RNA binding	nucleolus cytoplasm polysome
<u>201850_at</u>	capping protein (actin filament), gelsolin-like	CAPG	protein complex assembly response to pest, pathogen or parasite barbed-end actin filament capping	actin binding	nucleus F-actin capping protein complex
<u>202325_s_at</u>	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6	ATP5J	generation of precursor metabolites and energy ATP synthesis coupled proton transport proton transport	transporter activity hydrogen-transporting ATP synthase activity, rotational mechanism hydrogen-transporting ATPase activity, rotational mechanism	membrane fraction mitochondrion mitochondrial inner membrane proton-transporting two-sector ATPase complex
<u>204633_s_at</u>	ribosomal protein S6 kinase, 90kDa, polypeptide 5	RPS6KA5	regulation of transcription, DNA-dependent protein amino acid phosphorylation proteolysis and peptidolysis response to stress epidermal growth factor receptor signaling pathway protein kinase cascade response to external stimulus histone phosphorylation response to chemical substance	catalytic activity protein serine/threonine kinase activity protein-tyrosine kinase activity ATP binding metalloproteinase activity zinc ion binding transferase activity	nucleus
<u>204727_at</u>	WD repeat and HMG-box DNA binding protein 1	WDHD1	regulation of transcription, DNA-dependent	DNA binding	nucleoplasm cytoplasm
<u>207260_at</u>	FEV (ETS oncogene family)	FEV	regulation of transcription, DNA-dependent transcription from Pol II promoter	transcription factor activity transcription corepressor activity	nucleus
<u>209343_at</u>	EF hand domain family, member D1	EFHD1	---	calcium ion binding	---
<u>210381_s_at</u>	cholecystokinin B receptor	CCKBR	G-protein coupled receptor protein signaling pathway phospholipase C activation positive regulation of cytosolic calcium ion concentration digestion sensory perception feeding behavior cell proliferation positive regulation of cell proliferation	rhodopsin-like receptor activity phosphoinositide phospholipase C activity receptor activity gastrin receptor activity phosphatidylinositol 3-kinase regulator activity	integral to plasma membrane
<u>212034_s_at</u>	exocyst complex component 7	EXOC7	exocytosis protein transport	---	exocyst
<u>212410_at</u>	EF hand domain family, member A1	EFHA1	muscle development	calcium ion binding	---
<u>212600_s_at</u>	ubiquinol-cytochrome c reductase core protein II	UQCRC2	electron transport oxidative phosphorylation proteolysis and peptidolysis aerobic respiration	metalloendopeptidase activity ubiquinol-cytochrome-c reductase activity oxidoreductase activity	mitochondrion mitochondrial electron transport chain

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 19</b>					
<u>200908_s_at</u>	---	---	---	---	---
<u>203914_x_at</u>	hydroxyprostaglandin dehydrogenase 15-(NAD)	HPGD	prostaglandin metabolism metabolism	alcohol dehydrogenase activity prostaglandin-D synthase activity electron transporter activity 15-hydroxyprostaglandin dehydrogenase (NAD+) activity oxidoreductase activity	---
<u>204189_at</u>	retinoic acid receptor, gamma	RARG	regulation of transcription, DNA-dependent development	transcription factor activity steroid hormone receptor activity retinoic acid receptor activity	nucleus
<u>204928_s_at</u>	solute carrier family 10 (sodium/bile acid cotransporter family), member 3	SLC10A3	transport sodium ion transport organic anion transport	bile acid:sodium symporter activity symporter activity	integral to membrane
<u>205094_at</u>	peroxisomal biogenesis factor 12	PEX12	protein ubiquitination	ubiquitin-protein ligase activity protein binding zinc ion binding	ubiquitin ligase complex peroxisome integral to peroxisomal membrane
<u>205734_s_at</u>	lymphoid nuclear protein related to AF4	LAF4	regulation of transcription, DNA-dependent development	DNA binding	nucleus
<u>205767_at</u>	epiregulin	EREG	regulation of cell cycle angiogenesis epidermal growth factor receptor signaling pathway cell-cell signaling cell proliferation	epidermal growth factor receptor binding growth factor activity	extracellular space integral to plasma membrane
<u>206390_x_at</u>	platelet factor 4 (chemokine (C-X-C motif) ligand 4)	PF4	immune response negative regulation of angiogenesis cytokine and chemokine mediated signaling pathway platelet activation immune cell chemotaxis negative regulation of megakaryocyte differentiation	chemokine activity heparin binding	extracellular region
<u>207108_s_at</u>	Nipped-B homolog (Drosophila)	NIPBL	cell cycle	---	nucleus
<u>207572_at</u>	---	---	---	---	---
<u>207814_at</u>	defensin, alpha 6, Paneth cell-specific	DEFA6	xenobiotic metabolism response to pest, pathogen or parasite antimicrobial humoral response (sensu Vertebrata) defense response to bacteria defense response to fungi	---	extracellular region
<u>211211_x_at</u>	SH2 domain protein 1A, Duncan's disease (lymphoproliferative syndrome)	SH2D1A	cellular defense response intracellular signaling cascade cell-cell signaling	SH3/SH2 adaptor protein activity	cytoplasm
<u>213443_at</u>	---	---	---	---	---
<u>213873_at</u>	---	---	---	---	---

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 74</b>					
<u>203850_s_at</u>	kinesin family member 1A	KIF1A	microtubule-based movement anterograde axon cargo transport	microtubule motor activity ATP binding	microtubule associated complex
<u>203859_s_at</u>	paralemmin	PALM	cell motility regulation of cell shape	---	integral to plasma membrane cytoplasmic vesicle
<u>205887_x_at</u>	mutS homolog 3 (E. coli)	MSH3	mismatch repair	damaged DNA binding ATP binding	---
<u>206186_at</u>	membrane protein, palmitoylated 3 (MAGUK p55 subfamily member 3)	MPP3	signal transduction	guanylate kinase activity protein binding	integral to plasma membrane
<u>207096_at</u>	serum amyloid A4, constitutive	SAA4	acute-phase response	lipid transporter activity	extracellular region
<u>208008_at</u>	DKFZP434O047 protein	DKFZP434O047	---	---	---
<u>208755_x_at</u>	H3 histone, family 3A	H3F3A	---	---	---
<u>210650_s_at</u>	---	---	---	---	---
<u>210785_s_at</u>	chromosome 1 open reading frame 38	C1orf38	cell adhesion	---	---
<u>211154_at</u>	thrombopoietin (myeloproliferative leukemia virus oncogene ligand, megakaryocyte growth and development factor)	THPO	development cell proliferation	cytokine activity hormone activity growth factor activity	extracellular region
<u>213650_at</u>	golgin-67	GOLGIN-67	---	---	Golgi apparatus

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 41</b>					
<a href="#">201895_at</a>	v-raf murine sarcoma 3611 viral oncogene homolog	ARAF	protein amino acid phosphorylation intracellular signaling cascade	protein serine/threonine kinase activity protein-tyrosine kinase activity receptor signaling protein activity ATP binding transferase activity diacylglycerol binding	---
<a href="#">204168_at</a>	microsomal glutathione S-transferase 2	MGST2	signal transduction cell-cell signaling leukotriene biosynthesis antimicrobial humoral response (sensu Vertebrata)	glutathione transferase activity enzyme activator activity transferase activity	membrane fraction microsome integral to membrane
<a href="#">205844_at</a>	vanin 1 vanin 1	VNN1	nitrogen compound metabolism cell motility	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	---
<a href="#">206066_s_at</a>	RAD51 homolog C ( <i>S. cerevisiae</i> )	RAD51C	DNA repair DNA recombination	damaged DNA binding ATP binding DNA-dependent ATPase activity	nucleus
<a href="#">206143_at</a>	solute carrier family 26, member 3	SLC26A3	transport anion transport excretion sulfate transport	transcription factor activity transcription cofactor activity transporter activity sulfate porter activity antiporter activity	membrane fraction integral to membrane
<a href="#">206370_at</a>	phosphoinositide-3-kinase, catalytic, gamma polypeptide	PIK3CG	G-protein coupled receptor protein signaling pathway	inositol or phosphatidylinositol kinase activity phosphatidylinositol 3-kinase activity transferase activity phosphatidylinositol-4,5-bisphosphate 3-kinase activity	phosphoinositide 3-kinase complex
<a href="#">207737_at</a>	---	---	---	---	---
<a href="#">209971_x_at</a>	JTV1 gene	JTV1	protein biosynthesis	damaged DNA binding	---
<a href="#">210431_at</a>	alkaline phosphatase, placental-like 2	ALPPL2	metabolism phosphorylation	magnesium ion binding alkaline phosphatase activity hydrolase activity	membrane
<a href="#">210609_s_at</a>	tumor protein p53 inducible protein 3	TP53I3	induction of apoptosis by oxidative stress	alcohol dehydrogenase activity, zinc-dependent zinc ion binding	---
<a href="#">211062_s_at</a>	carboxypeptidase Z carboxypeptidase Z	CPZ	proteolysis and peptidolysis development	carboxypeptidase A activity transmembrane receptor activity peptidase activity	integral to membrane
<a href="#">211369_at</a>	---	---	---	---	---
<a href="#">213188_s_at</a>	MYC induced nuclear antigen	MINA	---	---	nucleus
<a href="#">213507_s_at</a>	karyopherin (importin) beta 1	KPNB1	protein-nucleus import, docking protein-nucleus import, translocation NLS-bearing substrate-nucleus import protein transport	binding nuclear localization sequence binding zinc ion binding protein transporter activity	nucleus nuclear pore cytoplasm
<a href="#">213517_at</a>	Poly(rC) binding protein 2	PCBP2	mRNA metabolism	nucleic acid binding DNA binding RNA binding	nucleus cytoplasm ribonucleoprotein complex
<a href="#">214207_s_at</a>	---	---	---	---	---

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 69</b>					
<u>202964_s_at</u>	regulatory factor X, 5 (influences HLA class II expression)	RFX5	regulation of transcription, DNA-dependent transcription from Pol II promoter	transcription factor activity transcription coactivator activity	nucleus
<u>203969_at</u>	hypothetical protein LOC153914	LOC153914	---	---	---
<u>203971_at</u>	solute carrier family 31 (copper transporters), member 1	SLC31A1	transport copper ion transport	copper ion transporter activity	integral to plasma membrane
<u>206308_at</u>	DNA (cytosine-5-)-methyltransferase 2	DNMT2	DNA methylation	DNA binding DNA (cytosine-5-)-methyltransferase activity	nucleus
<u>206691_s_at</u>	protein disulfide isomerase, pancreatic	PDIP	electron transport protein folding protein-ER retention	protein disulfide isomerase activity electron transporter activity isomerase activity	endoplasmic reticulum
<u>206692_at</u>	potassium inwardly-rectifying channel, subfamily J, member 10	KCNJ10	ion transport potassium ion transport	voltage-gated ion channel activity potassium channel activity ATP binding ATP-activated inward rectifier potassium channel activity	integral to plasma membrane
<u>206923_at</u>	protein kinase C, alpha	PRKCA	regulation of cell cycle protein amino acid phosphorylation cell surface receptor linked signal transduction intracellular signaling cascade induction of apoptosis by extracellular signals	protein kinase C activity protein-tyrosine kinase activity calcium ion binding ATP binding transferase activity diacylglycerol binding	membrane fraction
<u>206924_at</u>	interleukin 11	IL11	cell-cell signaling positive regulation of cell proliferation platelet activation B-cell differentiation megakaryocyte differentiation adipocyte differentiation	cytokine activity interleukin-11 receptor binding	extracellular region
<u>208902_s_at</u>	FLJ46061 protein	FLJ46061	---	---	---
<u>209024_s_at</u>	synaptotagmin binding, cytoplasmic RNA interacting protein	SYNCRIP	nuclear mRNA splicing, via spliceosome RNA splicing	RNA binding	nucleus endoplasmic reticulum ribonucleoprotein complex
<u>209252_at</u>	histidyl-tRNA synthetase-like	HARSL	protein biosynthesis histidyl-tRNA aminoacylation	histidine-tRNA ligase activity ATP binding ligase activity	cytoplasm
<u>212016_s_at</u>	polypyrimidine tract binding protein 1	PTBP1	nuclear mRNA splicing, via spliceosome RNA splicing	nucleic acid binding RNA binding poly-pyrimidine tract binding	nucleus nucleoplasm nucleolus heterogeneous nuclear ribonucleoprotein complex
<u>212986_s_at</u>	tousled-like kinase 2	TLK2	regulation of chromatin assembly or disassembly protein amino acid phosphorylation response to DNA damage stimulus cell cycle intracellular signaling cascade chromatin modification	protein serine/threonine kinase activity ATP binding transferase activity	nucleus
<u>214082_at</u>	carbonic anhydrase VB, mitochondrial	CA5B	one-carbon compound metabolism	carbonate dehydratase activity zinc ion binding lyase activity	mitochondrion

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 86</b>					
<u>203741_s_at</u>	adenylate cyclase 7	ADCY7	cAMP biosynthesis intracellular signaling cascade	magnesium ion binding adenylate cyclase activity guanylate cyclase activity lyase activity	integral to plasma membrane
<u>204122_at</u>	TYRO protein tyrosine kinase binding protein	TYROBP	cellular defense response intracellular signaling cascade	receptor activity receptor signaling protein activity	integral to plasma membrane
<u>204381_at</u>	low density lipoprotein receptor-related protein 3	LRP3	receptor mediated endocytosis	receptor activity	coated pit integral to membrane
<u>205339_at</u>	TAL1 (SCL) interrupting locus	SIL	cell proliferation	---	---
<u>207355_at</u>	solute carrier family 1 (glutamate transporter), member 7	SLC1A7	transport dicarboxylic acid transport	L-glutamate transporter activity symporter activity sodium:dicarboxylate symporter activity	membrane integral to membrane
<u>207606_s_at</u>	Rho GTPase activating protein 12	ARHGAP12	---	GTPase activator activity	---
<u>207723_s_at</u>	killer cell lectin-like receptor subfamily C, member 3	KLRC3	cellular defense response	transmembrane receptor activity sugar binding	integral to membrane
<u>209590_at</u>	Bone morphogenetic protein 7 (osteogenic protein 1)	BMP7	skeletal development growth	cytokine activity growth factor activity	---
<u>209751_s_at</u>	trafficking protein particle complex 2	TRAPPC2	skeletal development transcription regulation of transcription, DNA-dependent transport ER to Golgi transport	intracellular transporter activity transcription factor binding	intracellular endoplasmic reticulum Golgi apparatus
<u>210794_s_at</u>	maternally expressed 3 hypothetical gene supported by BX161452	MEG3 LOC440199	---	---	---
<u>211326_x_at</u>	hemochromatosis	HFE	protein complex assembly transport iron ion transport iron ion homeostasis receptor mediated endocytosis antigen presentation antigen presentation, endogenous antigen antigen processing, endogenous antigen via MHC class I	MHC class I receptor activity	cytoplasm integral to plasma membrane MHC class I protein complex
<u>212169_at</u>	FK506 binding protein 9, 63 kDa	FKBP9	protein folding	peptidyl-prolyl cis-trans isomerase activity calcium ion binding isomerase activity	endoplasmic reticulum
<u>212579_at</u>	KIAA0650 protein	KIAA0650	glycolysis chromosome organization and biogenesis	pyruvate kinase activity protein binding ATP binding	nucleus chromosome
<u>213143_at</u>	hypothetical protein LOC257407	LOC257407	---	---	---
<u>213433_at</u>	ADP-ribosylation factor-like 3	ARL3	G-protein coupled receptor protein signaling pathway small GTPase mediated signal transduction	signal transducer activity GTP binding	---

ProbeSet ID	Gene Title	Gene Symbol	GO Biological Process	GO Molecular Function	GO Cellular Component
<b>Mgene 35</b>					
<u>202539_s_at</u>	3-hydroxy-3-methylglutaryl-Coenzyme A reductase	HMGR	lipid metabolism cholesterol biosynthesis germ cell migration gonad development biosynthesis	hydroxymethylglutaryl-CoA reductase (NADPH) activity oxidoreductase activity	membrane fraction peroxisome endoplasmic reticulum membrane integral to membrane
<u>203766_s_at</u>	leiomodlin 1 (smooth muscle)	LMOD1	---	tropomyosin binding	membrane fraction cytoskeleton
<u>206912_at</u>	forkhead box E1 (thyroid transcription factor 2)	FOXE1	negative regulation of transcription from Pol II promoter regulation of transcription, DNA-dependent transcription from Pol II promoter ectoderm development morphogenesis	transcription factor activity RNA polymerase II transcription factor activity	nucleus
<u>211165_x_at</u>	EPH receptor B2	EPHB2	protein amino acid phosphorylation transmembrane receptor protein tyrosine kinase signaling pathway neurogenesis	protein serine/threonine kinase activity protein-tyrosine kinase activity receptor activity ephrin receptor activity transmembrane-ephrin receptor activity ATP binding transferase activity	membrane integral to membrane
<u>213575_at</u>	Transformer-2 alpha	TRA2A	nuclear mRNA splicing, via spliceosome	RNA binding pre-mRNA splicing factor activity	nucleus

**Supplementary Table 3. Logistic regression analysis on the Duke (n = 89), ACOSOG (n = 25), and CALGB cohorts (n = 84) demonstrating the added prognostic value (OR – odds ratio) of the genomic model in comparison to known clinical predictors of outcome.**

**Duke Cohort (n = 89)**

Predictor	Univariate OR [95%CI]	p-value	Multivariate OR [95%CI]	p-value
Age (yrs)	1.01 [0.96, 1.05]	0.750	n/a	---
Gender: Male	1.34 [0.56, 3.18]	0.509	n/a	---
Cell Type: Squamous cell	1.10 [0.47, 2.53]	0.831	n/a	---
Tumor size >3cm	2.29 [0.98, 5.39]	0.057	2.99 [0.52, 17.10]	0.219
Lymph node status: Positive	2.47 [0.88, 6.89]	0.085	4.87 [0.69, 34.33]	0.112
Pathologic stage: Non-stage I	3.28 [1.13, 9.48]	0.029	7.27 [1.02, 51.80]	0.048
Genomic prediction (>50% probability of recurrence)	137 [29, 650]	<0.0001	207 [32, 1348]	<0.0001

**ACOSOG Cohort (n = 25)**

Predictor	Univariate OR [95%CI]	p-value	Multivariate OR [95%CI]	p-value
Age (yrs)	0.97 [0.89, 1.04]	0.391	n/a	---
Gender: Male	2.57 [0.47, 14.10]	0.277	n/a	---
Cell Type: Squamous cell	1.20 [0.25, 5.84]	0.821	n/a	---
Tumor size (> 3cm)	0.84 [0.49, 1.44]	0.529	n/a	---
Lymph node status: Positive	3.13 [0.47, 20.58]	0.236	n/a	---
Pathologic stage: Non-stage I	3.13 [0.47, 20.58]	0.236	n/a	---
Genomic prediction (>50% probability of recurrence)	35.90 [2.78, 463]	0.006	35.90 [2.78, 463]	0.006

**CALGB Cohort (n = 84)\*\***

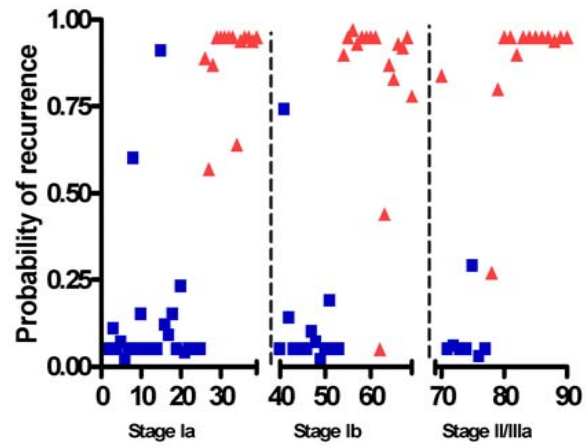
Predictor	Univariate OR [95%CI]	p-value	Multivariate OR [95%CI]	p-value
Age (yrs)	1.02 [0.98, 1.07]	0.304	n/a	---
Gender: Male	5.33 [1.87, 15.18]	0.002	4.95 [1.35, 18.116]	0.016
Pathologic stage: Non-stage I	3.14 [1.26, 7.87]	0.014	4.15 [1.29, 13.41]	0.018
Lymph node status: Positive	3.45 [1.27, 9.37]	0.015	4.59 [1.32, 15.98]	0.017
Genomic prediction (>50% probability of recurrence)	16.1 [4.78, 54.23]	<0.0001	15.81 [4.19, 59.75]	<0.0001

*In both the Duke and CALGB datasets tumor size & lymph node status were analyzed separately from non-stage I variable in the multiple logistic regression analysis due to co-linearity.*

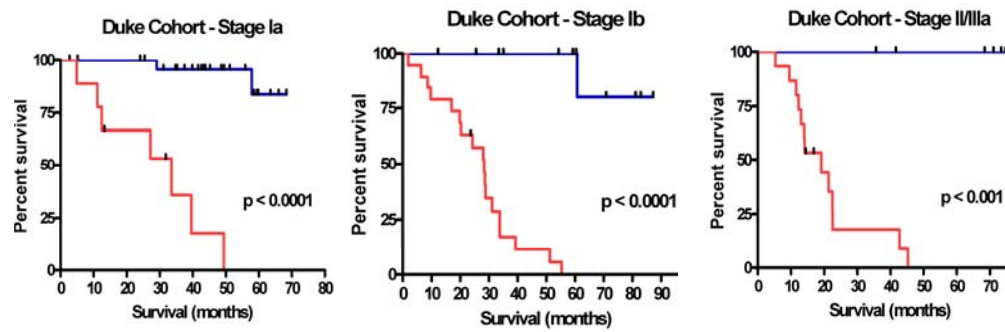
*\*\*Exact tumor size data was not available for the CALGB data. All samples included in the CALGB cohort were adenocarcinomas.*

# Supplementary Figure 1

A.

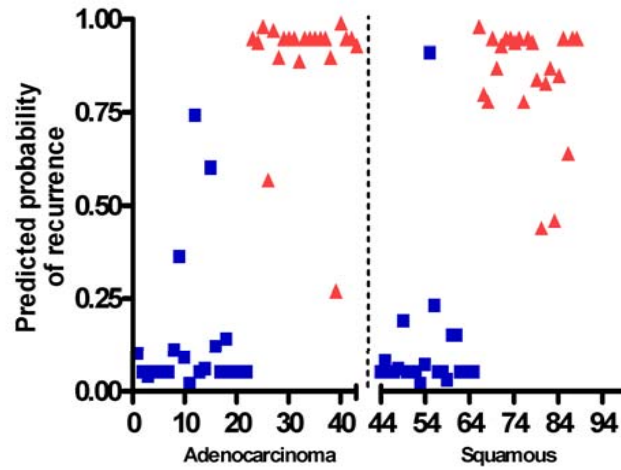


B.

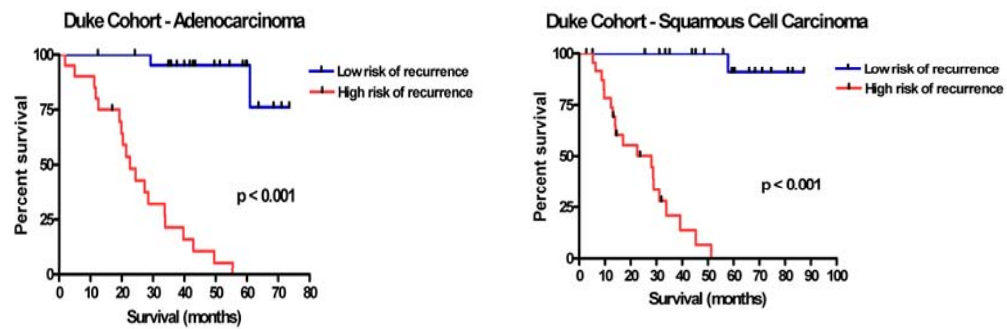


## Supplementary Figure 2

A.



B.



### Supplementary Figure 3

