Discovery of Biomarkers
Pharmaceutical Users Software Exchange

Nusrat Rabbee, PhD

September 26 2013

Professional Background in Biomarker Development/ Bioinformatics

- 10 years experience in biomarker development in drug and molecular diagnostics industry
- Academia: Co-inventor of the BRLMM algorithm used by Affymetrix in accurate calling SNP genotypes for their SNP arrays (2005)
- Industry: Co-inventor of Affirma – a molecular diagnostics test by Veracyte in accurately predicting thyroid cancer from thyroid biopsies

Professional Background in Biomarker Development (cont'd)

- Industry: Analytics for T-DM1 – the new ADC compound in breast cancer or new generation Herceptin™ at Genentech
- Industry: Analytics for anti-IL13 compound Lebrikizumab – a humanzied monoclonal antibody with high specificity and affinity for IL13. Anti-IL13 activity can impact multiple pathways in asthma. Periostin is a serum based biomarker of Il-13 activity.
- Industry: Analytics for cMET drug-diagnostic combination in NSCLC at Lilly

Why look for a companion diagnostic?

- Personalized Medicine – using diagnostics to identify the members of a population that will best benefit from a targeted therapy
- Particularly
  - Compound is directed to a specific target
    - Trastuzumab/HER2 vs taxol or cisplatin
  - Target is only expressed by a subset of patients (<50%)
    - HER2 is 20% of invasive breast cancer,6-8% of gastric cancer
    - EML4-ALK mutation in 3-5% of NSCLC/crizotinib
    - BCR-AbI mutation in 90% of CLL
  - Benefit is associated with modulating the specific target

Why look for a companion diagnostic? – cont'd

- Benefits to Drug Development
  - Faster proof of concept
  - Higher response rates
  - Smaller trials, lower cost, faster recruitment

Biomarker Platforms for Biomarkers/ Molecular Diagnostics are heterogeneous

- Blood, plasma or serum biomarker
  - PSA in prostate cancer
  - CA125 in ovarian cancer
  - Circulating tumor cells
- Genetic Test (PCR, FISH, etc) of tumor
  - EGFR mutation in lung and colon cancer
  - ALK mutation in melanoma
  - BRAF mutation in melanoma
  - KRAS mutation in colon cancer

Biomarker Platforms for Biomarkers/ Diagnostics heterogeneous – cont'd

- Immunohistochemistry of tumor
  - Estrogen receptor in breast cancer
  - Herceptest in breast cancer
- High Throughput Tumor Marker Profiling
  - Gene expression pattern
  - SNP expression pattern
  - High throughput sequencing

Co-development of Assay Development in Drug Development

- Preclinical/Translational Research – Biomarker Discovery and Hypothesis Generation
- Phase 1 – Dose escalation for safety and early signs of efficacy
  - Feasibility: include a wide range of diagnostic assays to look for limitations and associations
  - Define and finalize key reagents (antobodies, primers, etc)
  - Develop prototypes

Co-development of Assay Development in Drug Development – cont'd

- Phase 2a – Proof of concept
  - Design optimization: include assays as research kits in their final form
  - Assay all comers (or enriched populations) to identify the correct population to target / refine assay cutpoint
- Phase 2b – Dose ranging to optimize the therapeutic window
  - Mostly in non-oncology studies
  - Design in verification: Finalize assay SOP and define target population
- Phase 3 – Validation of efficacy and safety
  - Design validation: validate assay and the association of efficacy with targeting the population *

Bioconductor
Open Source Software for Bioinformatics

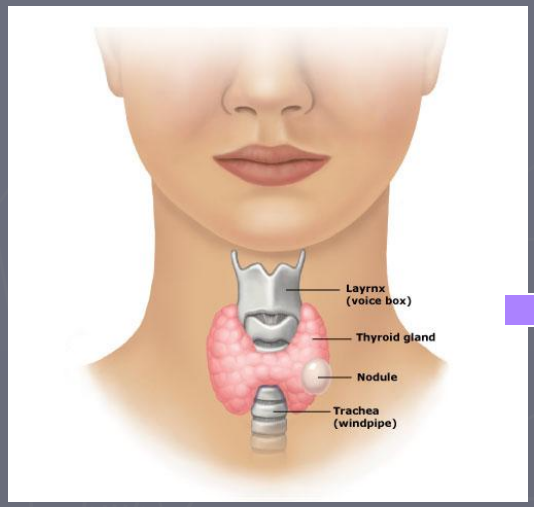
- Advanced analytics facilities for microarray platforms (e.g., Affymetrix, Illumina, Nimblegen, Agilent, etc)
- Platforms
  - Affymetrix 3’ biased arrays
  - Affymetrix Exon ST arrays
  - Affymetrix Gene ST Arrays
  - Affymetrix SNP arrays
  - Nimblegen Arrays
  - Illumina Expression Microarrays
  - More…
- Ref: http://www.bioconductor.org/help/workflows/oligo-arrays/

DEVELOPING HIGH DIMENSIONAL BIOMARKER ALGORITHM

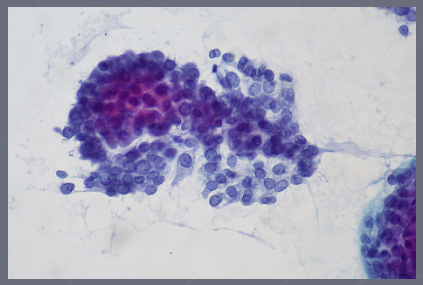
High Dimensional Biomarker Algorithm

- I headed up Biostatistics and Bioinformatics Team at Veracyte
- Objective
  - Discriminate between B and M in thyroid tumor biopsy samples
  - Develop algorithm & assay into a commercial product
  - Design clinical trials (V & V) and regulatory strategy

Evaluation of Thyroid Nodule – Clinical Process

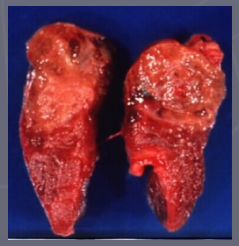


Fine Needle Aspiration (FNA)

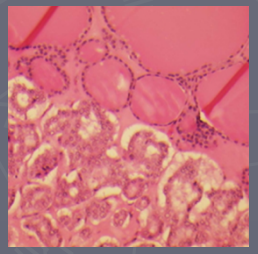


Papillary Thyroid Carcinoma (PTC) from cytology

Thyroidectomy (Surgery)



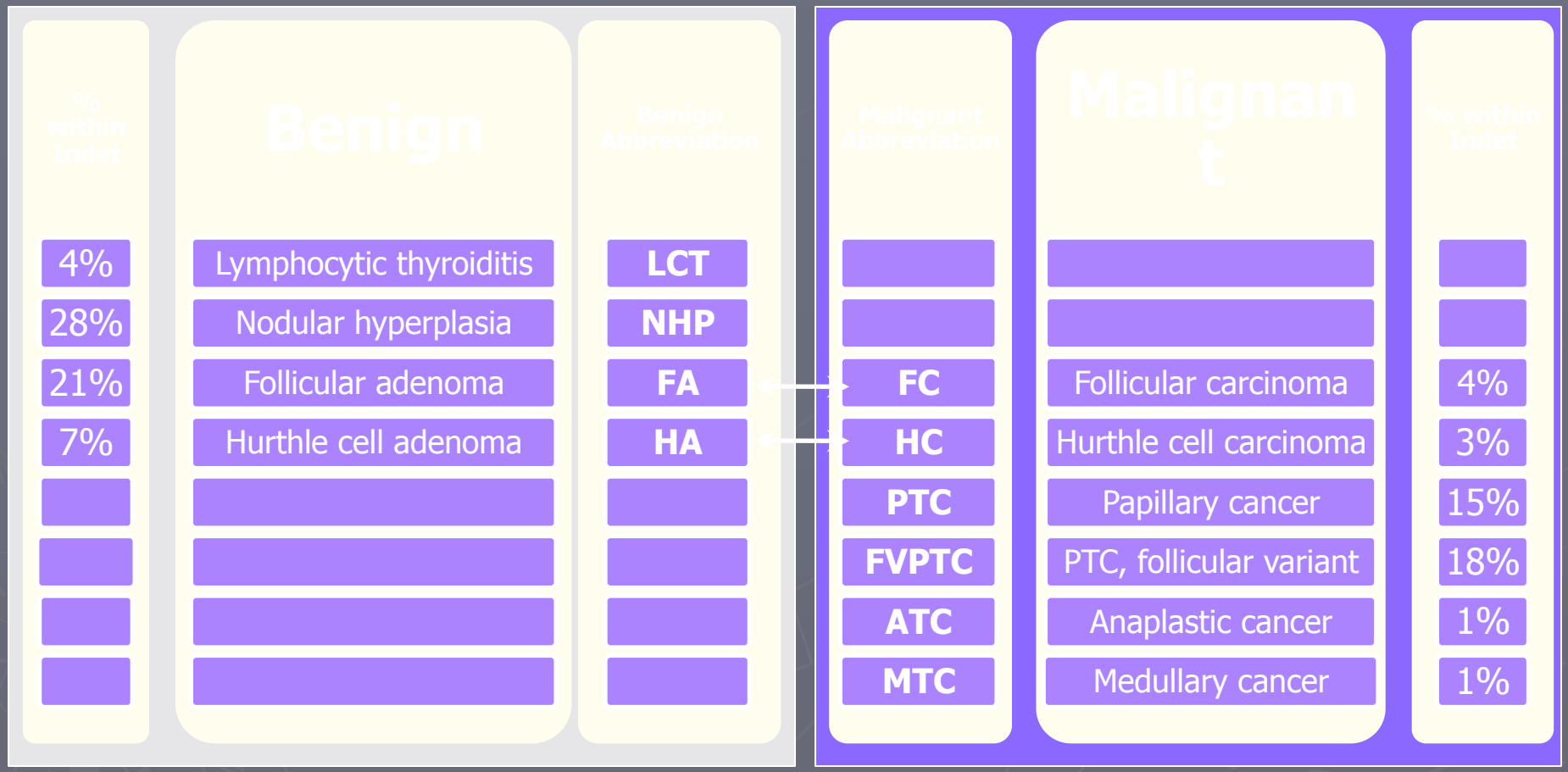
PTC after resection



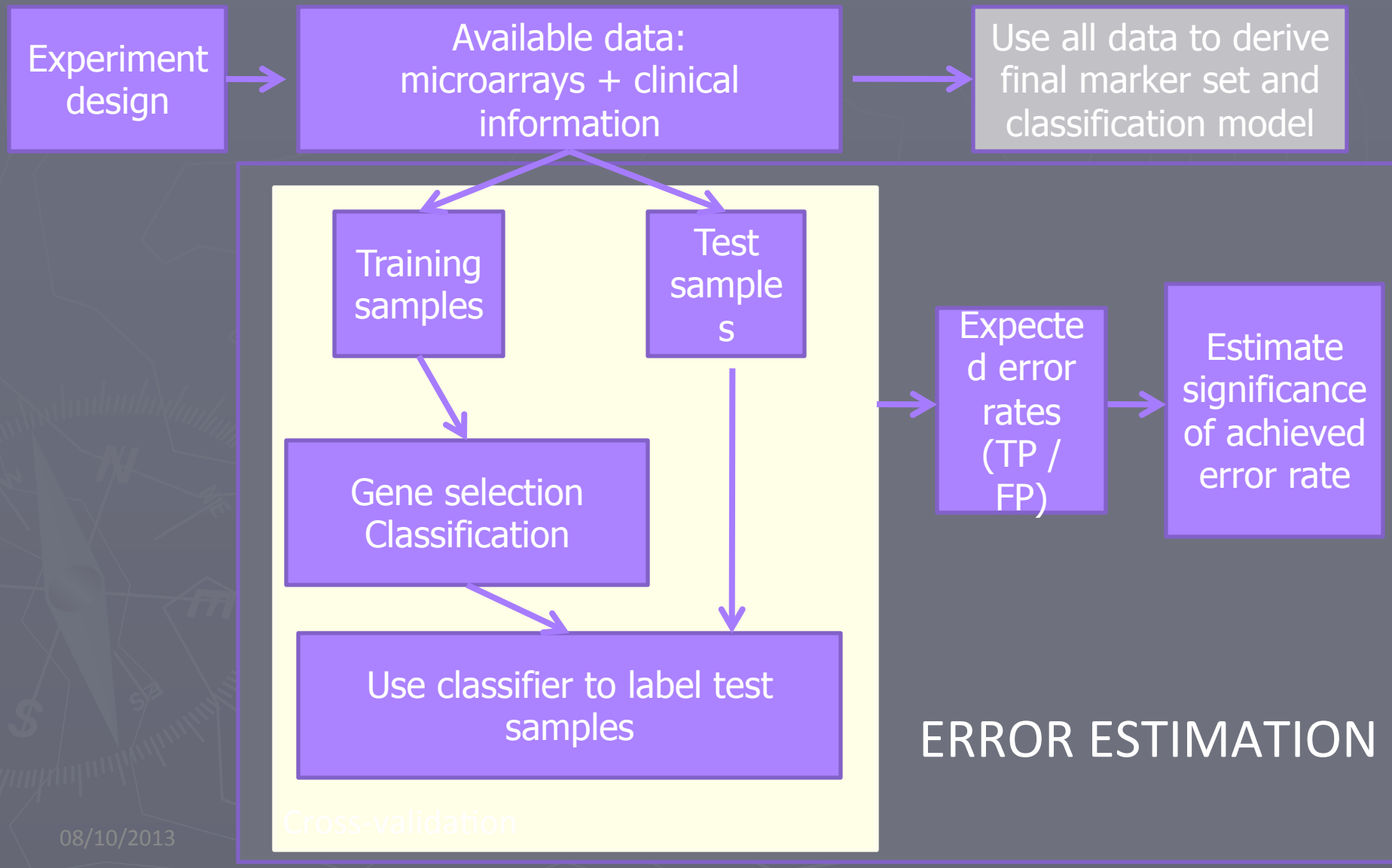
PTC on histology

Source - http://www.uptodateonline.com/online/content/images/endo_pix/Thyroid_nodule.jpg; http://medicalimages.allrefer.com/large/cervix-needle-sample.jpg; http://farm4.static.flickr.com/3207/3153794593_0abe7e7a84.jpg; http://www.medgadget.com/archives/thyroidectomy.jpg; http://www.pathology.med.ohio-state.edu/paxit/depthase/Paxit/Images/10531/PAXIT009.JPG

Subtypes of thyroid nodules



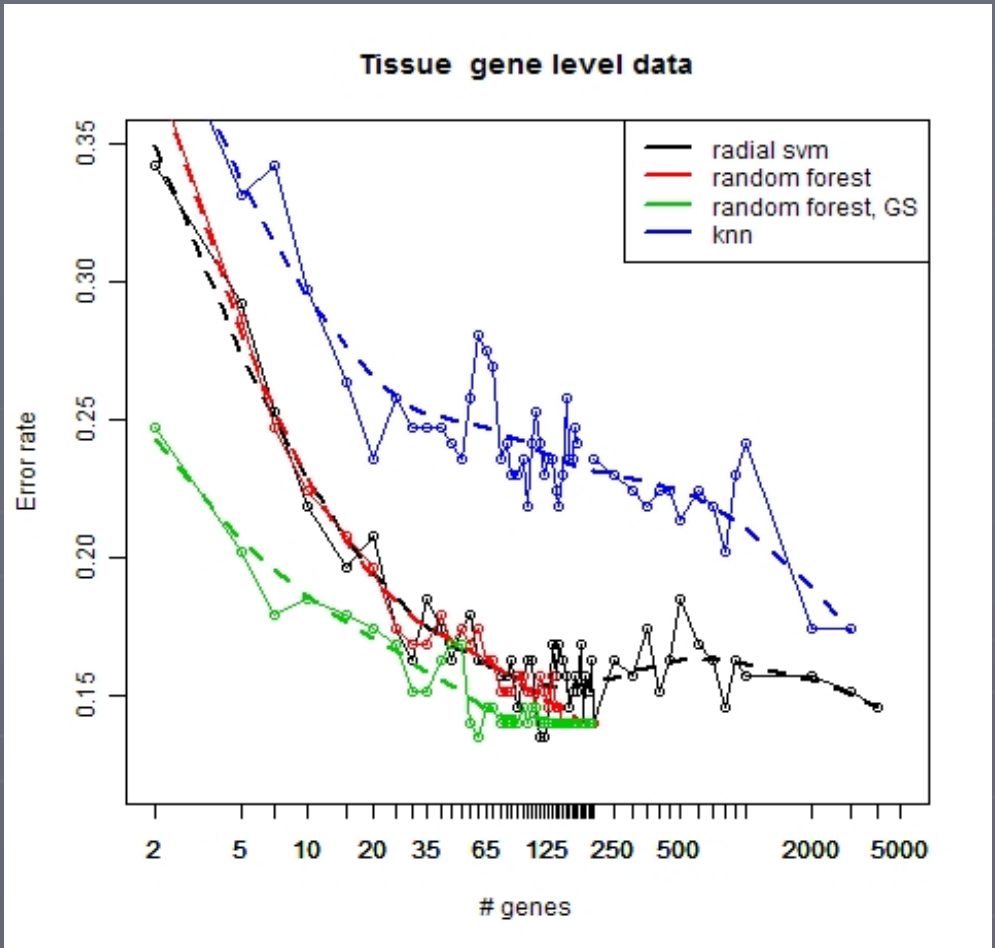
NEPTUNE Classification framework: single classifier



High Dimensional Biomarker Algorithm – R&D

- Research & Discovery
  - Platforms (Exon array (Affy), miRNA array (Agilent and Illumina), DNA array (Affy 6.0))
  - Literature markers available on thyroid cancer ("candidate gene" approach)
  - Obtained Tissue samples for initial discovery ("whole genome" approach)
  - Obtained FNA samples for initial discovery
- Discovery & Feasibility
  - Measured classifier performance on banked FNA samples
  - Marker transferability from Tissue to Banked FNA
  - Measures taken to improve FNA sample quality
  - Measured classifier performance on prospective FNA samples
    - Experimental Design
    - Key classification results
- Development
  - Clinical Trial Design
  - Obtain FDA blessing for trial
  - Refine Assay Development
  - Design custom array
  - Develop software (VTM)

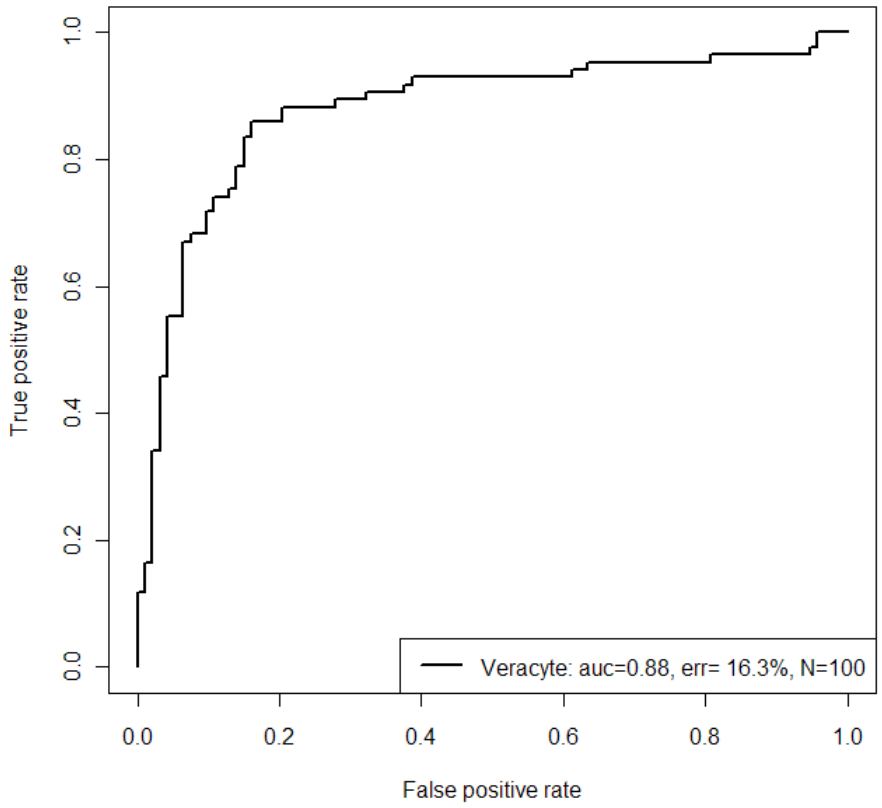
Classifier performance on Tissue data



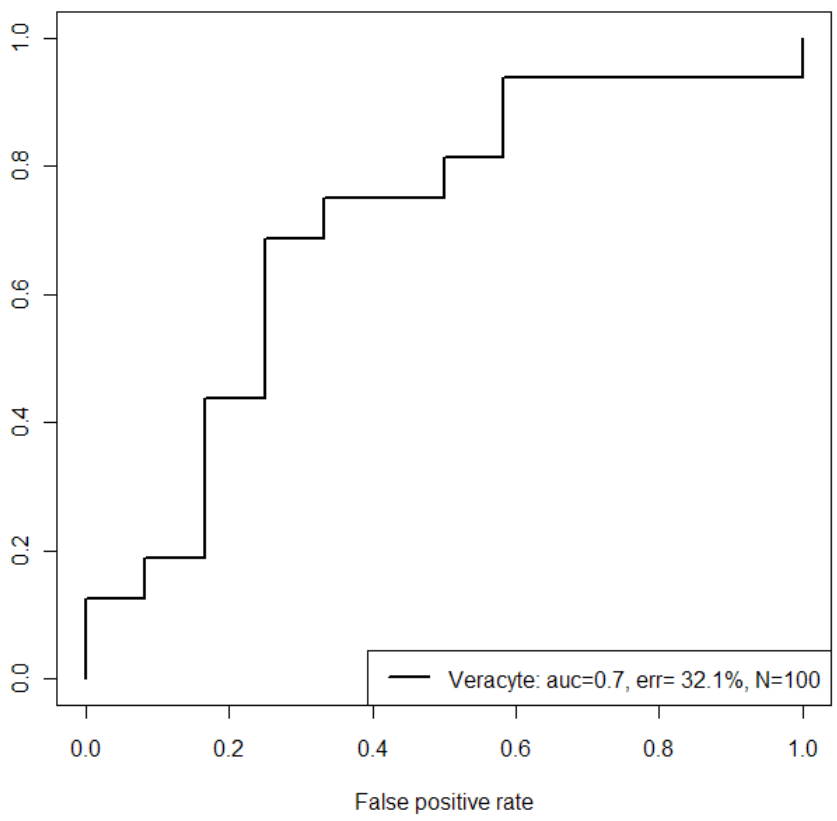
Performance in good-quality FNAs (Ex/In > 0.75) is poor

- Excluding poor-performing arrays (with Exon / intron separation below 0.75) still results in inferior classification performance on FNA data

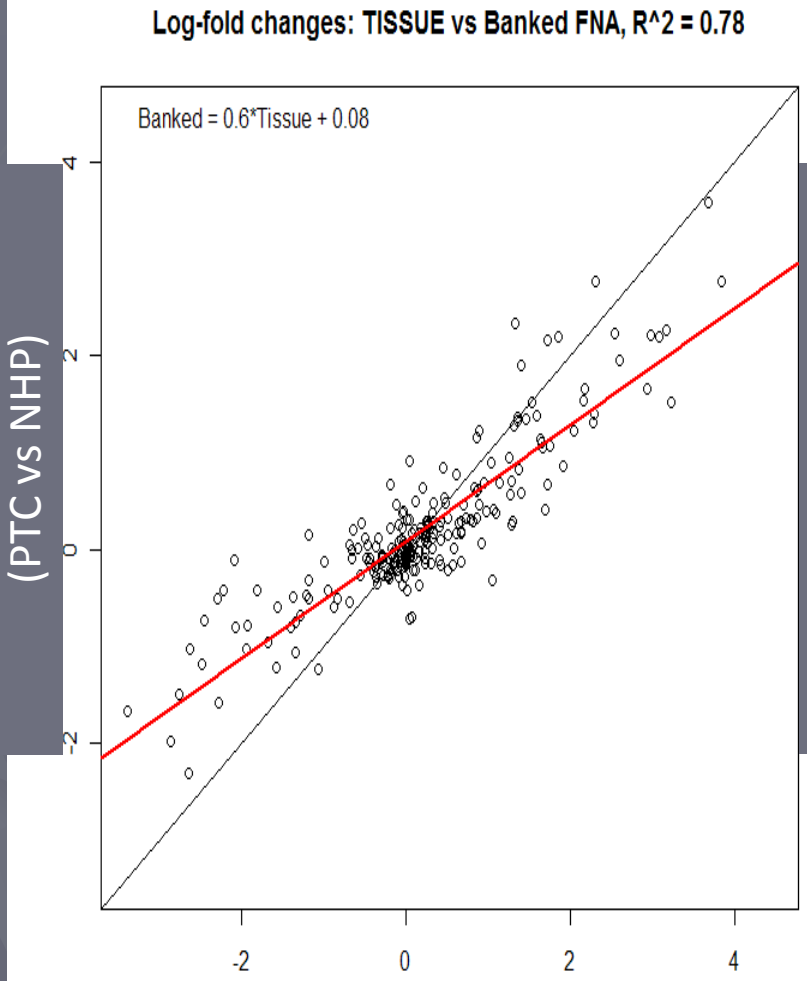
Tissue data (Gene level)

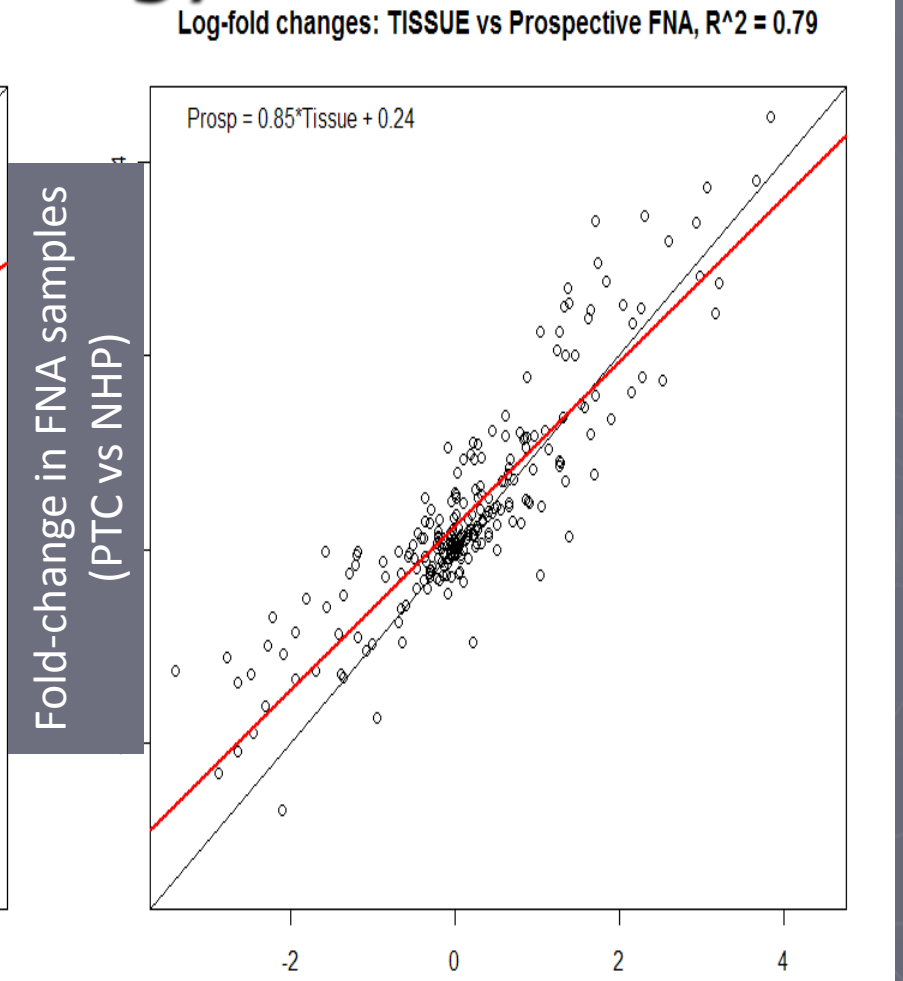


FNA data (Gene level)



Still…Preliminary evidence for similarity of biology



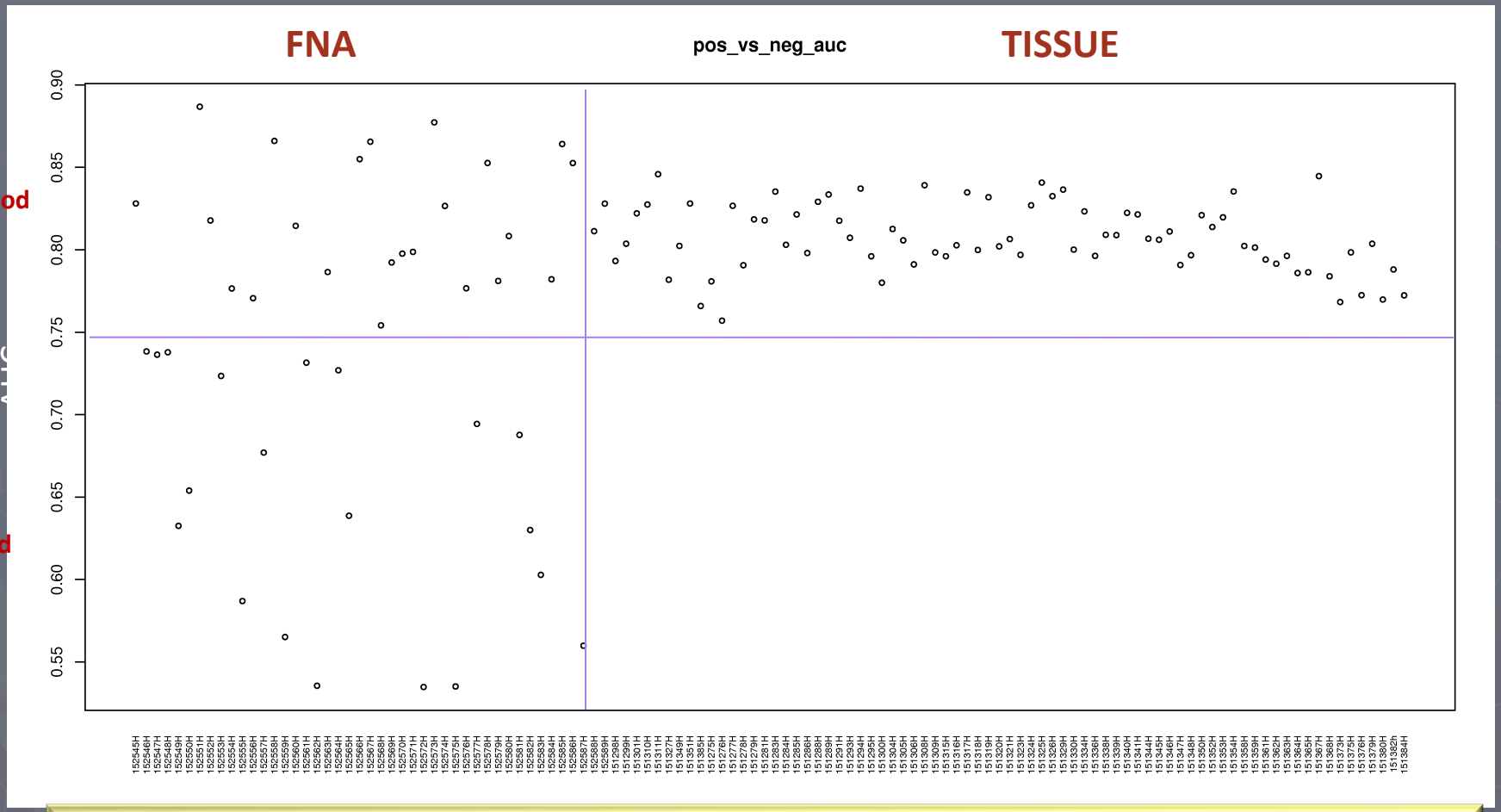


A Few Early Questions: the FNA story

- Question 1: Degradation Impact
  - How does degradation affect performance and how can information be extracted from degraded samples?
    - Pre-hybridization
      - RNA Degradation (RIN)
      - RNA concentration/yield
    - Post-hybridization
- Question 2: Transferability of Biology
  - Do intact / mildly degraded FNA carry the same biological signal as surgical tissues?
- Question 3: Marker Selection
  - How can we leverage disparate sample sources for marker selection?

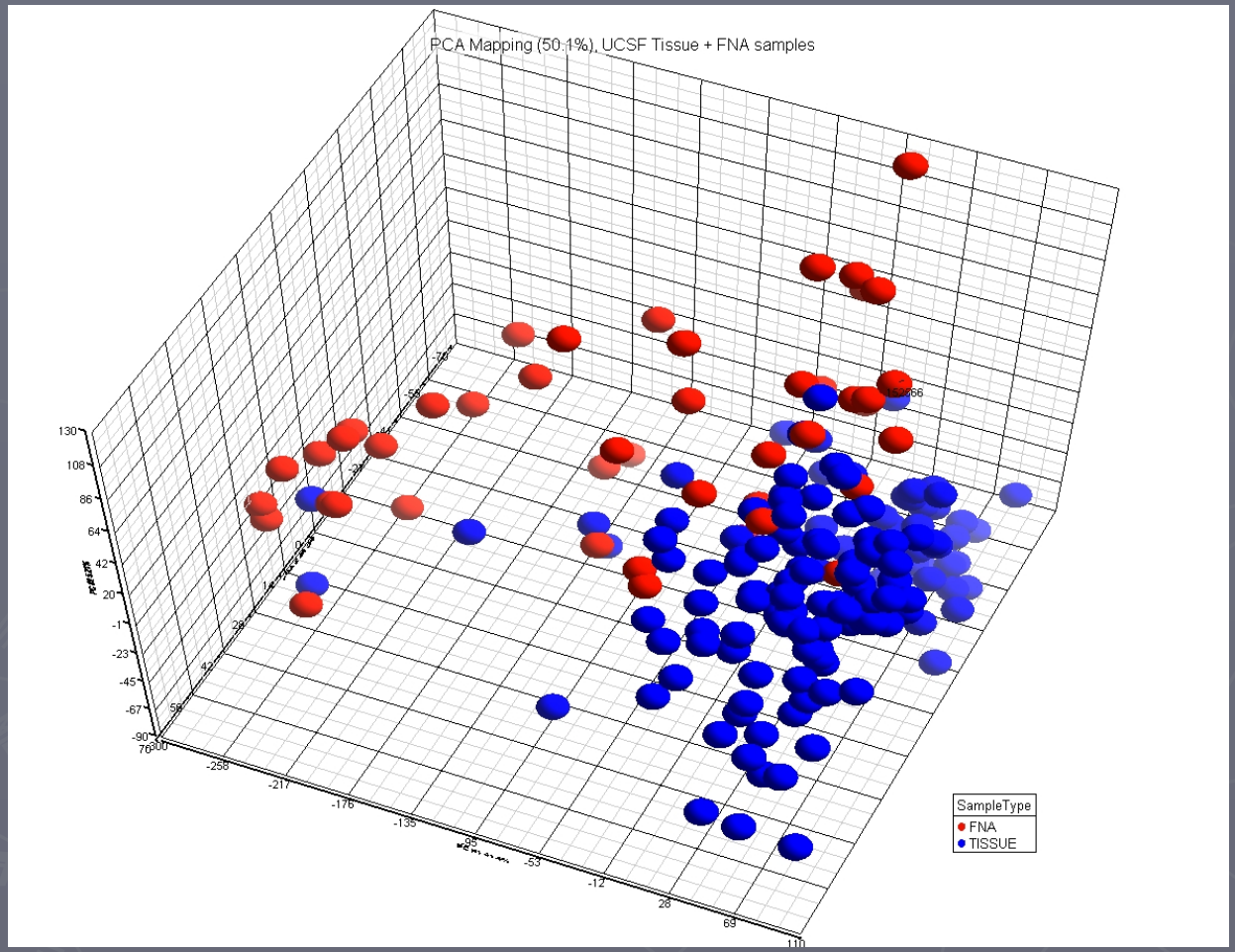
Answer: Collect more FNA samples, design experiment on FNA samples, make improvements in assay and classifier

Comparison of AUC Values for FNA and Tissue Samples



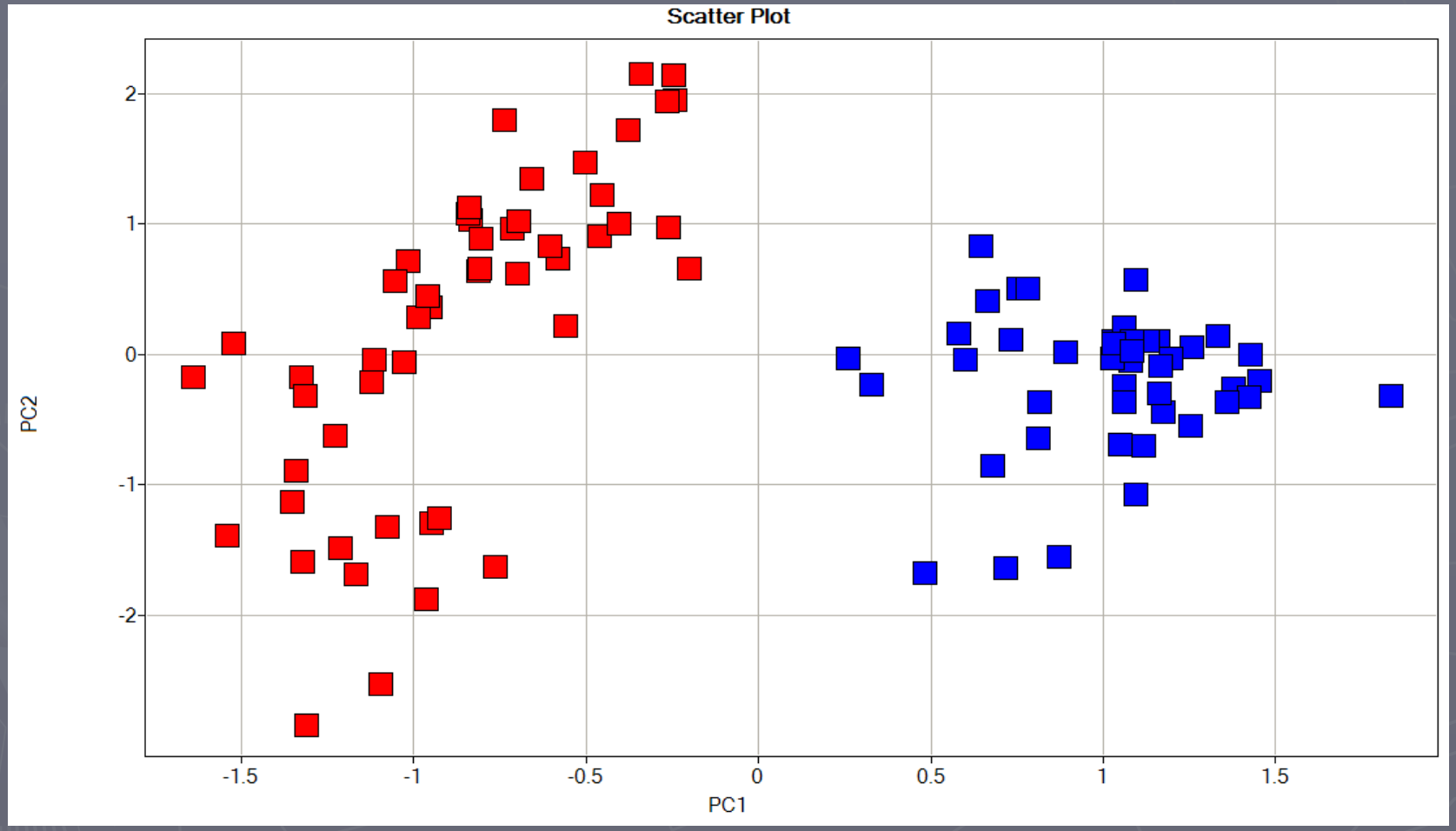
FNA samples show higher variability than tissues

PCA on FNA and Tissue Samples Using All Genes show large differences (PARTEK)



Large differences between FNAs and tissue: Biological and/or Technical?

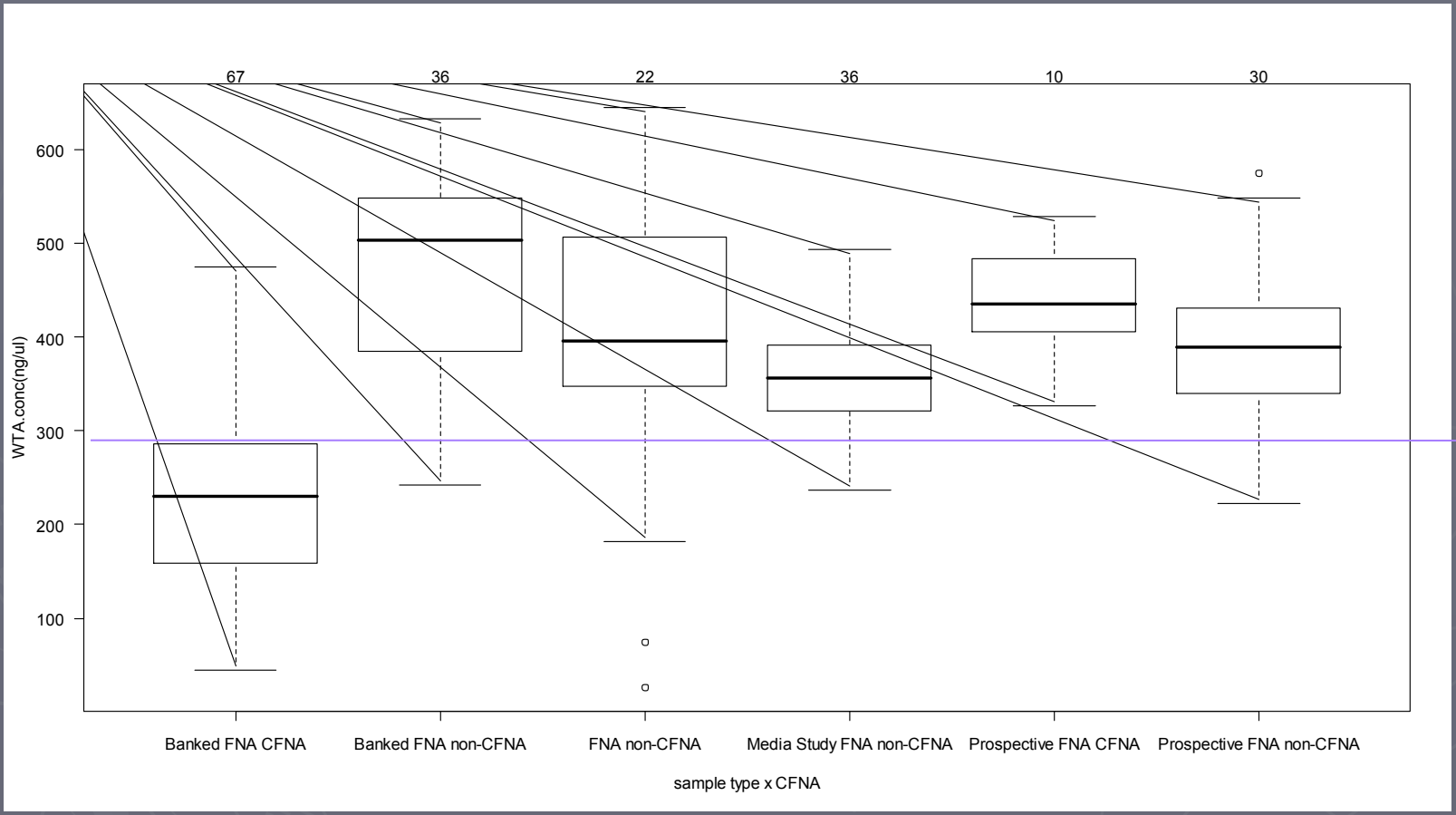
FNA Data impacted heavily by site-to-site differences (SpotFire)



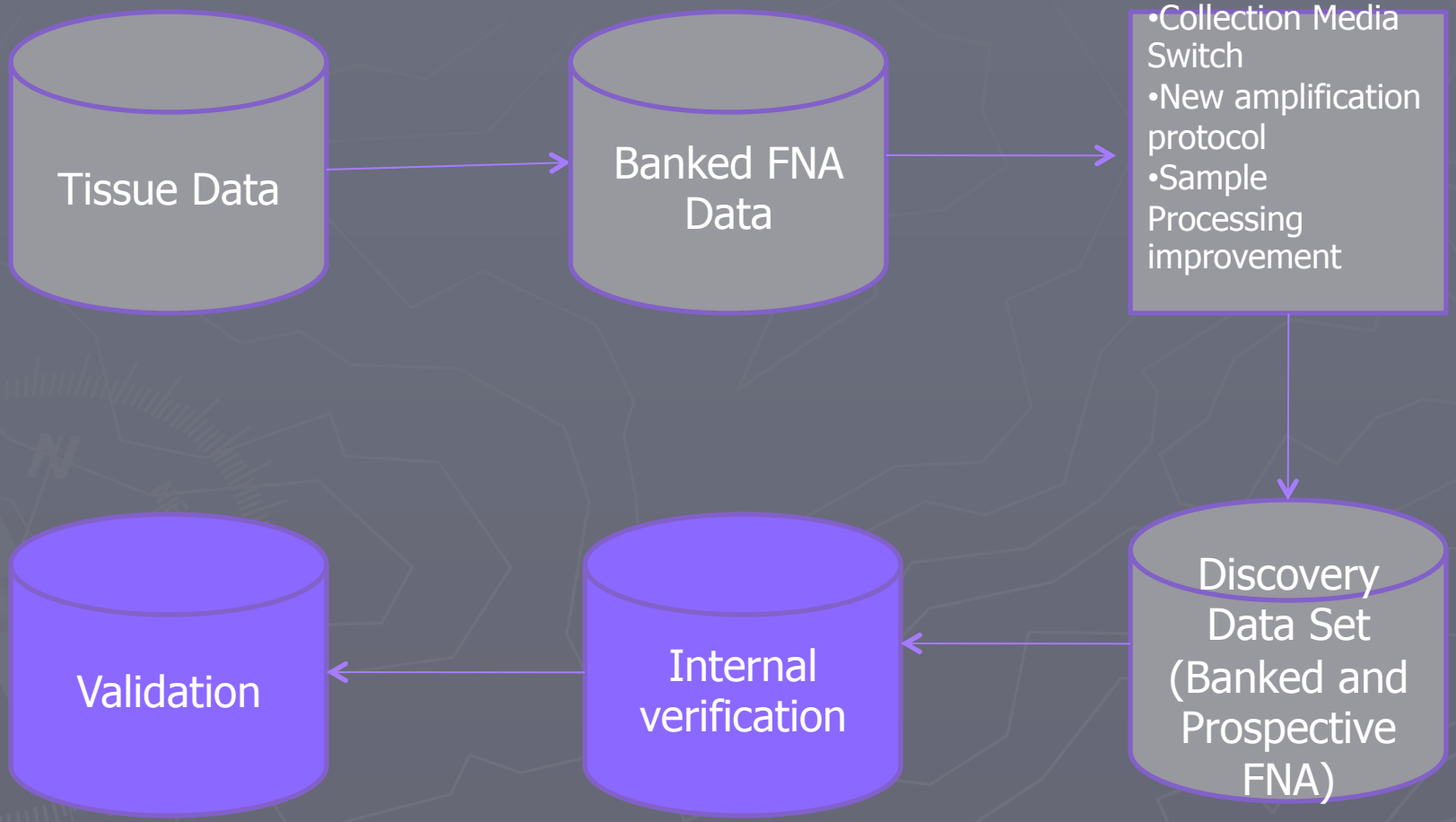
Site 1

Site 2

FNA Sample - Yields after amplification shows marked difference from banked FNA samples of one investigator vs rest



Data sets: Discovery, Verification, Validation



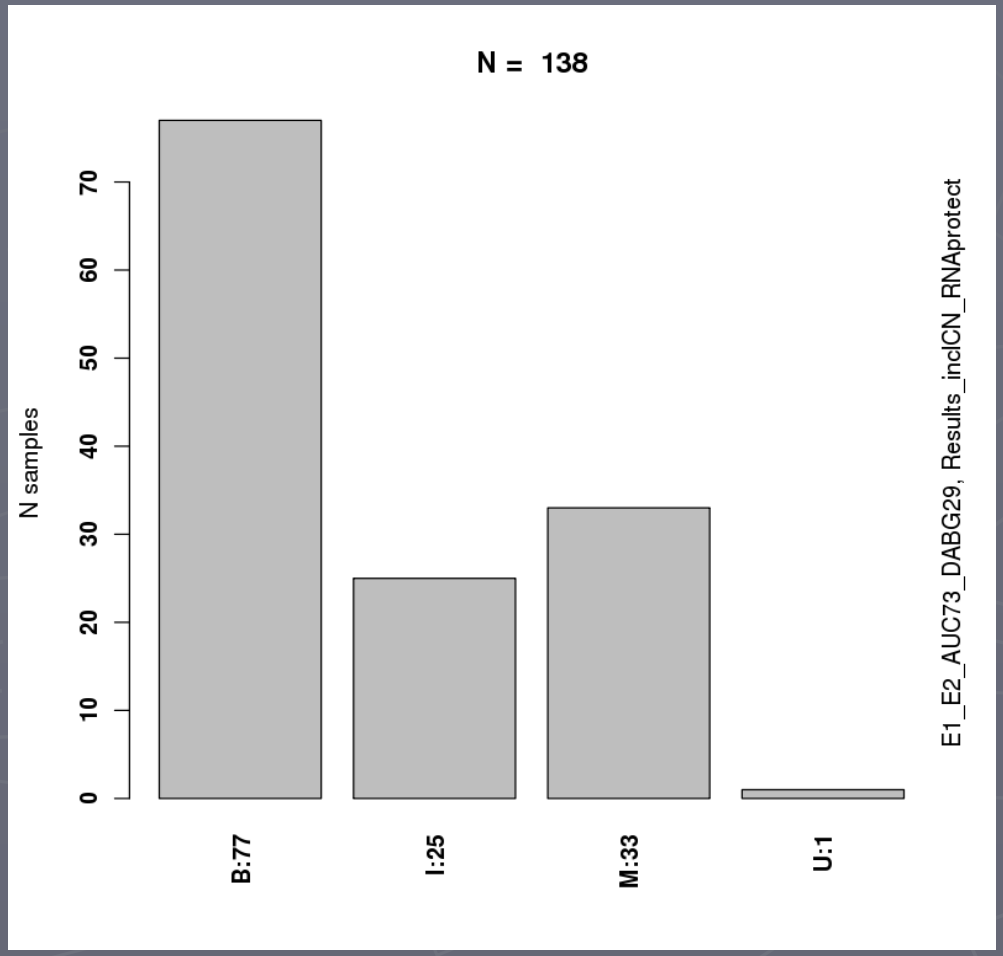
Statistical Challenges

- Decompose sources of gene expression variation in complex clinical and experimental settings
  - Known factors (yield, 260:280 ratio, RIN)
  - Unknown Factors
- Produce single marker list from several biological sources
  - Feature selection via a repeatability along with technical factor removal
  - Combination of three separate methods and sources of discovery
- Produce GOLD STANDARD in Pathology Labels

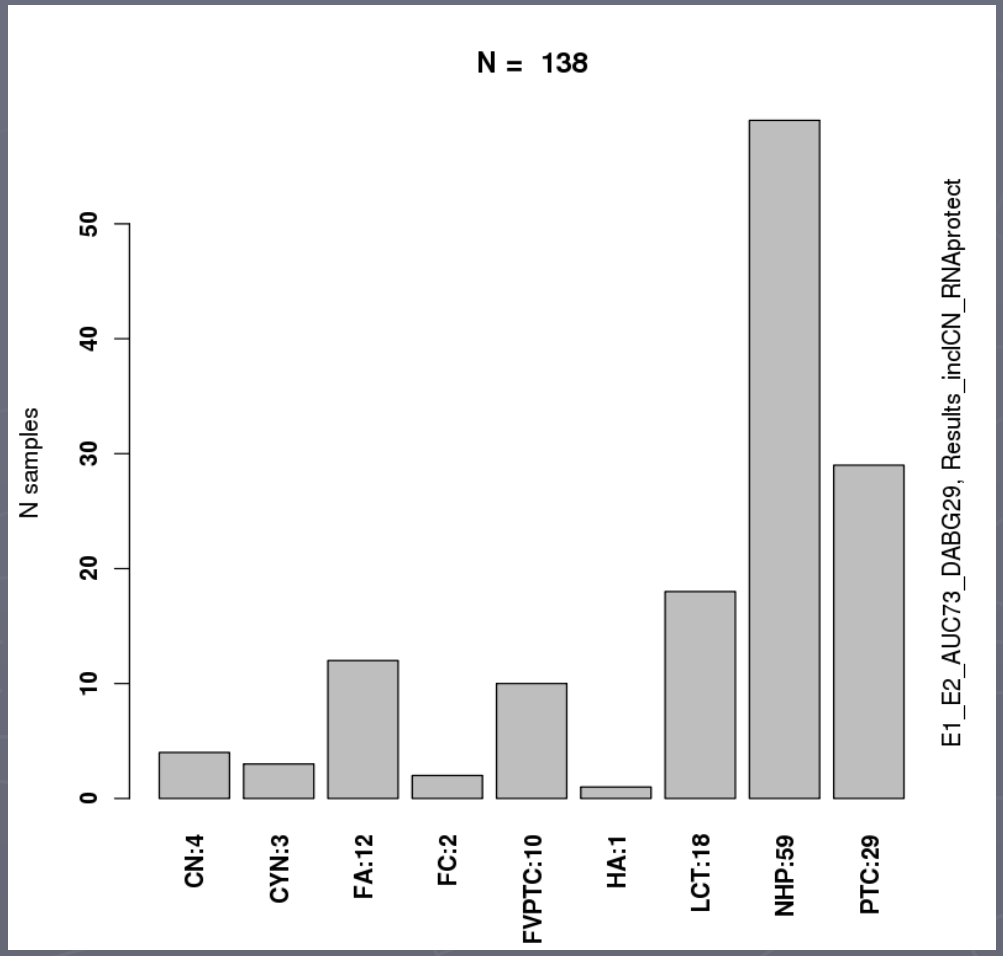
After all the improvements…

ADJUDICATED LABELS, RNA PROTECT ONLY (N = 138) 2009/11/04

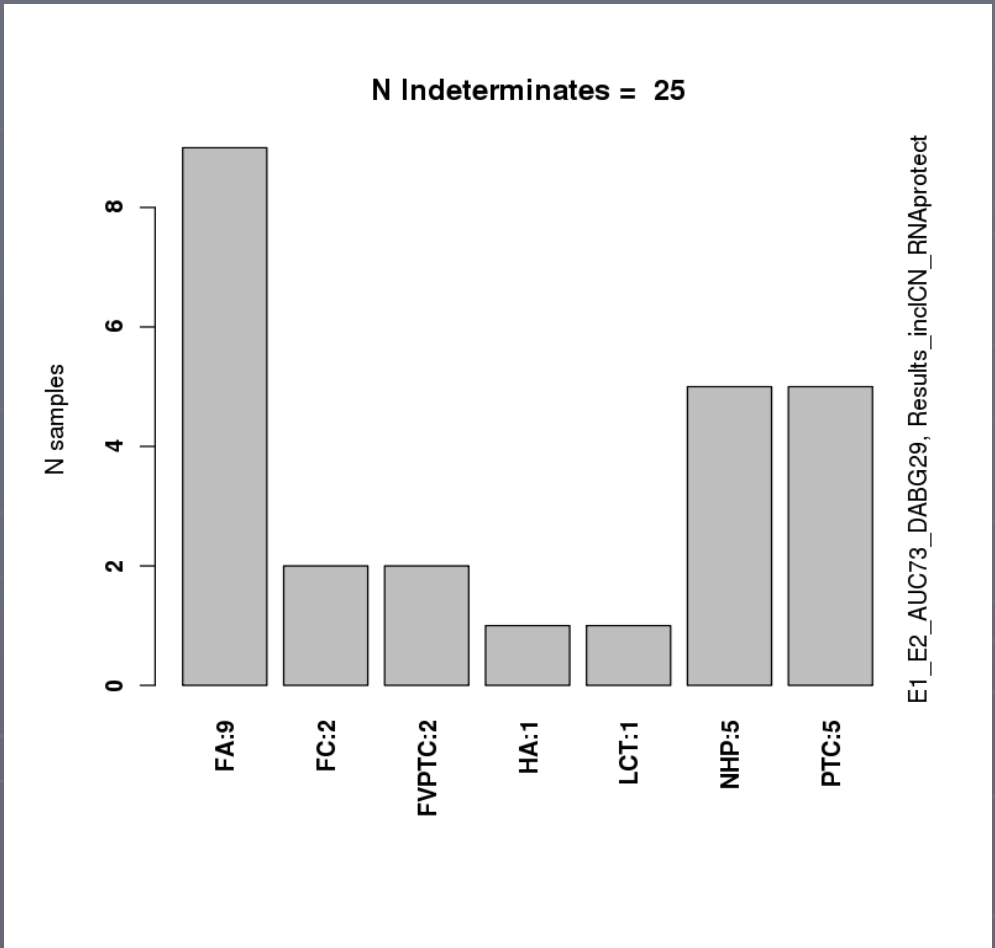
Cytology B/I/M/U



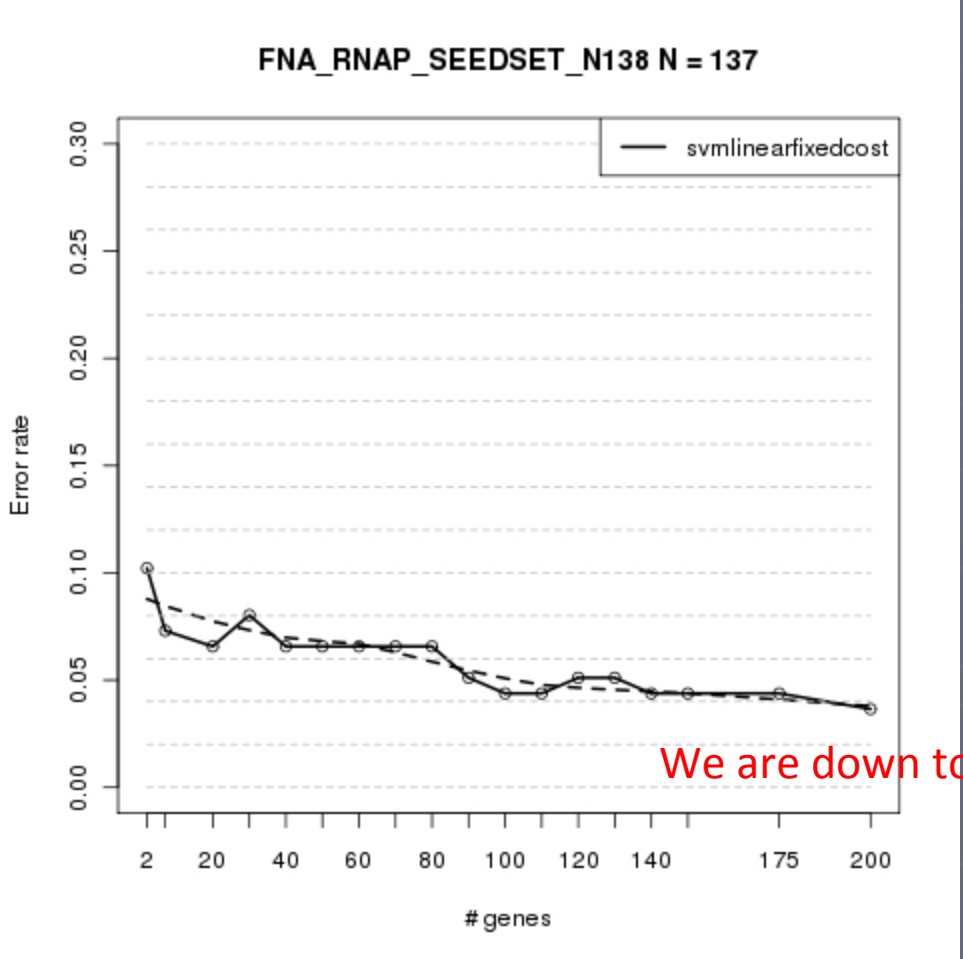
Pathology


N = 138
N samples
0
10
20
30
40
50
CN:4
CYN:3
FA:12
FC:2
FVPTC:10
HA:1
LCT:18
NHP:59
PTC:29
E1_E2_AUC73_DABG29; Results_inclCN_RNAprotect

Pathology within I group



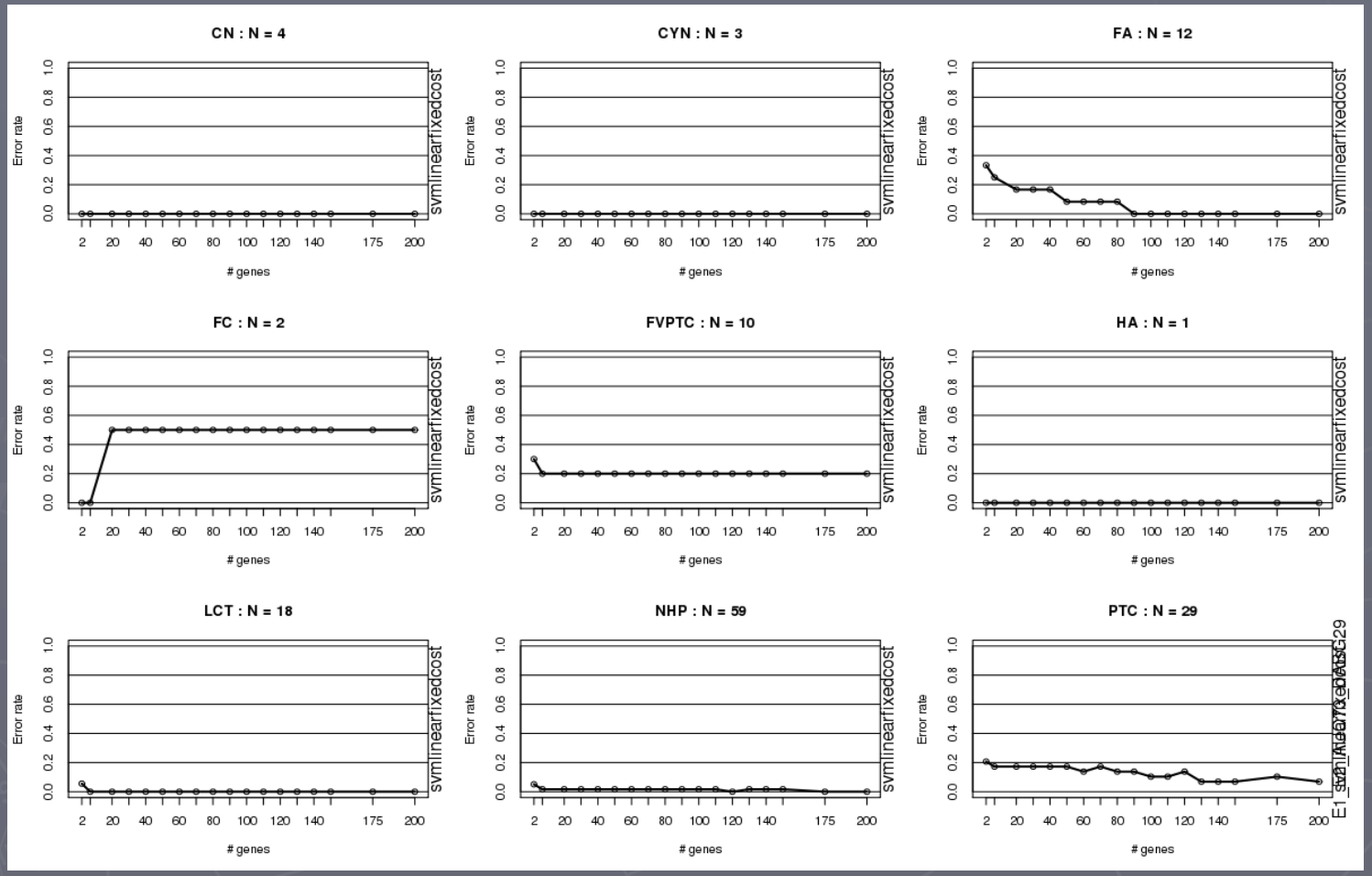
N137 Error Titration Curve SVM using K=230 (Repeatable FS)



We are down to 4% error rate!

BN	CN	CYN	FA	FC	FVPTC	HA	LCT	NHP	PTC
22	4	3	10	3	7	2	18	37	31

SVM linear fixed cost



Performance by subtype (n=138)
No adjustment for subtype prevalence
Threshold = .5

	Total	Benign						Malignant					Other
		FA	HA	LcT	NHP	CN	CYN	FC	PTC	fvPTC	MTC**	HC	
Zeiger Est. Prevalence among indeterminates(%)	**100**	**21**	**7**	**4**	**28**			**3.5**	**15**	**14**	**1.5**	**2.0**	**4.0**
Specimens tested	138	12	1	18	59	18	3	2	29	10	6	tbd	n/a
VTM test result is benign	102	12	1	18	59	18	3	1	2	2	0	tbd	n/a
VTM test result is continued indeterminate	36	0	0	0	0	0	0	1	27	8	6	tbd	n/a
%Accuracy	**96**	100	100	100	100	100	100	50	93	80	100	tbd	n/a

Submitted to FDA in December 2009

Adjusting for subtype prevalence = TRUE

Using subtype prevalence as in Zeiger

Sensitivity = 95.1 %

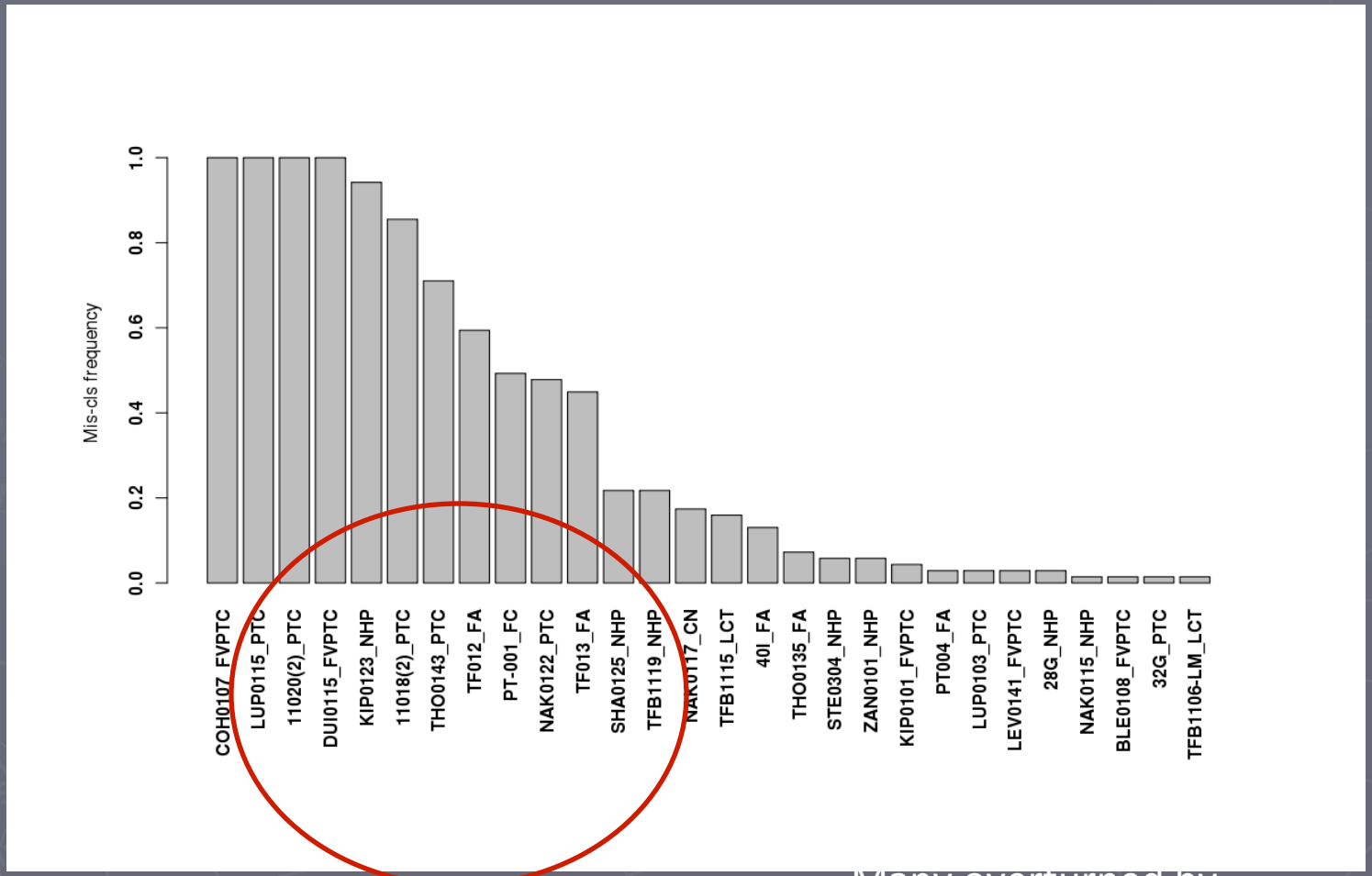
Specificity = 77.9 %

NPV = 97.4 %

PPV = 64.8 %

ights for
ld
calculation

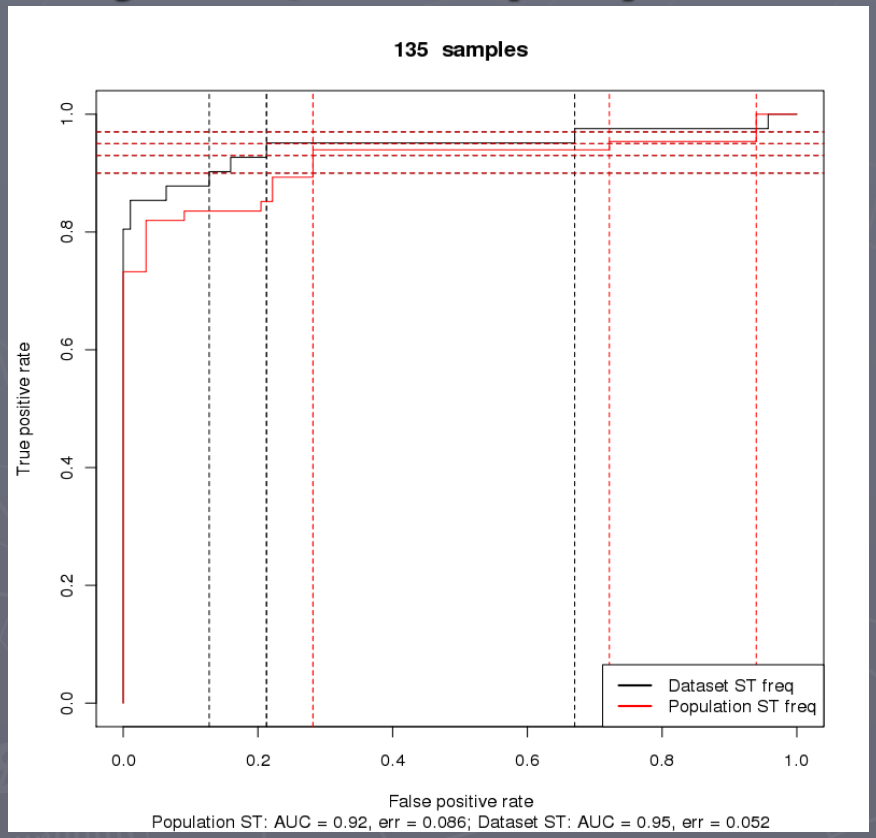
Mis-classification frequency



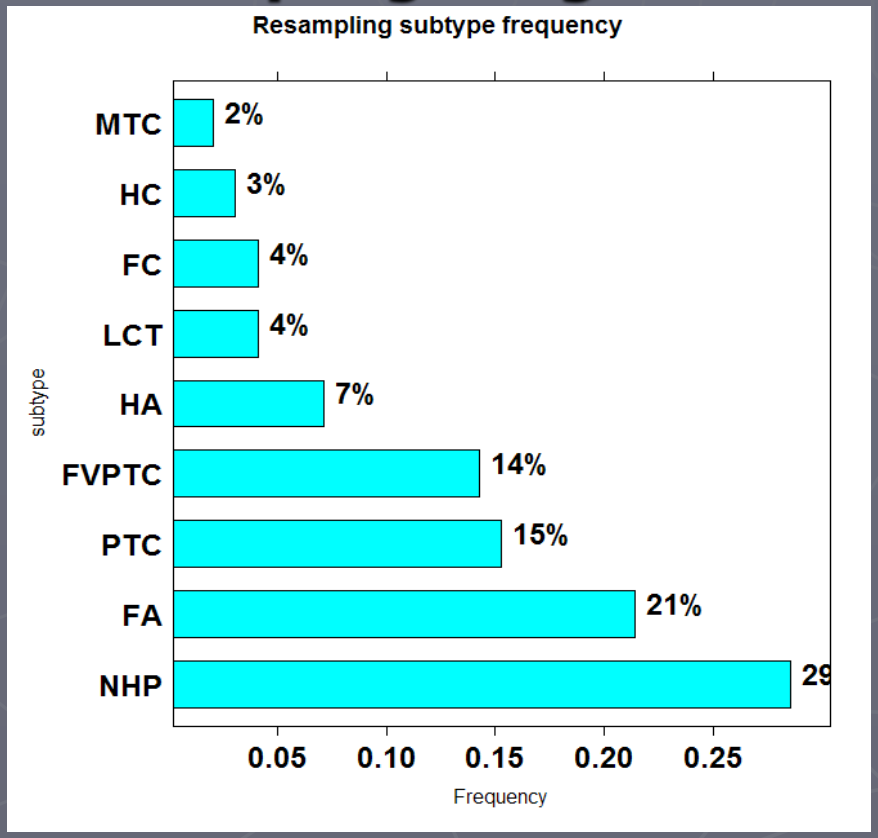
Many overturned by expert re-reads

Statistical Challenge: Metrics adjusted (or re-sampled) by subtype prevalence in cytology=I class

ROC curves (unadjusted and adjusted/re-sampled)

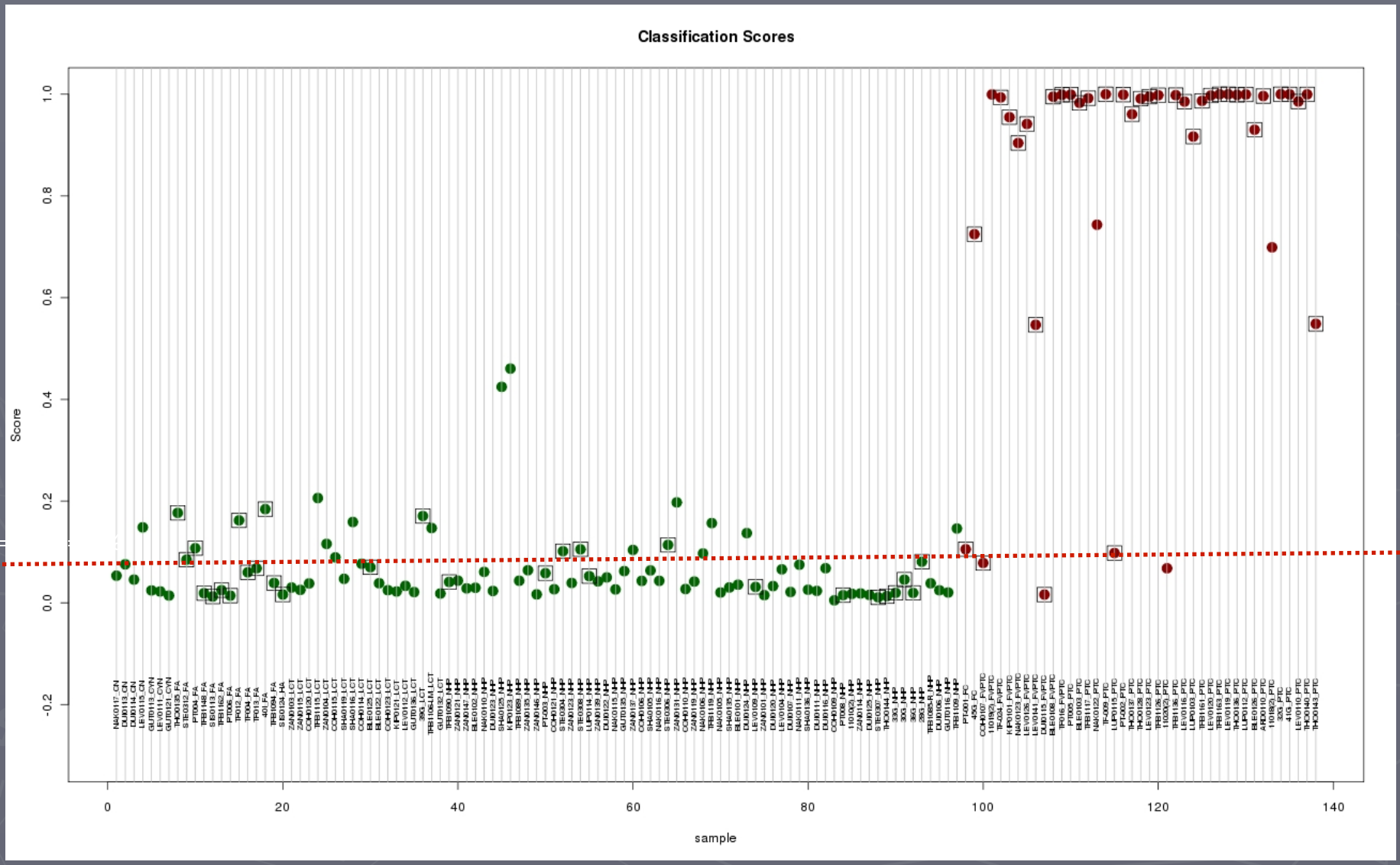


Re-sampling weights



Statistical Challenge: Choose Model cutoff

SVM linearfixedcost, NG = 200



Performance at various cut-offs

Cut-off = 0.075	True B	True M
Test B	0.8	0.1
Test M	0.2	0.9

Cut-off = 0.075	True B	True M
Test B	77	4
Test M	20	37

Cut-off = 0.5	True B	True M
Test B	97	5
Test M	0	36

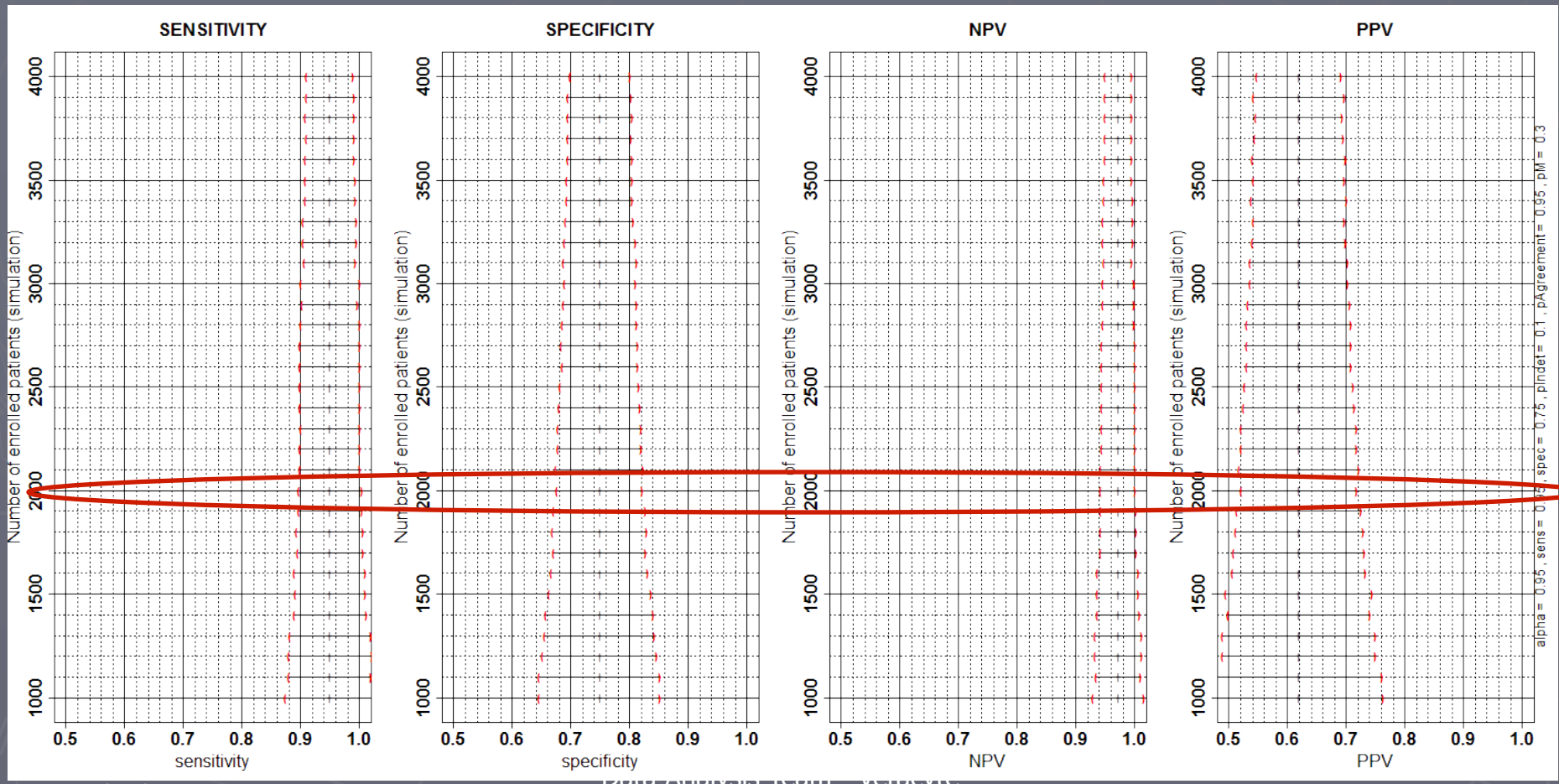
Cut-off	Sens	Spec
0.05	0.97	0.6
0.07	0.95	0.73
0.1	0.9	0.78

	CN	CYN	FA	FC	FVPTC	HA	LCT	NHP	PTC
B	4	3	12	1	2	1	18	59	2
M	0	0	0	1	8	0	0	0	27
%acc	100	100	100	50	80	100	100	100	94

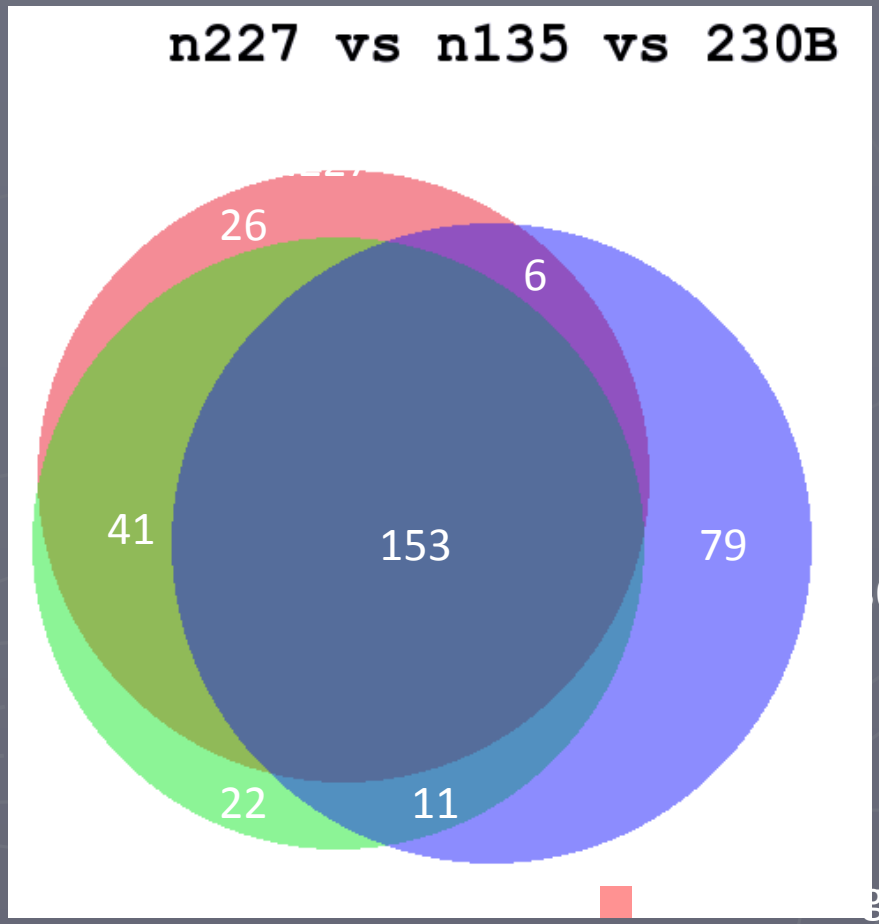
Statistical Challenge: Sample size calculation sens =0.95, spec = 0.75

N=200 I/S/A: recommended for FDA validation trial

N=2000→ N(I/S/A)=200 → enables us to exclude sens=.9, spec=.66 and NPV=.94!



Comparison of different marker lists show a stable marker list emerging!



0B

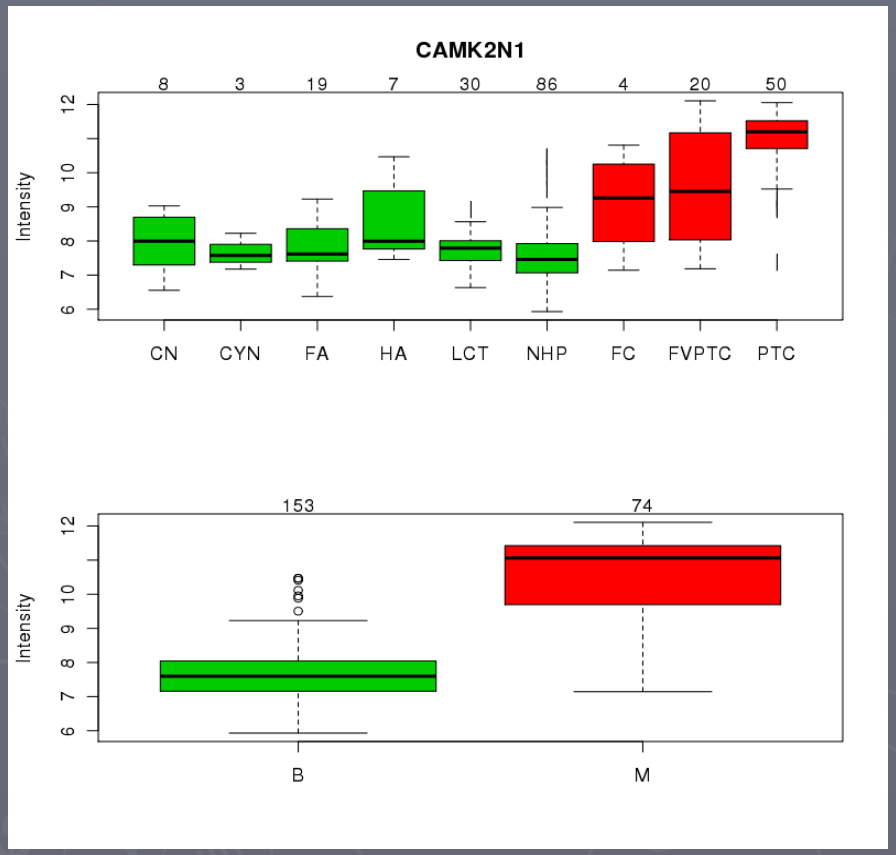
genes

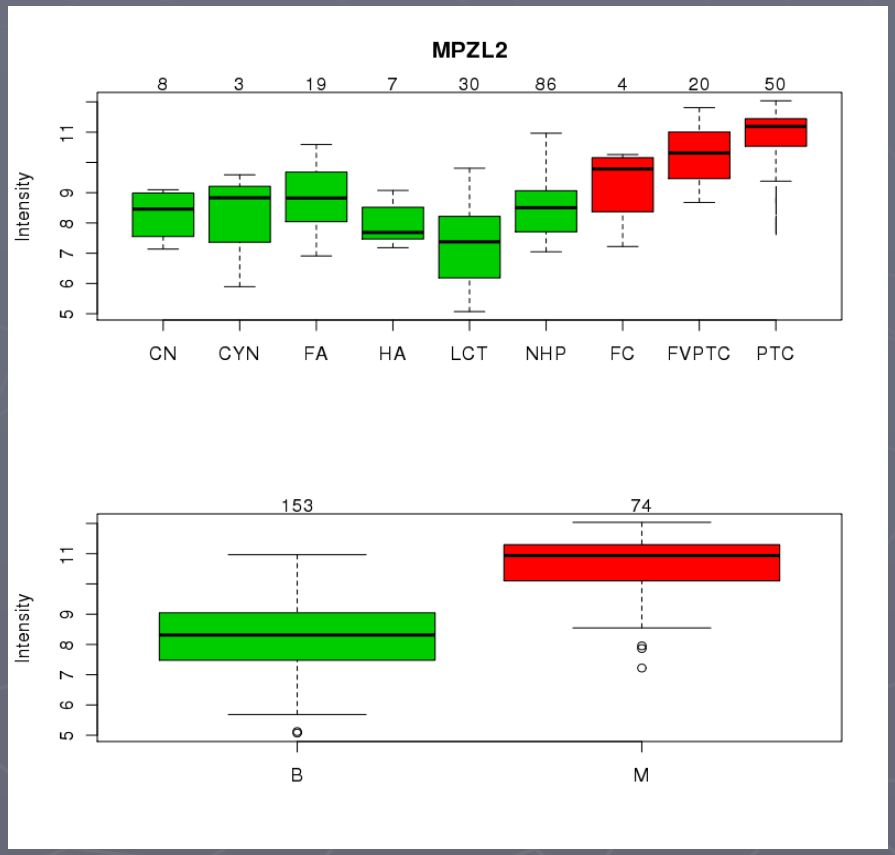
N135: 227 genes ← Our data set!

230B: 249 genes ← Older marker set which was derived from Tissue and ENA

Top 2 markers

Known Thyroid markers





Statistical Challenge – Design Custom Array!

Feature Size	Array Format	Minimum Array Order Size	# Probes	# Probe Sets*
	49-5241	35 - 48	221,832	55,458
	49-7875	35 - 48	506,952	126,738
	400	3700 - 3990	49,140	12,285
11 micron	169	1450 - 1680	223,732	55,933
	100	765 - 891	529,976	132,494
	64	55 - 63	961,400	240,350
	49	35 - 48	1,347,320	336,830
5 micron	100-2187	170 - 198	182,115	45,528
	100-3660	170 - 198	523,572	130,893
	49-5241	35 - 48	1,079,766	269,941
	49-7875	35 - 48	2,444,520	611,130
	400	3700 - 3990	247,186	61,796
	169	1450 - 1680	1,083,917	270,979
	100	765 - 891	2,564,275	641,068
	64	55 - 63	4,636,384	1,159,096
	49	35 - 48	6,480,226	1,620,056

* Each probe set contains 4 probes.

Probeset Tally

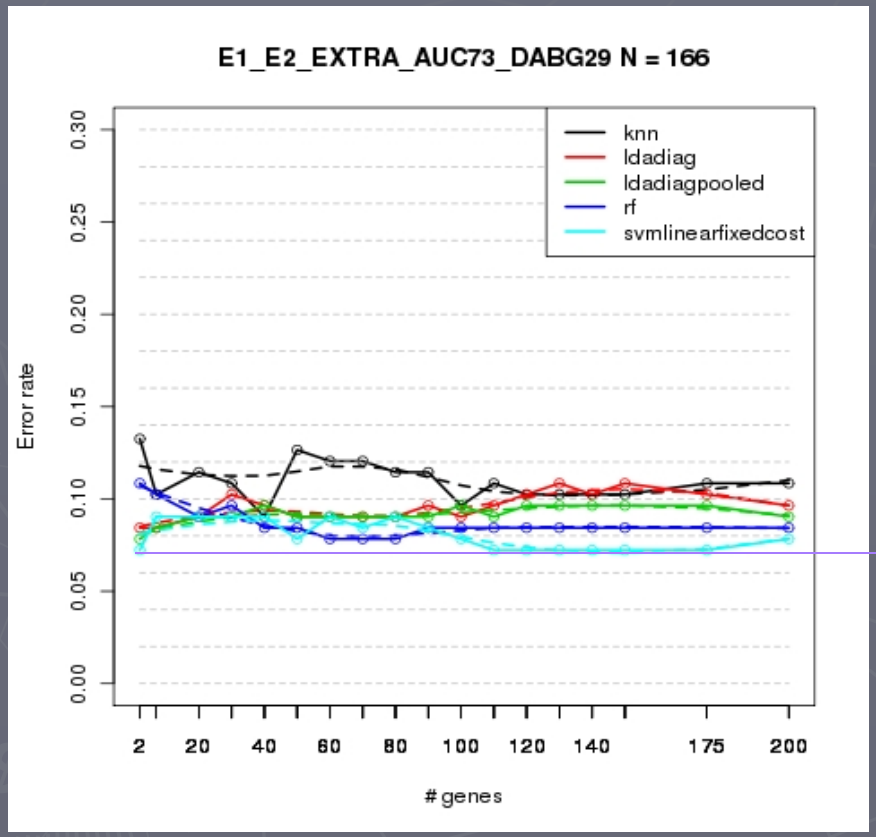
Category	Transcripts	Probe Sets	ProbesStatus
230B List	247	4323	16681DONE
232 List	232	3890	15006DONE
Top Tissue Lists *	922	13776	55580DONE
Top FNA Lists	1061	16716	64214DONE
Top FNA (multivariate)	325	5588	21575DONE
Top FNA RNAprotect Lists*	1740	25433	97651DONE
Top OB/OM Lists	477	6887	26496DONE
Biology Network Lists**	391	8169	31375DONE
Literature Markers 12-02-2009**	178	2513	9676DONE
combine.TISSUE.BAYES.REPEATABLE.marker.list.FNA.N227.csv	210	3484	13432DONE
Cell-Type Markers (Hemoglobin & LCTness & Lung)	18	142	442DONE
Alternative Splicing B vs M Probesets (ALLMEDIA+RNAp)	283	5615	21710DONE
Total Union of all of the above	**3207**	48326	185636DONE
Probesets responding to various experimental factors		442	1540DONE
Additional probesets proprietary to Veracyte		12383	48850DONE
FNA B vs M Probeset-level Top 500		545	2132DONE
FNA B vs M Probeset-level Top 500 - RNAProtect		559	2197DONE
Tissue B vs M Probeset-level Top 500 Rep>.1		642	2506DONE
Gender Determining Probesets		25	94DONE
Thyroid vs non-Thyroid Probesets		203	782DONE
Removing Non-Responsive Probesets (min.dabg >= 0.0001)		-6608	-21442DONE
Total Union of all Gene Level + Probeset Level Markers		52137	205031DONE
QC Probesets	4130	4130	15919DONE
Anti-genomic & controls		45	16943DONE
GRAND TOTAL			**238903**

Statistical Challenge: Designing custom array and in-silico assessment of classifier performance

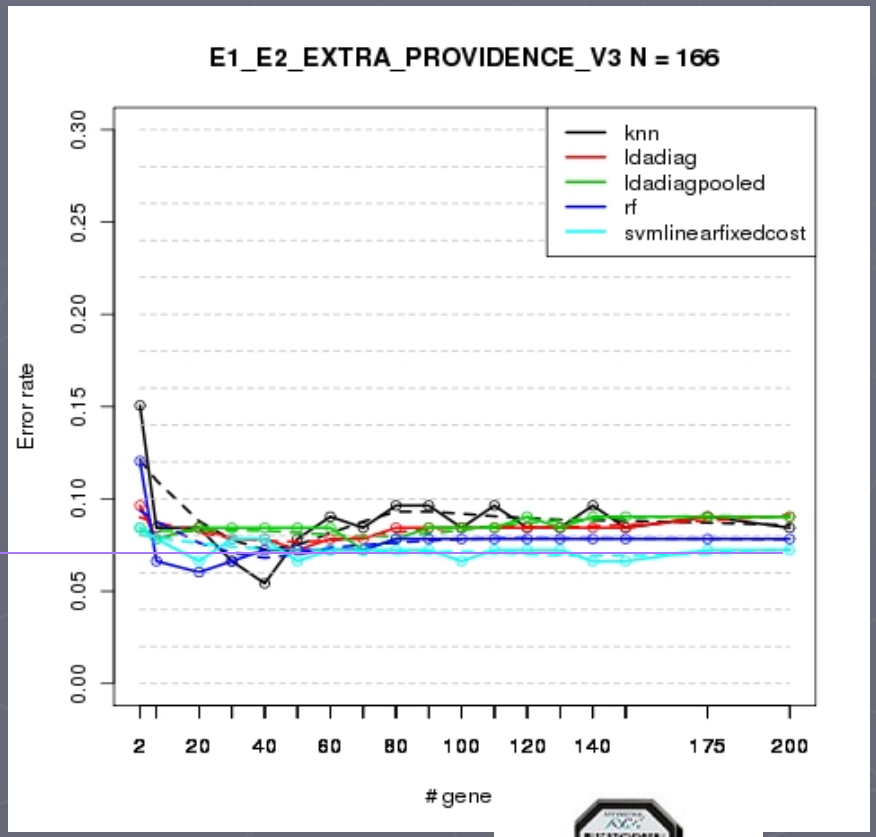
EXON Array

In-Silico Custom Array

5.4M Probes



~246K Probes



Summary of Statistical Challenges and Solutions performed in R and PERL

- Generate marker list from heterogeneous sources of samples
- Remove non-biological effects from data
- Accurately assess classifier performance across subtypes
- Software: Choose model cutoff, marker list and model coefficients and implement solution in PERL
- FDA buyoff on clinical trial design
  - Eventually Affirma is marketed as a CLIA test
- Design custom array

The entire diagnostic development process was in open software platform

- R
- Bioconductor packages
- Advanced model for teasing out non-biological effect on array through academically developed and contributed R package
- PERL chosen as software platform for final model
- Spotfire and Partek used for some modelling and visualization
- JMP and JMP genomics considered but not used

Acknowledgement

- Scientific staff at Veracyte, Inc.
- Shariq Alavi at Cytel, Inc.

Thank you!!