Discovery & development of biomarker candidates for drug development

V. Devanarayan, Ph.D.

AbbVie Inc., USA

Biopharmaceutical Applied Statistics Symposium
Orlando, November 4, 2013

abbvie

Disclosure statement

This presentation was sponsored by AbbVie. AbbVie contributed to the writing, reviewing, and approving the publication.

Viswanath Devanarayan is an employee of AbbVie.

– *V. Devanarayan, November 4, 2013*

Outline

Biomarker Overview

Finding needles in a haystack

- *biomarker discovery*

Creating optimal combinations

- *biomarker signatures*

Metrics / Scoring

- *biomarker performance*

Prognostic & Predictive Signatures

- *Patient Subgroup selection*

Endpoints & Biomarkers
Definitions

Primary Endpoint

- Characteristic that reflects how a subject feels, functions or survives

Surrogate Endpoint

- Biomarker intended to substitute for a primary endpoint

Biomarker

- **Characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic response to a therapeutic intervention**

Typical uses of biomarkers in clinical trials

Predict responders & non-responders to a drug

- Patient subgroup selection (Most Oncology programs, DAC, MTX, …)

Predict safety/AEs (e.g., liver, skin, kidney)

- *(e.g., Daclizumab, + compounds tested in discovery using Toxico-Gx)*

Patient-selection for clinical trial.

- Better specificity in disease diagnosis (e.g., AD vs. FTD vs. VD)
- Identify which patients are likely to progress in disease
- Reduce variability / placebo response (e.g., AD clinical trials)

Dose selection (PK-PD modeling)

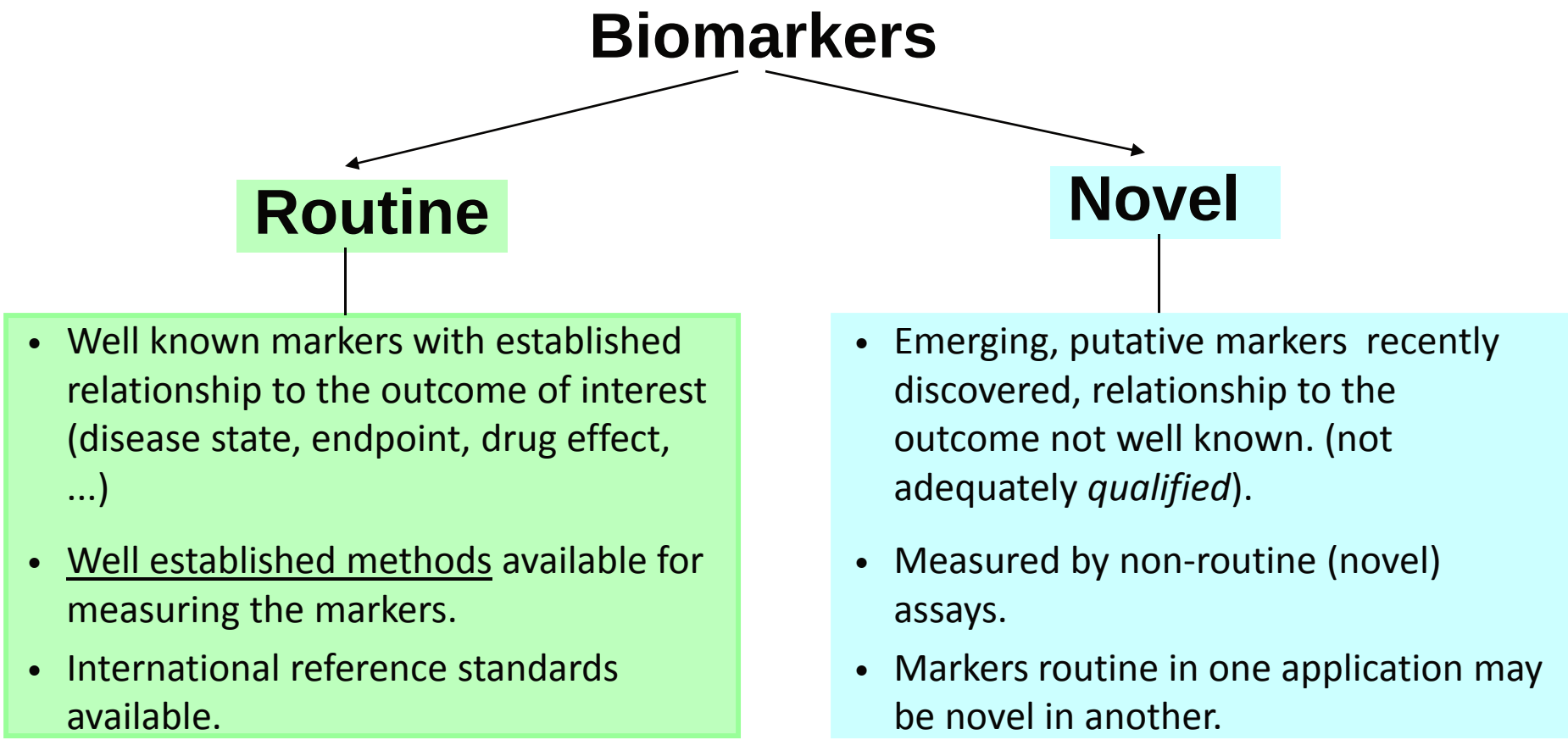
Proof of Mechanism & Concept in early drug development

Examples of Biomarkers in Current Use

Biomarker	Current Use	Classification	Qualification
HER2, EGFR, K-RAS mutations	Directing treatment in oncology	Predictive Biomarker	Defines indication in label, diagnostic development required
P450 enzymes (CYP2D6, CYP2C9, CYP2C19 polymorphisms)	Known to affect drug metabolism (e.g., for NSAIDs)	Predictive Biomarker	Can appear in label as risk factor. Prior testing suggested, dose adjustment
UGT1A1, TMPT, HLA-B*5701 polymorphisms	Predisposition to certain toxicities (e.g., liver, bone marrow)	Predictive Biomarker	Can appear in label as risk factor. Prior testing suggested, dose adjustment
AB1-42	Diagnosis of prodromal Alzheimer's Disease	Prognostic marker	Used to enrich clinical trial populations. Example of qualification procedure
Gene signature chips (e.g., Oncotype, Mammaprint)	Prognosis prediction in oncology	Prognostic marker (also predictive in certain cases)	Diagnostic qualification process applies
CRP, IL-6, TNFa in blood samples	Proof of principle in inflammatory diseases	Pharmacodynamic biomarker	Formal qualification not required, but fit for purpose assay validation
FDG-PET (SUVmax) Functional imaging	Proof of concept (e.g., in tumour metabolism)	Pharmacodynamic biomarker	Formal qualification not required, but collaborative opportunities
LDL cholesterol	Confirmatory trials in coronary heart disease	Surrogate Endpoint	Appears in label, used for approval. Any such new markers require qualification
HbA1c	Represents glycemic control in diabetics	Surrogate Endpoint	Appears in label, used for approval. Any such new markers require qualification

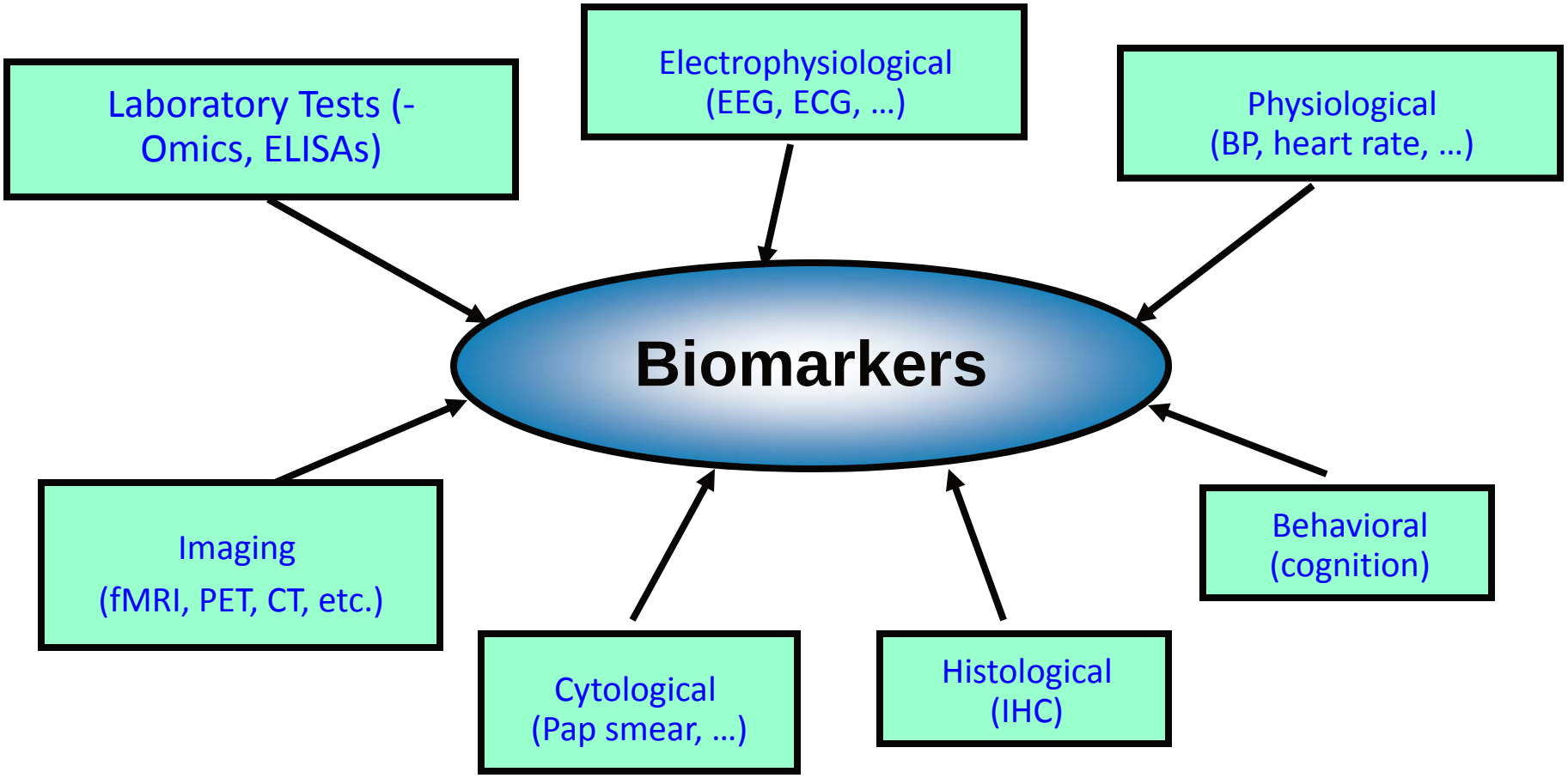
Jenkins et al., 2012

Routine versus Novel Biomarkers



Many biomarkers during development are Novel

Where do these Biomarkers come from?



Biomarker may be a Gene, Protein, Imaging, Clinical measure, etc.

Biomarker Discovery, Qualification & Method Validation Definitions

Biomarker Discovery

- **Process of identifying candidate markers for response of interest (target, disease, drug, clinical endpoints, …): hypothesis generation!**
- **Typically from Genomics/Proteomics/other-omics & targeted panels**

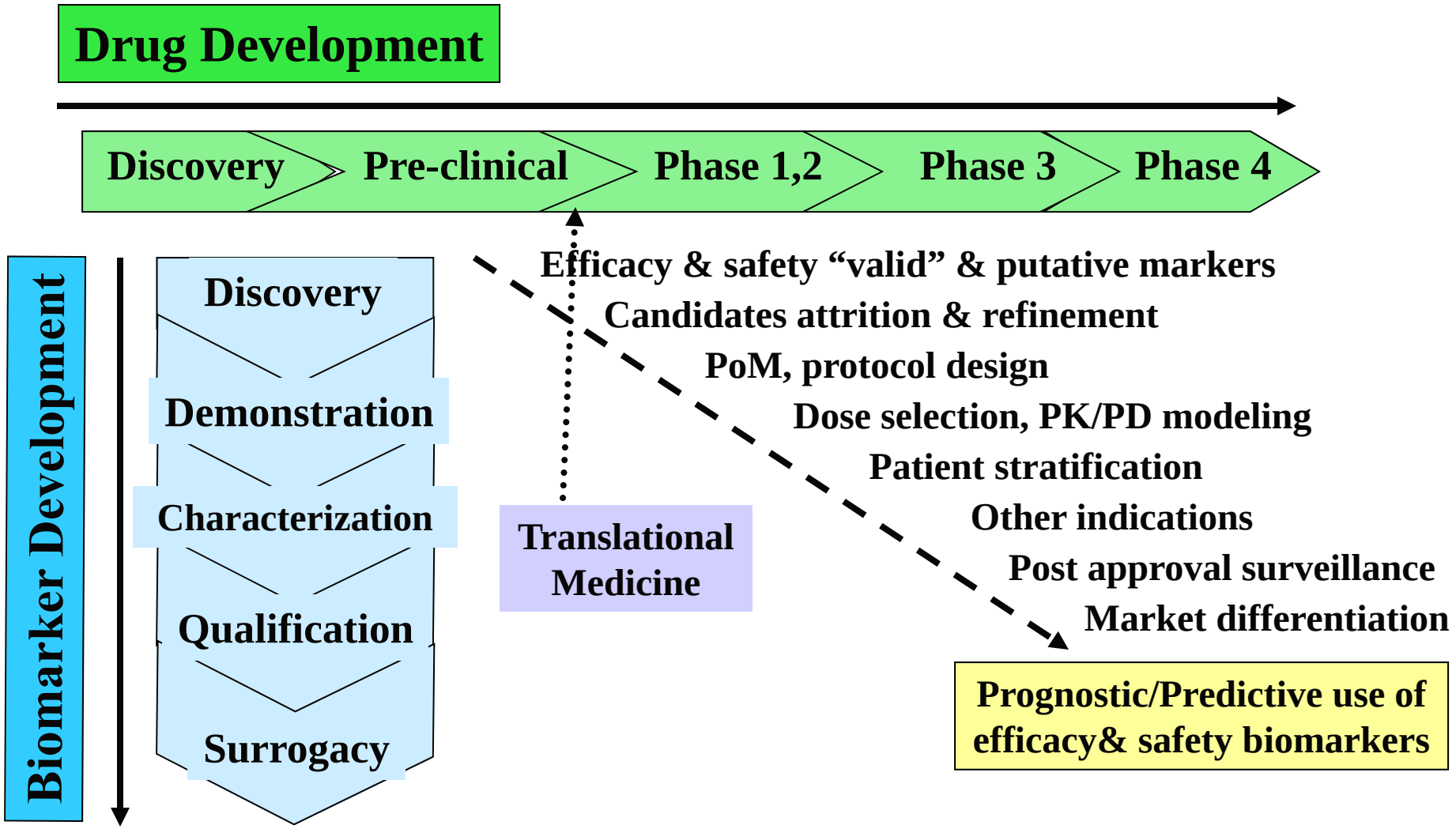
Biomarker Qualification

- **Evidentiary and statistical process linking biomarkers to biologic characteristics, pathologic state and/or clinical endpoints.**
  - **Degree of qualification achieved impacts the *fit for purpose*.**
- **More sensitive and targeted assays/platforms are used for quantitation.**

Biomarker Method Validation

- **Process of assessing the performance characteristics of a BM platform. (imaging platform, assay/multi-plex platforms, etc.)**
- **Includes all of the procedures required to demonstrate that a particular method is fit for its intended purpose.**
  - ***Industry white paper by Lee et al. (2006, Pharm. Research)***

Biomarker & Drug development are intertwined



Recommended Process for Biomarker Qualification

1. At least *two independent studies*, providing "predictive significance", PPV, NPV, etc., as appropriate
2. Performance should meet or exceed current state of art.
3. Studies should be well designed, powered and conducted by qualified investigators, and published!
4. Should describe control subjects & related disease groups. Demographics should mimic target population.
5. After a marker is "accepted", follow-up data should be collected to routinely monitor accuracy and value.

Evidentiary Standards for Biomarkers & Diagnostics

PhRMA/FDA paper, Altar et al., Clin Pharm & Therap. 2008

nature publishing group

PUBLIC POLICY

A Prototypical Process for Creating Evidentiary Standards for Biomarkers and Diagnostics

CA Altar, D Amakye, D Bounos, J Bloom, G Clack, R Dean, V Devanarayan, D Fu, S Furlong, L Hinman, C Girman, C Lathia, L Lesko, S Madani, J Mayne, J Meyer, D Raunig, P Sager, SA Williams, P Wong and K Zerba

A framework for developing evidentiary standards for qualification of biomarkers is a key need identified in the Food and Drug Administration's Critical Path Initiative.[1] This article describes a systematic framework that was developed by PhRMA committees and tested at a workshop

its use. The key driver that is used for the weight of evidence is a classic tolerability of risk argument: the value of a true result with a biomarker compared with the harm from a false result defined by the relevant stakeholders. If the value of the true result is high, and the harm of falsehood is low, then the weight of evidence

Outline

Biomarker Overview

Finding needles in a haystack

- ***biomarker discovery***

Creating optimal combinations

- *biomarker signatures*

Metrics / Scoring

- *biomarker performance*

Prognostic & Predictive Signatures

- *Patient Subgroup selection*

Biomarker Discovery

- Process of identifying candidate markers for response of interest (target, disease, drug, clinical endpoints, …).
- Typically from Genomics & Proteomics studies (Small n, Large p datasets)
  - 20-100 patients (n), but 500 to 3 million variables (p)
    - CGH Arrays: 150K, 500K, or 2-3 million variables
    - mRNA arrays: 20,000 – 50,000 variables
    - miRNA arrays: ~500 to 1000 variables
    - Proteomics: 500 – 5,000 peptides/proteincs

Goals:

1. Identify individual markers of strong ***significance***
2. Identify highly ***predictive*** composite/signature markers of outcome

Outline

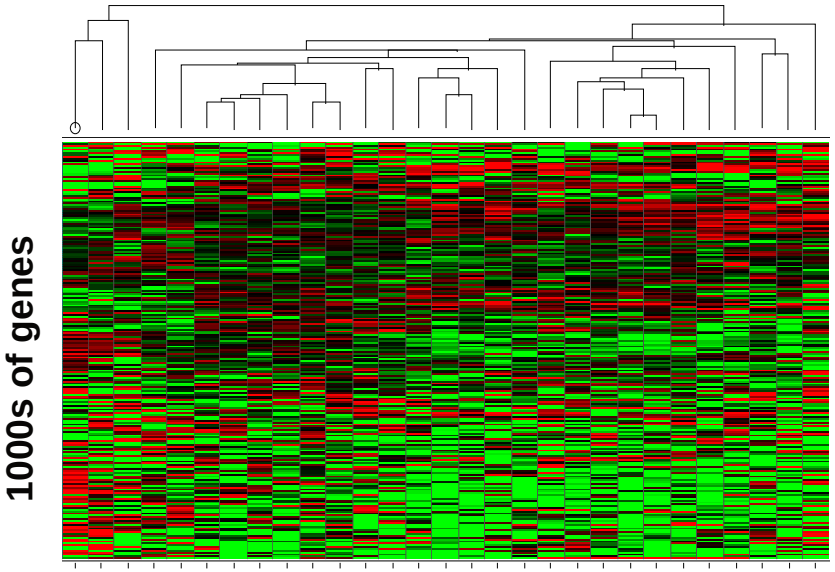
Single marker discovery & evaluation

- Data Normalization
- False Discovery Rate

Composite/Signature marker discovery & evaluation

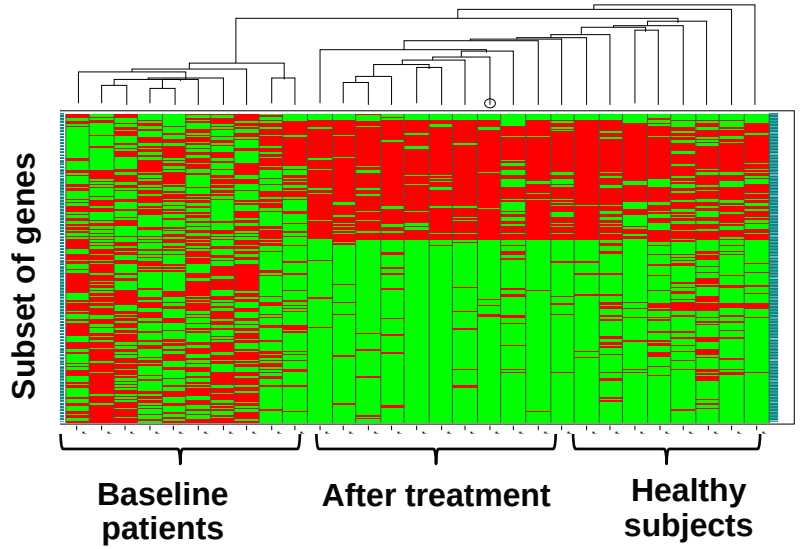
- Statistical Significance & Individual markers aren't enough!
- Filtering & Signature derivation
- Multivariate predictive modeling algorithms
- Inference via Cross-Validation

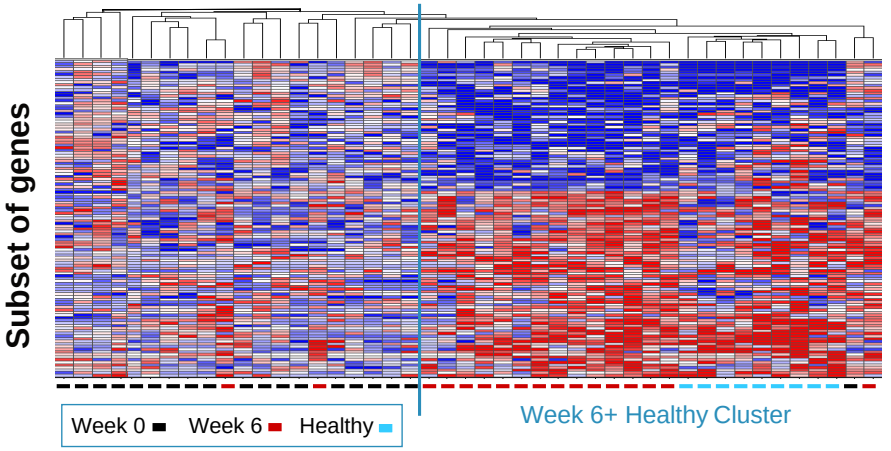
Example: Genomics data



Where are the needles?

What are some of the key statistical considerations?

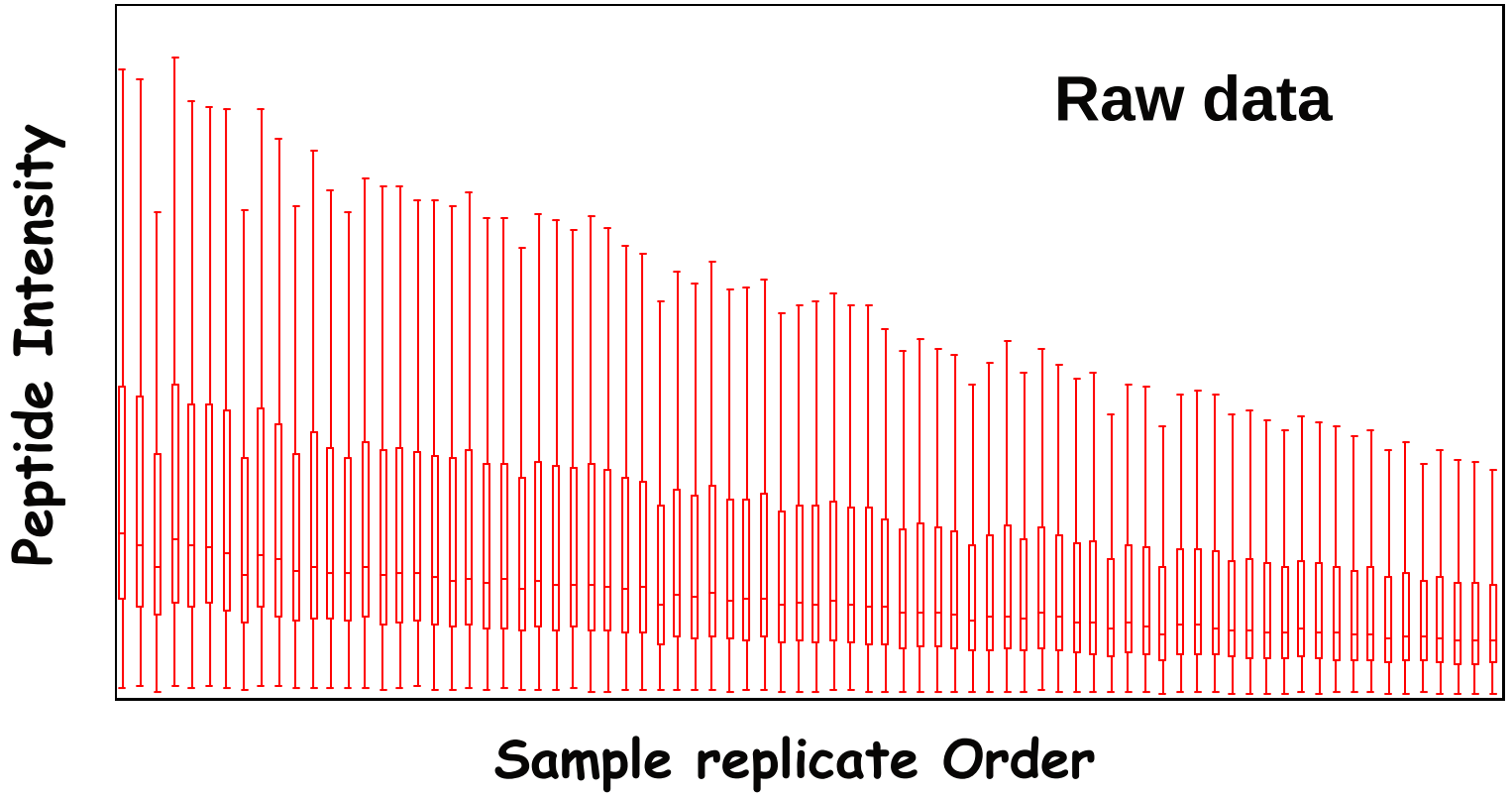




Raw data from a proteomics study

>11,000 proteins

80 replicates of the same sample



Kerry Bemis

Normalized data

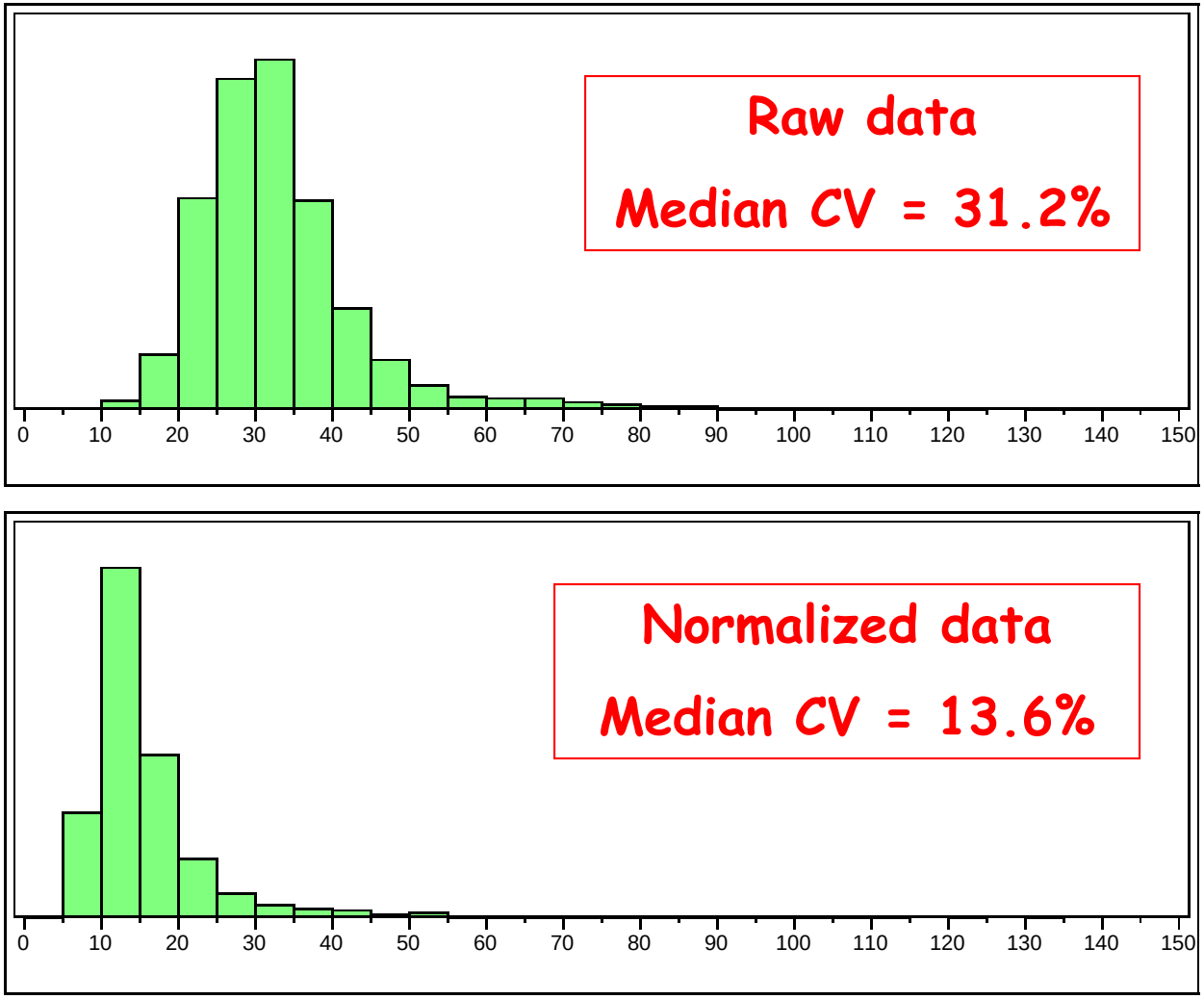
(Quantile method, Bolstad, et.al., Bioinformatics, 2003:185-193)

Log_2(Protein Intensity) = Average (Log_2 Normalized Peptide Intensities)

>11,000 proteins, 80 replicates of a sample

Kerry Bemis

Normalization Reduces Peptide Variability (%CV)



Kerry Bemis

Illustration:
FDR (q-value), FPR (p-value), etc.

Suppose that half the genes/proteins are truly positive, i.e., the ***True Positive Rate = 50%***

True

	# of Markers
Positive	5000
Negative	5000
Total	10000

Illustration:
FDR (q-value), FPR (p-value), etc. (contd.)

Called

True	Positive	Negative	# of Markers
Positive	4750	250	5000
Negative	250	4750	5000
Total	5000	5000	10000

False Neg. Rate
= 250/5000 = 5%

False Pos. Rate
= 250/5000 = 5%

p-value

q-value → FDR
= 250/5000
= 5%

Miss Rate
= 250/5000
= 5%

NB: Half the markers are truly positive in this illustration

Illustration:

What if the true positive rate = 5%?

Called

True	Positive	Negative	# of Markers
Positive	475	25	500
Negative	475	9025	9500
Total	950	9050	10000

False Neg. Rate = 25/500 = 5%

False Pos. Rate = 475/9500 = 5%

p-value

q-value → **FDR = 475/950 = 50%**

Miss Rate = 25/9050 = 0.3%

In most –omics applications, true positive rate ~ 5%

Is 50% confirmation acceptable ?

What threshold is appropriate for FDR?

Usually q < 0.05 or 0.1 is recommended. Markers that meet this criteria are considered to be ***robust*** with > 90 % confirmation rate!

For early discovery/exploratory studies, ***75% confirmation rate might be good enough.***

- **Identify all markers that have *q < 0.25* instead of q < 0.05.**
- **Identify all markers that meet *p < 0.05* criteria, but report the FDR as well so that you know the likelihood of confirmation.**

When is FDR (not) important?

Important for ***hypothesis generation*** from biomarker **discovery** (-omics) studies when the initial goal is to identify the top few markers of disease/endpoint/mechanism/etc.

- e.g., new disease targets for a drug discovery program

Not relevant for ***hypothesis confirmation*** or for a "targeted" evaluation of a few markers.

Not relevant when the goal is to identify **predictive** markers.

- **Predictive performance via cross-validation** is more relevant.
  - Hypothesis testing versus Predictive Inference; statistical significance doesn't imply good predictivity.
  - Some markers in a predictive composite can be non-significant!

Summary
Biomarker Discovery

- Data normalization impacts the ability to identify significant & predictive markers.
  - Analytical bias and variability are greatly reduced, thus increasing the ability to identify novel markers.
- False Discovery Rate estimates (q-values) are the primary focus in the *identification of individual markers* from large arrays.
  - p-values and q-values estimate entirely different parameters. Most importantly, q-value is not a "multiplicity-adjusted p-value".
  - p-values and fold-change aren't adequate, but should be considered in conjunction with q-values in biomarker discoveries.
  - Thresholds on q-values may be as high as 0.5, depending on the intended application.

Outline

Biomarker Overview

Finding needles in a haystack

- *biomarker discovery*

Creating optimal combinations

- ***biomarker signatures***

Metrics / Scoring

- *biomarker performance*

Prognostic & Predictive Signatures

- *Patient Subgroup selection*

Biomarker Signatures

Biomarker Signature is a combination of one or more markers, when applied to an empirical model or rule, *predicts* an outcome of interest.

- Outcome may be *disease process, drug efficacy, safety*, etc.
- Defined by list of markers & empirical model/rule

May be as complex as

- "25-genes in a Neural-Network model" for predicting patient response
- "15 proteins in a " for predicting tumor progression.

or as simple as

- Linear combination of 5 markers to predict cancer survival.
- Decision threshold on a marker to predict disease progression.

Biomarker Signatures (contd.)

Variety of possible data-sets (usually “small n, large p”)

- # of subjects: 10s to 100s (n)
- # of markers: 10s to 100s to 1000s to Millions (p)
  - Genomic arrays
    - CGH Arrays: 150K, 500K, or 2-3 million SNPs
    - mRNA arrays: 20,000 – 50,000 gene probe-sets
    - miRNA arrays: ~500 to 1000 genes
  - Proteomic Arrays: 10-100-1000s of peptides/proteins
  - Imaging: Voxel-level (several 1000s) or “region aggregates (10-100s).
  - Clinical observations: 10s to 100s of clinical measures

Often, multiple sources of biomarker data are available from the same set of subjects (e.g., clinical + genomic).

Biomarker Signatures (contd.)

Typically, signatures include markers from only one source

- gene signature, proteomic signature, imaging signature, etc.

Often appropriate, and possibly more powerful, to include markers from multiple sources.

- *"Extreme" example:* A 12-marker signature for predicting *patient prognosis, disease progression*, etc., may include
  - 3 proteins measured using ELISAs
  - 2 genes; mRNA expression from TaqMan
  - 3 imaging variables (e.g., k-trans for Cancer, vMRI measures for AD)
  - 2 genetic variables (e.g., ApoE for AD)
  - 2 clinical variables (baseline Age, Cognition, etc.)

Usually more practical to expect a combination of 2 to 3 sources (e.g., imaging+proteomic, genetic+clinical, etc.)

Predictive inference, Not statistical inference!

Interest is in how well a biomarker signature predicts an outcome of interest, usually at the individual patient-level.

- *Typically, "individual" refers to a*
  - *patient/subject (drug response, disease process, etc.),*
  - *compound (prediction of compounds likely to toxic), etc.*

Need minimal overlap of the individuals between the groups that are compared.

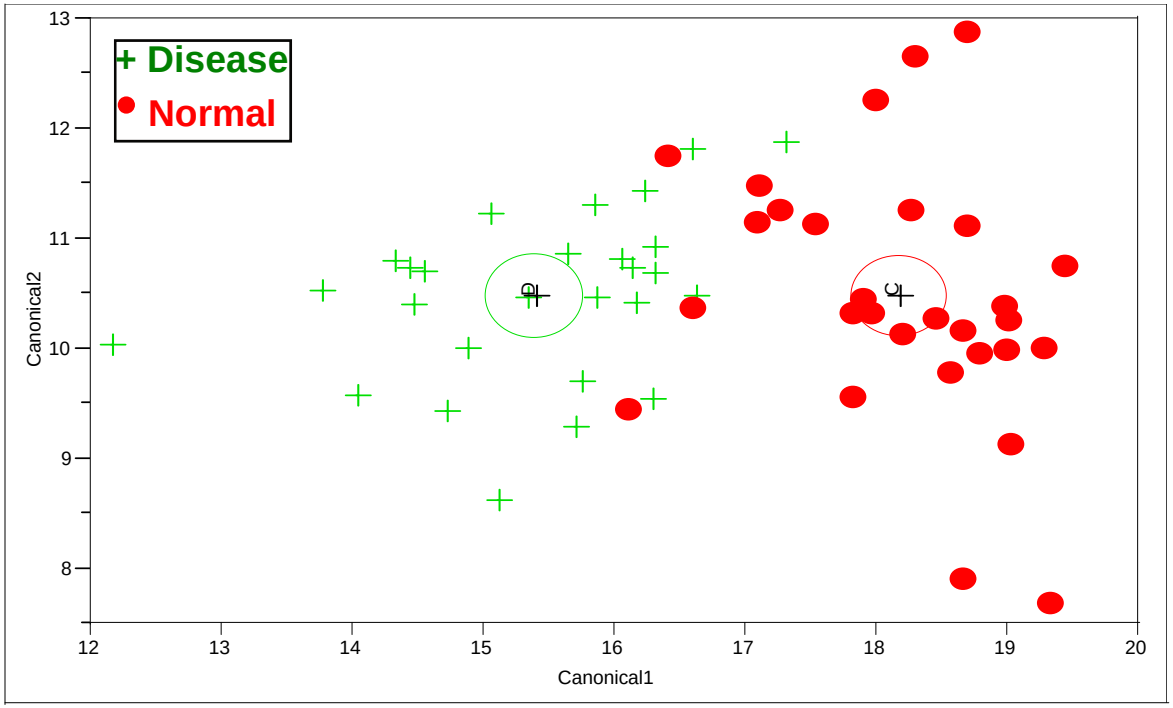
- Not adequate if the groups are different with respect to just their mean biomarker response ($p<0.05$ doesn't mean much!).

Predictive inference – *rigorous assessment of the predictive performance* ***via cross-validation*** */ resampling methods.*

- *p-values don't matter much.*

Individual vs. Multi-Analyte Performance

Statistical significance isn't enough!



Discriminant Analysis

- **This multivariate analysis output shows the ability of a composite of 6 markers to discriminate Disease from Normal.**
- **Predictive Accuracy of this "composite biomarker" = 94%**
- **But each marker, although p <0.01, is < 70% sensitive/specific!**

A marker useless on it's own, may be useful in a composite!

Case-Study:

A marker, "mk.1", is highly significant on it's own, and provides **75%** predictive accuracy.

Another marker, "mk.X", although **biologically relevant from pathway analysis**, is statistically useless on its own ($p>0.05$).

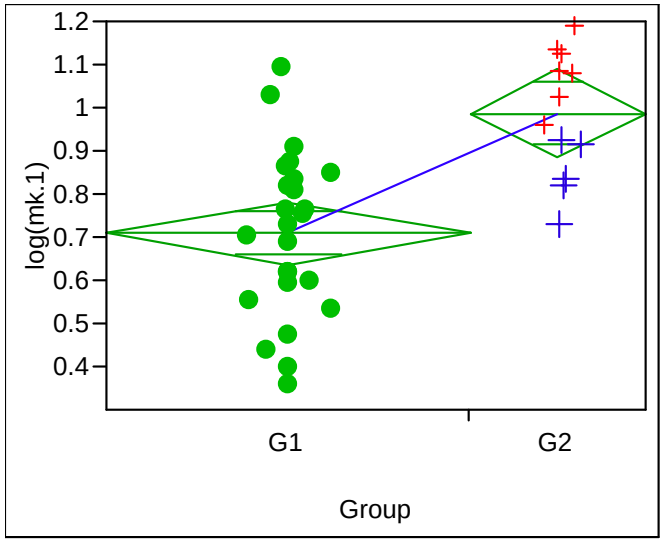
But when mk.X is combined with mk.1, predictive accuracy = **89%**.

Surprised?

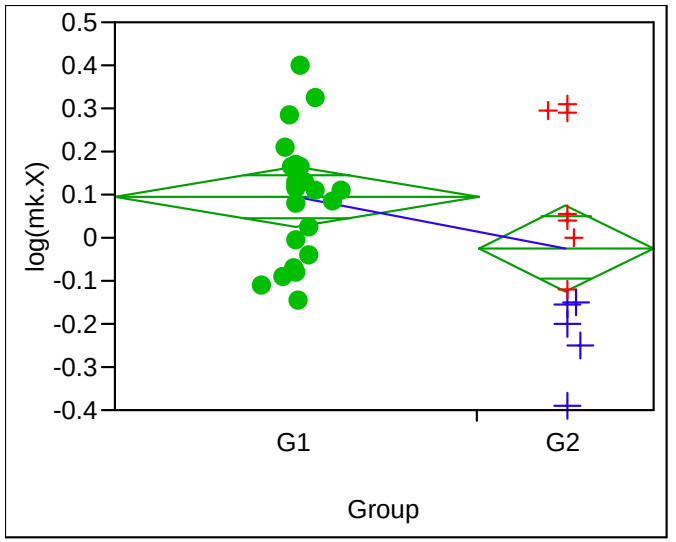
Let's look at a scatter-plot.

A marker useless on it's own, may be useful in a composite! (contd.)

mk.1



mk.X

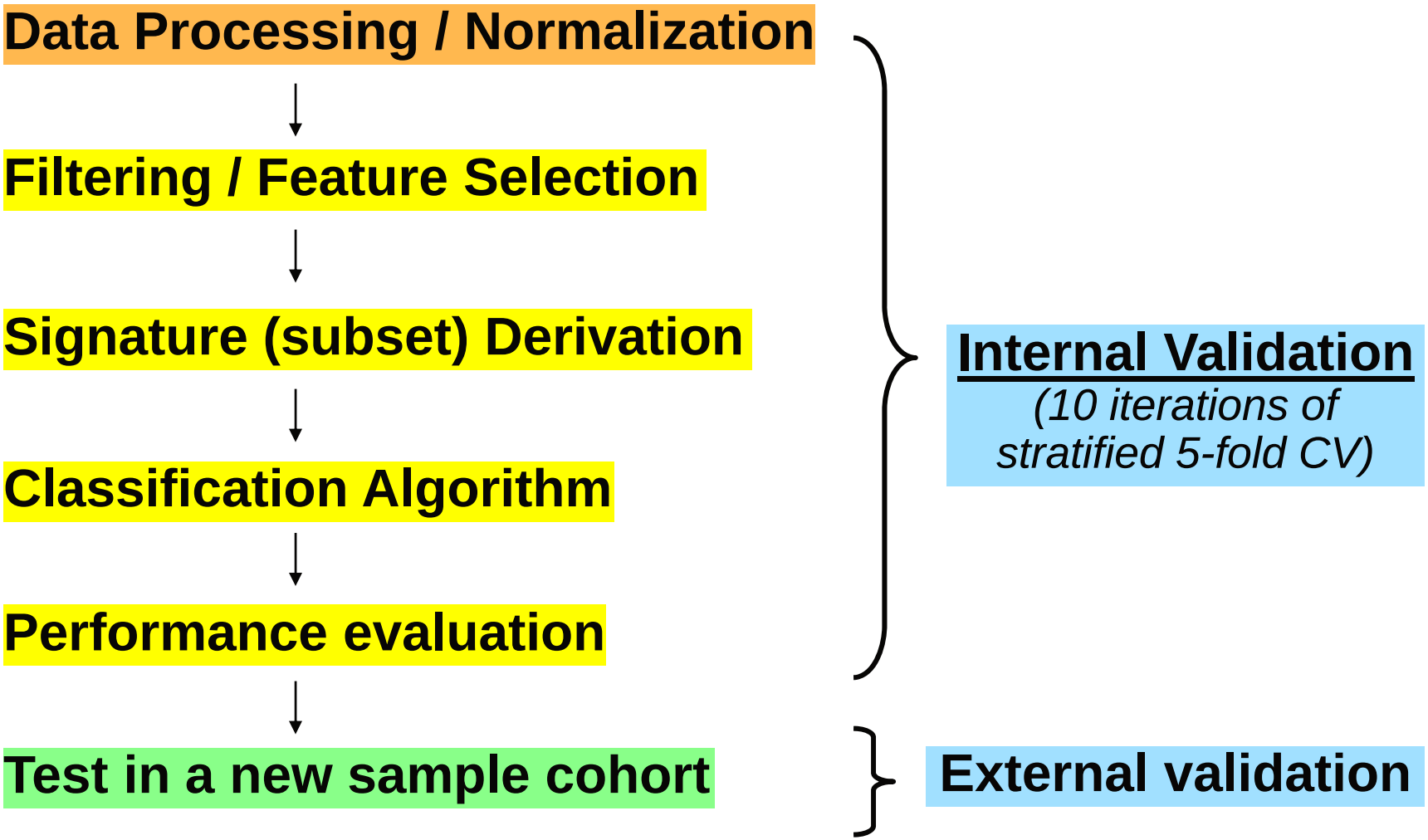


Patients that overlap with respect to mk.1 do not overlap in mk.X

Thus when combined into a signature, the overall prediction accuracy increases significantly. This has to be confirmed in subsequent studies.

mk.X is biologically relevant as well, hence further strengthening it's consideration in the model.

Process of deriving prognostic/predictive biomarker signatures



Filtering / Feature Selection

Score statistics and q-values from SAM may be used for selecting the Top-X genes.

- Significant Analysis of Microarrays; Tusher et al., PNAS, 2001.

Depending on the dataset, and the types of classification algorithm/model, different numbers of markers are selected.

Usually a small number (e.g., 25-500) is adequate!

Biological screening of markers based on literature or pathway analyses should also be considered.

Signature (final subset) Derivation

Goal: Need signatures with best predictive power.

Several options available, some examples:

- Forward selection, AIC, etc.
- Relative importance scores from various models (RF, LDA, PLS, etc).
- Simulated Annealing with nearest centroids.
- Lasso, Elastic-Net, etc.

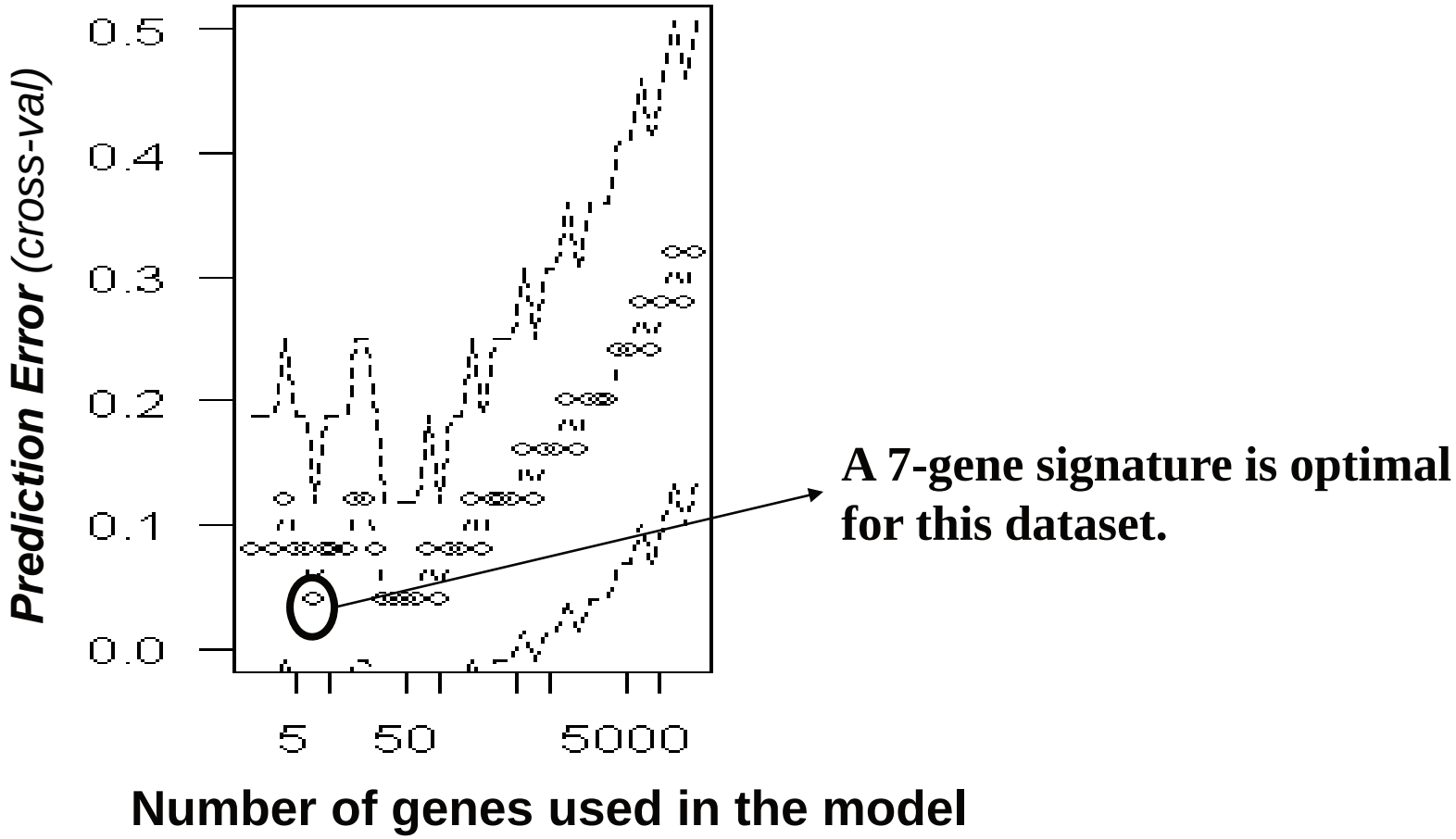
The optimal signatures/subsets are then applied to various classification algorithms.

Signature Derivation (contd.)

- Markers in the optimal signatures may not be the ones that are most highly significant on their own.
  - For example, the optimal 6-gene signature may not be 6 most significant genes; genes that are ranked much further down the list but are less correlated may yield greater predictive power.

Signature Derivation (contd.)

Graphical representation of composite selection using importance scores from RF. Reference: Diaz-Uriarte (2006, BMC Bioinformatics)



Multivariate models/ algorithms

- Forward stepwise linear/logistic/GAM models
- LDA, DLDA, DQDA, etc.
- Random Forests, AdaBoost, Bagging, SVM, Neural Nets, SAM, kNN, etc.

- Select an algorithm that provides the best predictive performance.
- Top 2 or 3 algorithms may provide different insights on the markers that may complement each other.
- Regularization-based methods (lasso, elastic-net) such as GLMNET and Shrunken Centroids are much faster and do both signature selection and model fitting simultaneously.

Outline

Biomarker Overview

Finding needles in a haystack

- *biomarker discovery*

Creating optimal combinations

- *biomarker signatures*

Metrics / Scoring

- ***biomarker performance***

Prognostic & Predictive Signatures

- *Patient Subgroup selection*

Cross-Validation, Model Reliability, etc.

Using the same data to identify and evaluate a biomarker signature will exaggerate its performance.

Leave one-out methods are not adequate either.

Typically need several replicates of "**k-fold cross validation**".

- Original data divided randomly into k equal parts
  - If N=1000, k=10, obtain 10 random subsets of 100 each.
- Leave out the first subset ("test data"), build the model on the remaining k-1 subsets ("training data") and use this model to predict the cases in the first/test subset.
- Repeat this procedure by leaving each of the other k subsets, one at a time.
- Determine the prediction error rates across these k permutations.
  - % sensitivity, % specificity, overall error rate, etc.
- Repeat this k-fold cross validation procedure 50 times.
- Aggregate the error rates across these 50 repetitions (*report Mean & SE*)

Cross-Validation, Model Reliability, etc.

Choice of k depends on N. Generally 5 to 10 is OK. If k is too small, say if k=2, it can be lead to highly fragile results & biased estimates of sensitivity, specificity, etc.

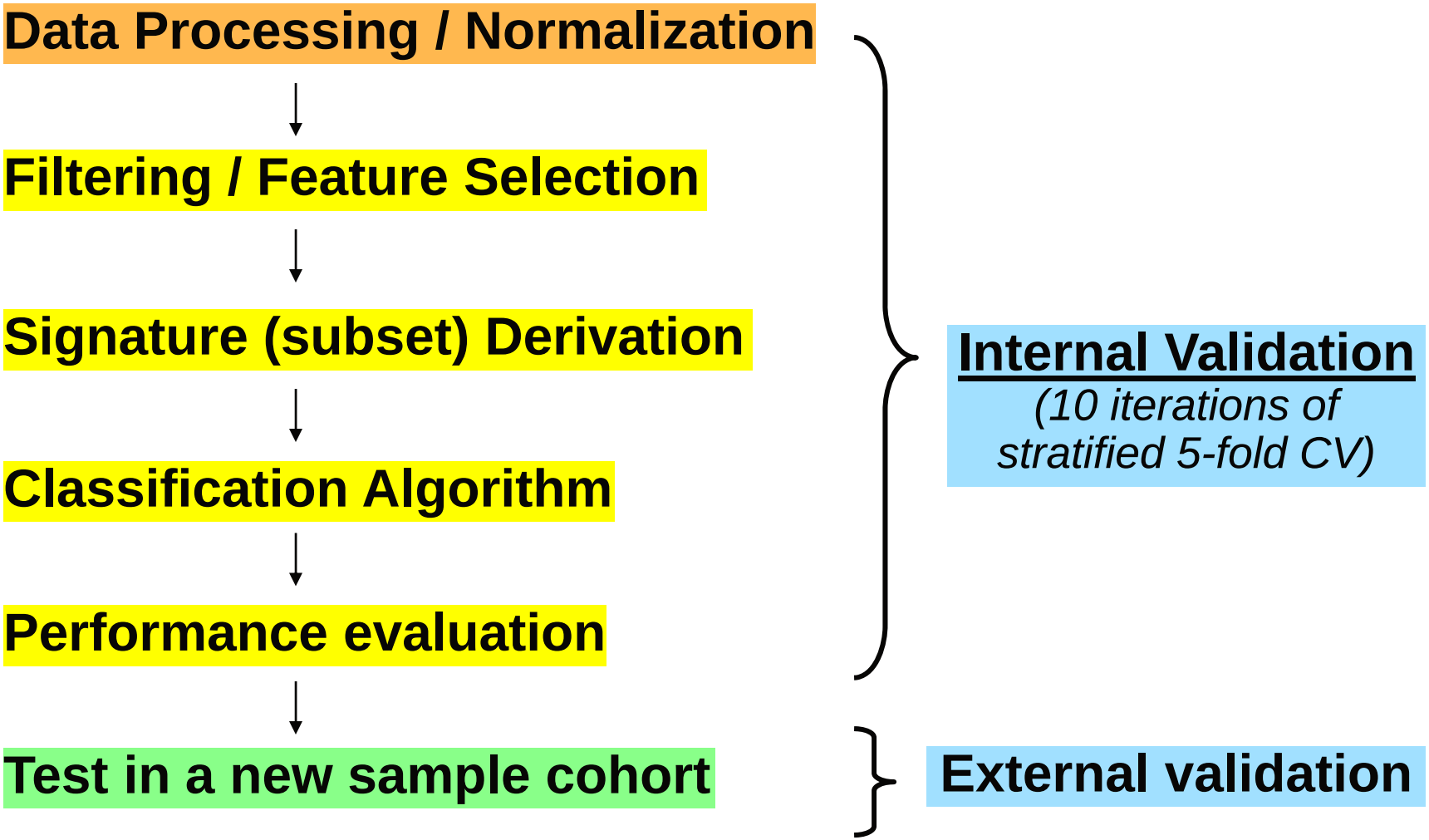
Example of Questionable results:

- Dave et al. *"Prediction of survival in follicular lymphoma based on molecular features of tumor infiltrating cells".* NEJM, Nov. 18, 2004 vol. 35set 2:2159-2169
- Reasons are explained and illustrated at:
  - http://www-stat.stanford.edu/~tibs/FL/report/index.html

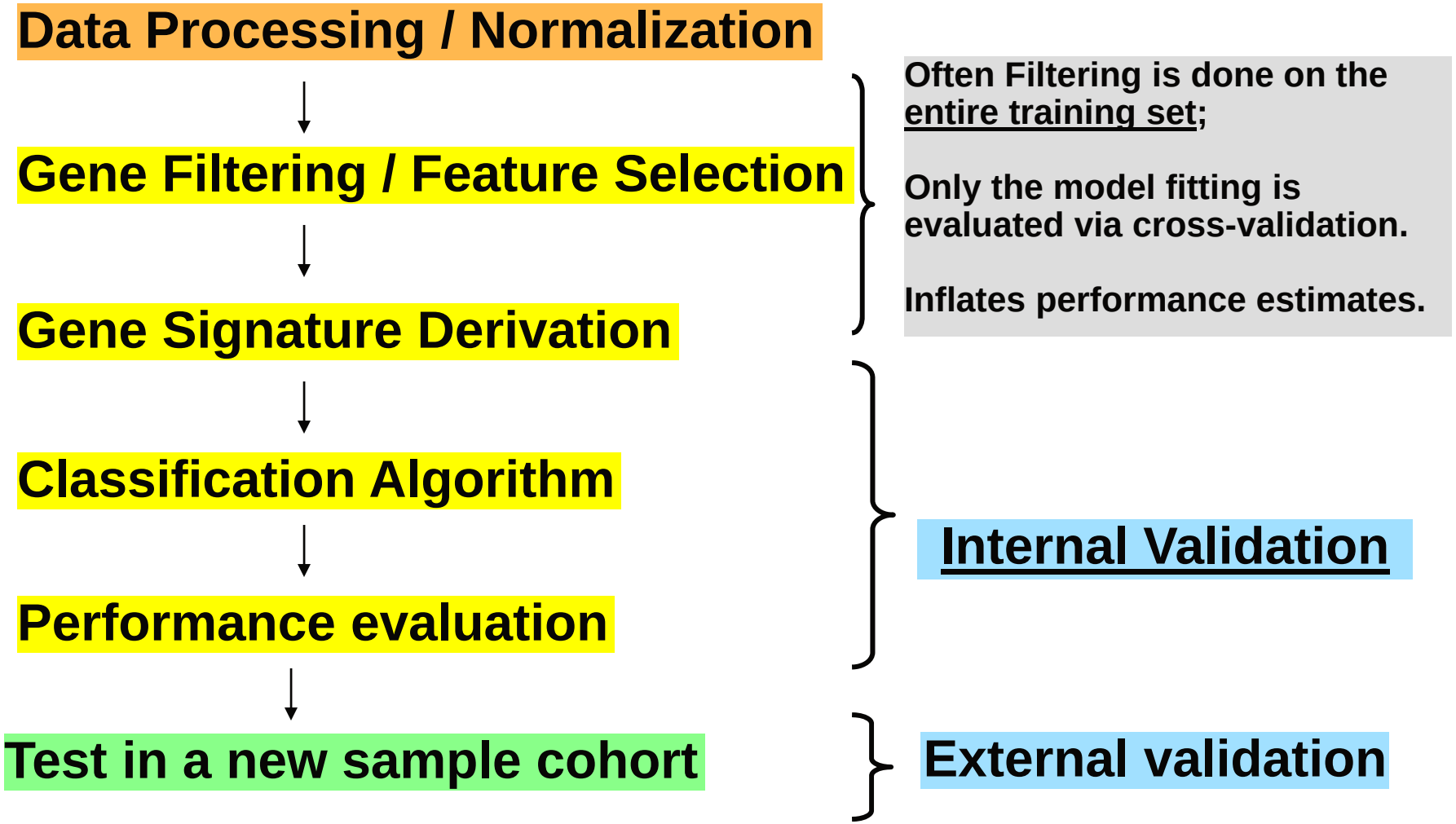
→ ***Unfortunately this is very common! Even Ph.D. statisticians not trained in biomarker data analysis make these mistakes.***

→ ***Don't take publication/literature claims for granted.***

Process of deriving predictive biomarker signatures (with full cross-validation)



Common mistake in performance evaluations (only partial cross-validation)



Biomarker Performance Evaluation
External Validation

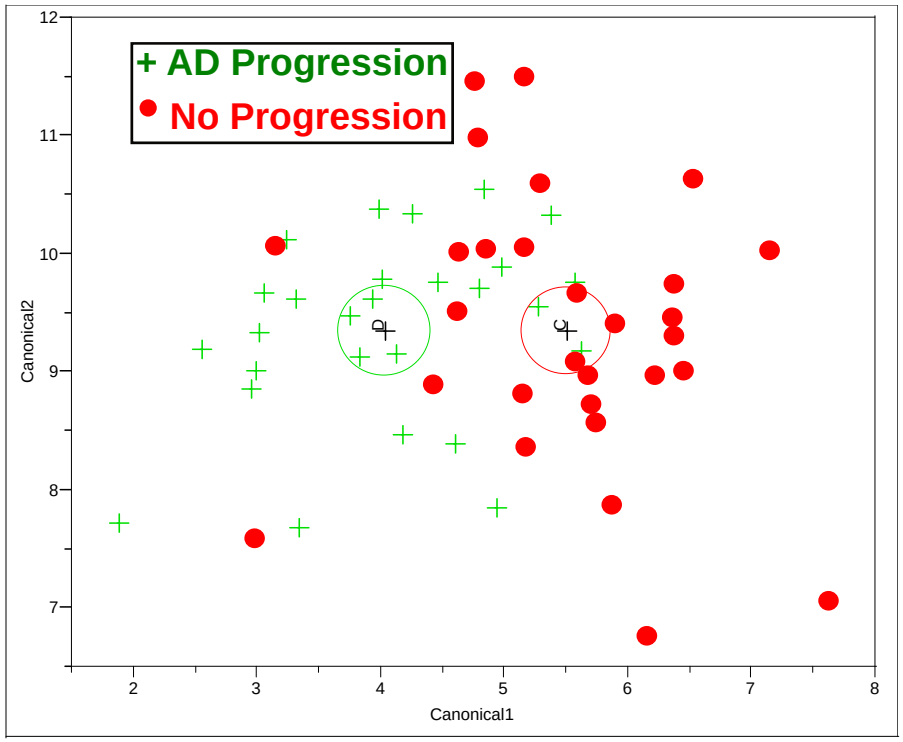
- After rigorous internal cross-validation, *test the signatures in independent external cohorts*.
- Should adequately represent the target population with respect to several features (gender, race, age, disease severity, …)

- Samples in training & external sets are seldom run together.

So *batch-effect normalization* may be necessary.

1. Normalize the training & external data.
   - A method that works well in my experience: *Eigen-Strat.*
2. Apply previously derived signature on the normalized training set.
3. Use this model on normalized external data to predict the response.

Example 1: Evaluation of Biomarker Performance



6-marker proteomic multiplex signature for possible use in selecting patients for a Clinical Trial

Predictive Accuracy:

- **Internal Cross-Validation:**
  - **No Cross-validation (CV): 84%**
  - **Partial CV: 72%**
  - **Full CV: 65%**
- ***External Validation (new study): 63%***

Improper Cross-Validation exaggerates biomarker performance.

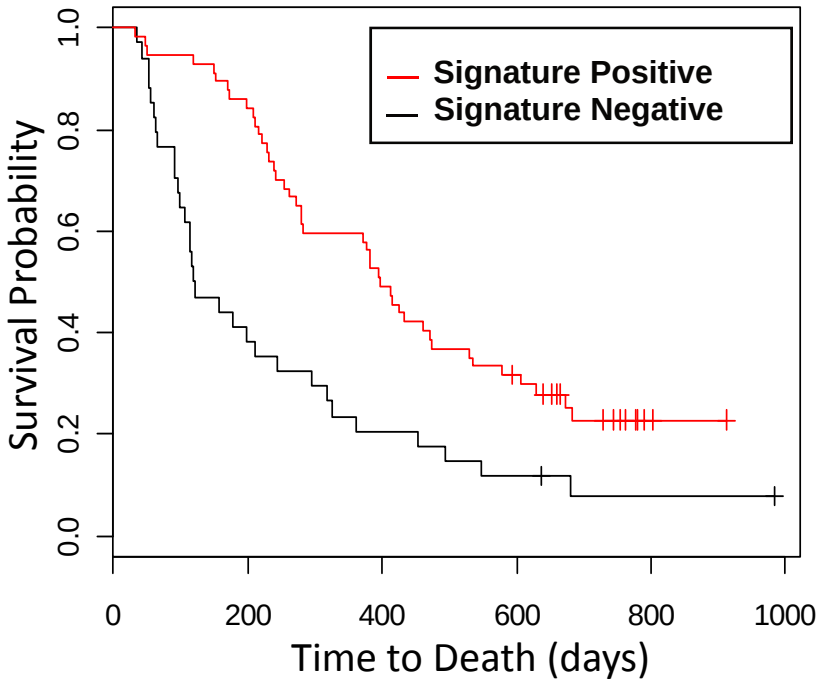
Example 2: Evaluation of Biomarker Performance

4-SNP Genotype Signature for Predicting Patient Response to a chemotherapy.

- **Derived from a large genotype array (100s of SNPs) via a Statistical Algorithm**

Signature Positive: ***SNP-1 ≠ WT,SNP-2 ≠ WT, SNP-3 = WT, SNP-4 ≠ WT***

- **Patients in this Signature Positive group are expected to respond better.**



p-value of Treatment Effect in Signature Positive vs. Negative:

- **Internal Validation:**
  - **No Cross-Validation: p < 0.0001**
  - **10-fold Cross-Val: p = 0.06**
- **External Validation: p =0.1**

Improper Cross-Validation exaggerates biomarker performance.

Hypothesis-driven vs. Hypothesis-free

Instead of using the whole array data of say 30K genes (hypothesis-free), a biologically targeted (hypothesis-driven) list of say 1000 genes can be used in the signature development process.

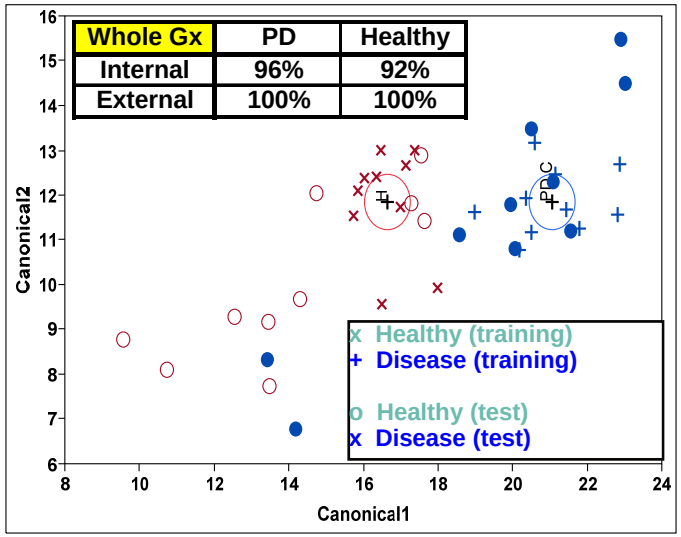
- Biologically relevant lists typically come from pathway analyses and literature.
- Often, not a considerable difference in the predictive performance between these signatures.

Clinicians & biologists may be more comfortable with signatures based on primarily biologically relevant markers.

Hypothesis-driven vs. Hypothesis-free Example

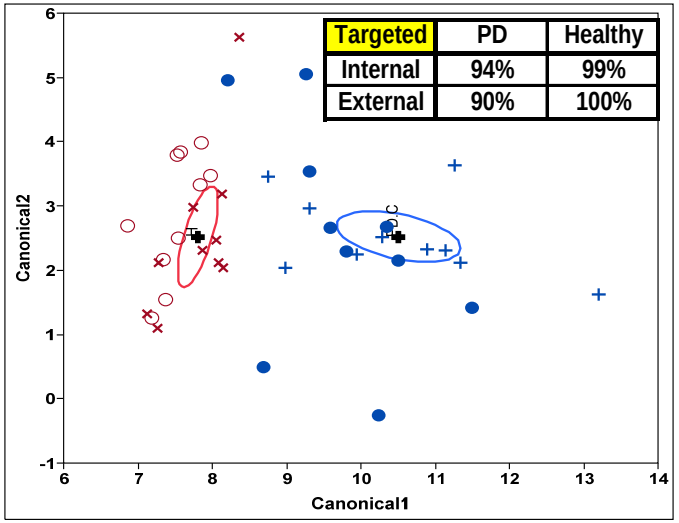
Optimal signature derived from the entire genomic array.

(hypothesis-free approach



Signature derived from only a subset of genes in the biological pathway

(hypothesis-driven approach)



Biological hypothesis-driven signature performs almost as well (in this example), and is more likely to be accepted than the hypothesis-free signature that includes unknown/novel genes.

Robustness

During a study, additional variability can be introduced (unavoidable factors)

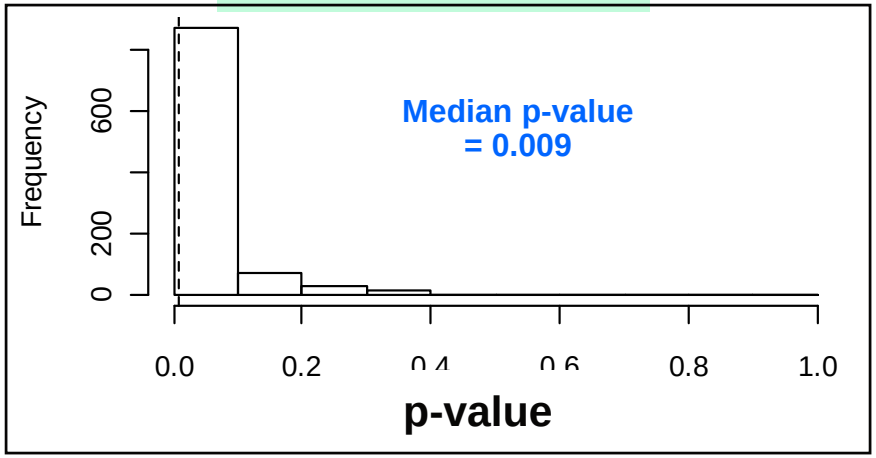
- **changes in reagents, instruments, operators, sample collection/storage, …**
- ***This is typically not accounted for during biomarker validation/evaluation.***

Example: **5-marker Signature for identifying patients more likely to respond to treatment. Robustness of this signature is evaluated via Simulations.**

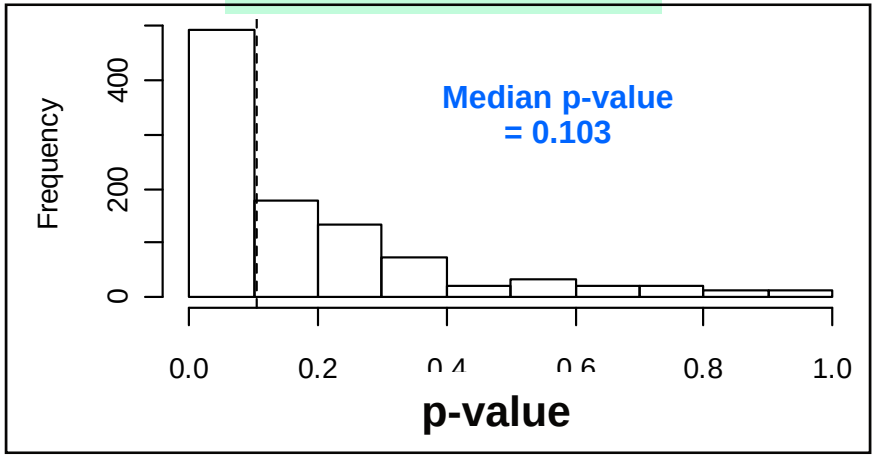
15% CV & 30% random noise are artificially added to the original data.

Distribution of p-values for Treatment Effect evaluated via 1000 iterations.

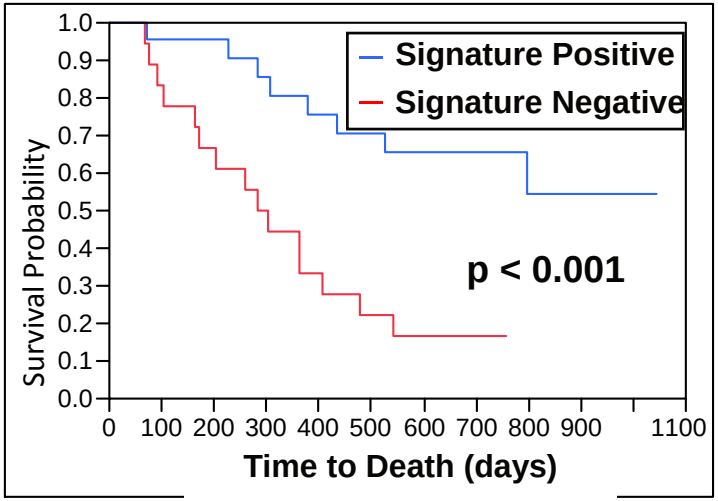
Additional 15% CV



Additional 30% CV

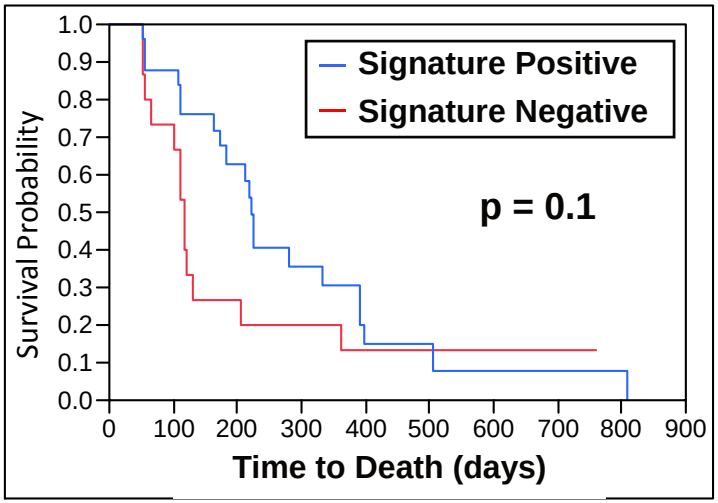


Translation



Biomarker Signature derived & evaluated in male cancer patients

Confirmed via external validation on same population

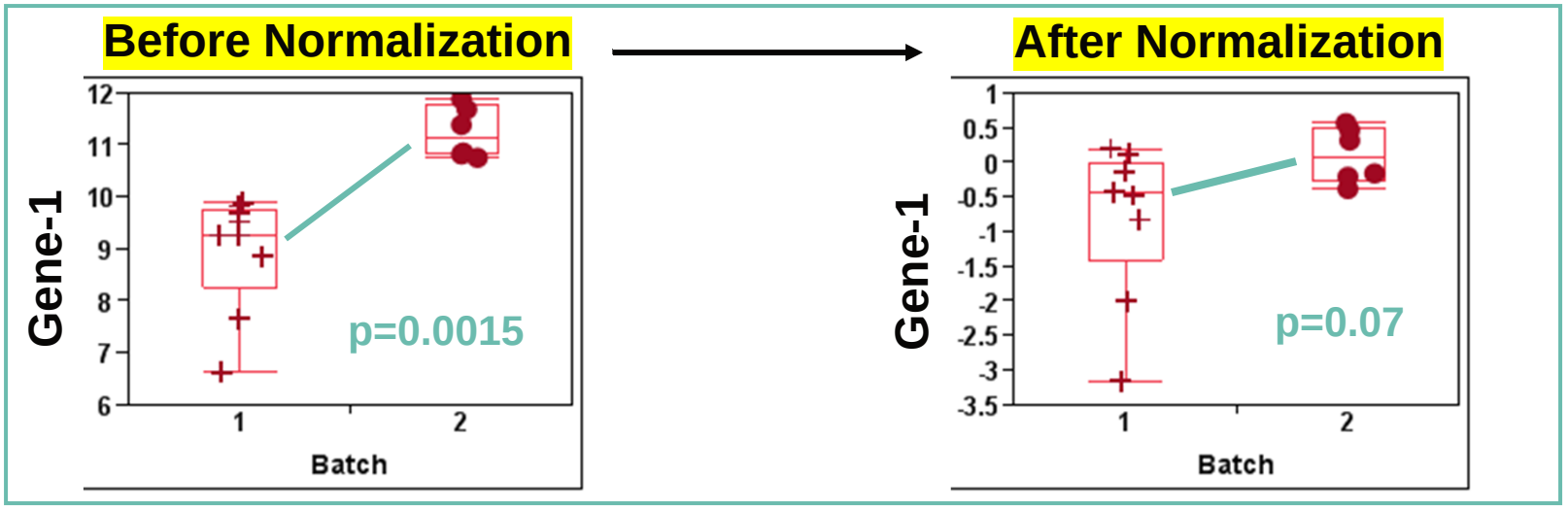


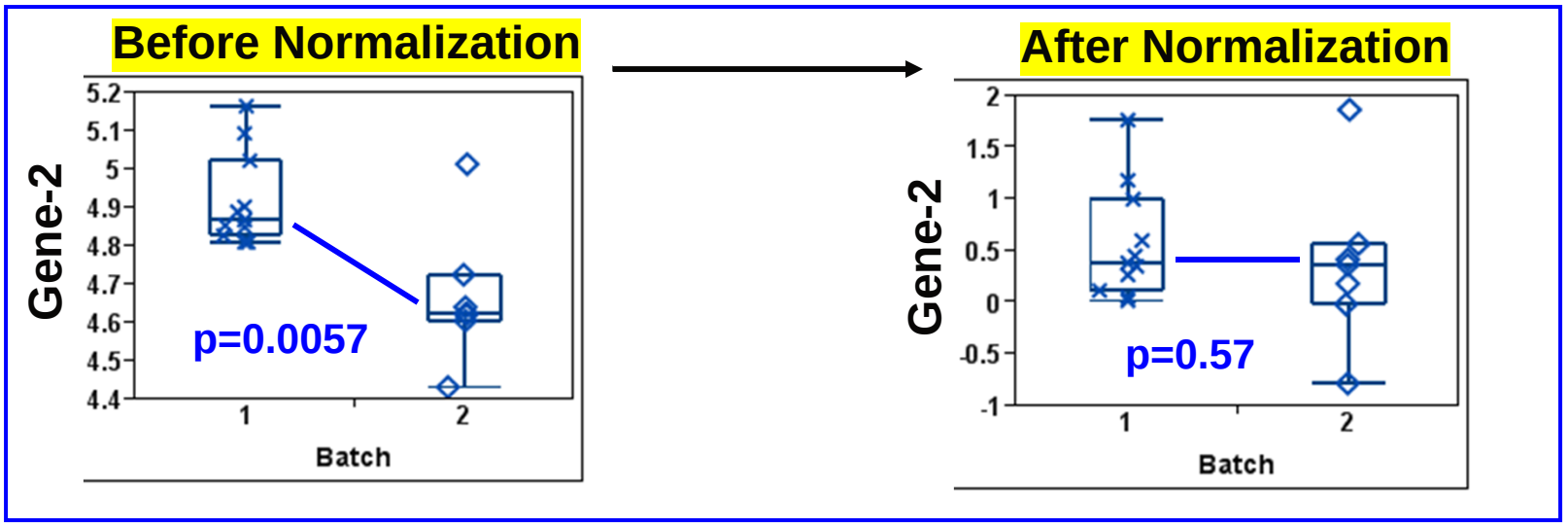
Same Biomarker Signature does not perform well when tested in a different study (females, older age group, more severe cancer)

This gets more challenging between animals & humans!

Batch-Effect in External Validation

Illustration from a genomics study

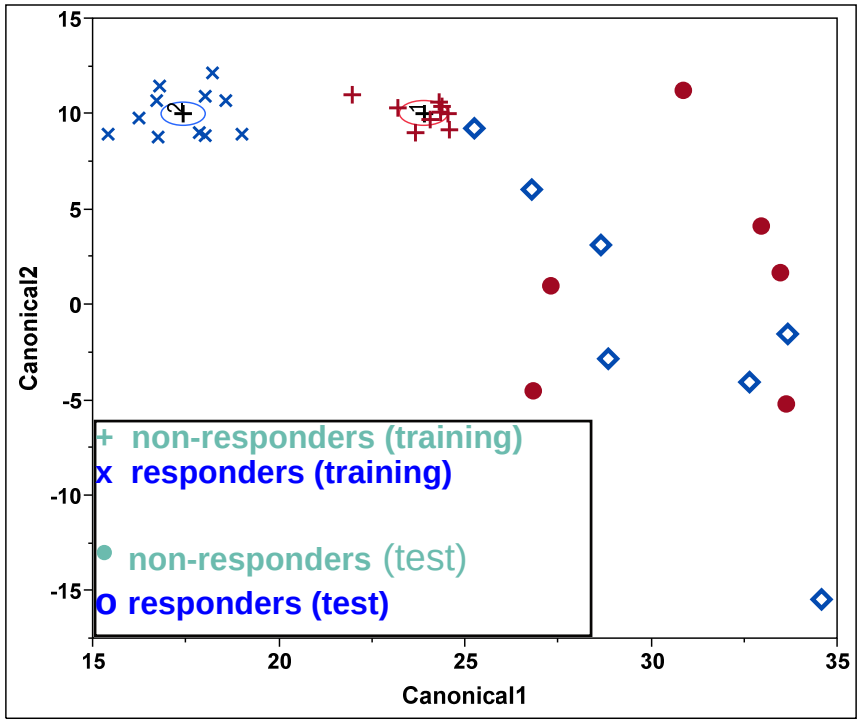




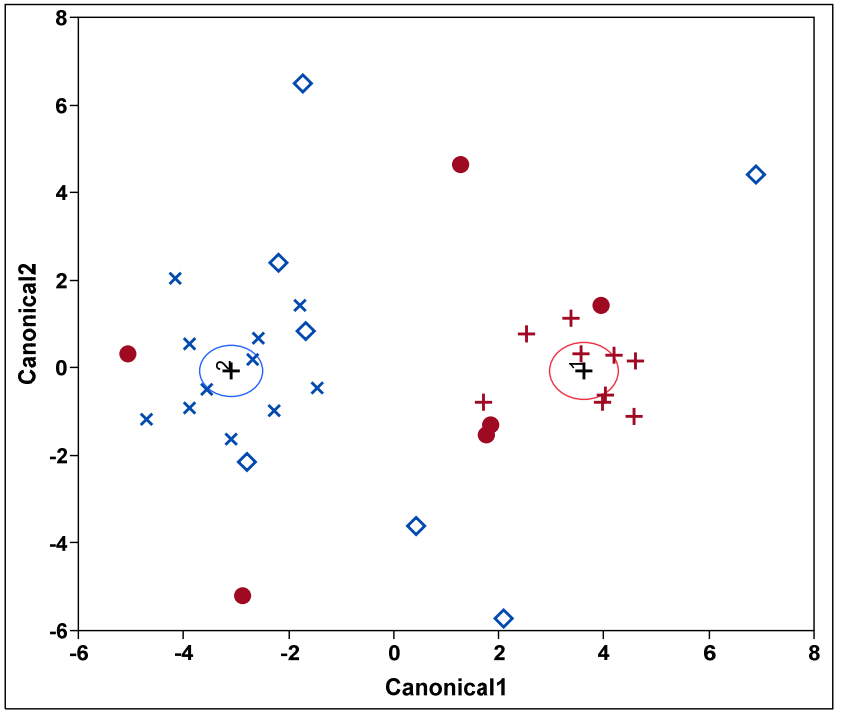
Impact of batch-effect on external validation

Illustration from a genomics study (contd.)

Before Normalization



After Normalization



Before normalization, all "responders" are incorrectly predicted.

Normalization yields significant improvement, although far from perfect due to other challenges (external set included a different disease state as well).

Outline

Biomarker Overview

Finding needles in a haystack

- *biomarker discovery*

Creating optimal combinations

- *biomarker signatures*

Metrics / Scoring

- *biomarker performance*

Prognostic & Predictive Signatures

- ***Patient Subgroup selection***

Prognostic vs. Predictive Signatures for patient subgroup selection

- **Predictive Signatures** - predict the response to a specific treatment compared to other treatments.
  - ➢ Identifies patients respond only to our drug, and not to the competitor.
- **Prognostic Signatures** - predict the disease outcome irrespective of the treatment.
  - ➢ Identifies patients that respond to our drug, but is not specific to our drug (i.e., these patients may respond to competitor drugs as well).

- Apply similar data analysis process as described in earlier slides.
- When the relationship is ~ step-wise or nonlinear, development of thresholds (cut-points) may be more attractive to clinicians.
  - ➢ Data-driven statistical methods based on decision-rules can be used for developing these signatures.

Subgroup Selection Methods

For single biomarkers: Resampling-based methods on tree-based or ROC-based thresholds

For thresholds using combination of biomarkers:

1.Adaptive Indexing Method - Tian & Tibshirani, 2010

2.PRIM (Patient Rule Induction Method) – Friedman, 1999

3.Hiernet (Lasso for hierarchical interactions) – Bien & Tibshirani, 2013

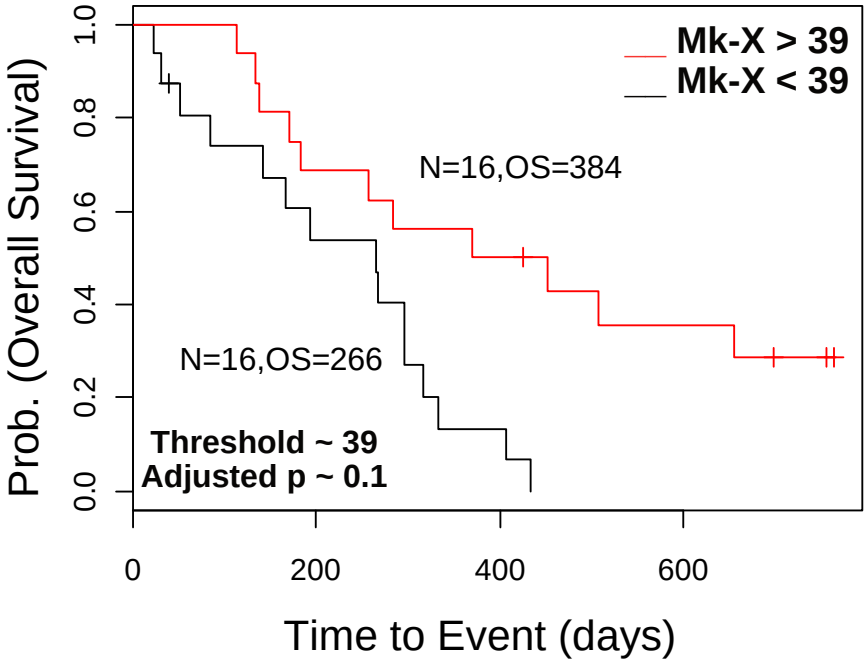
Active area of research. Some exciting methods coming soon.

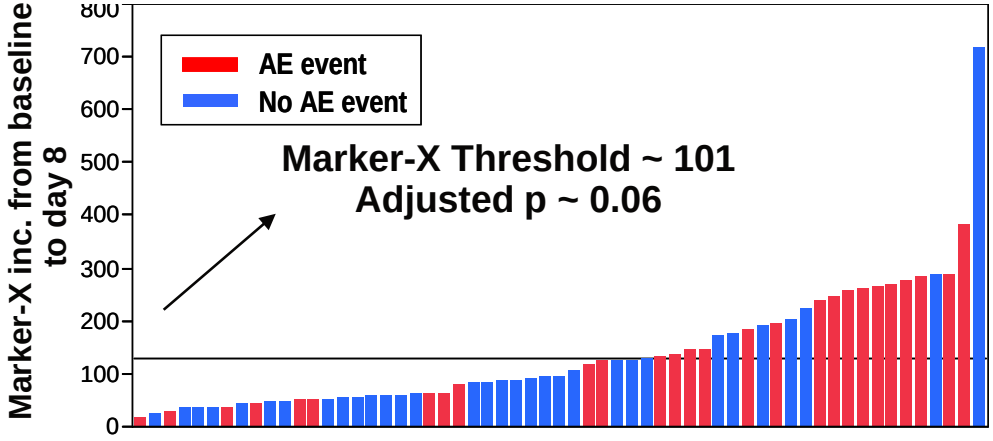
I highly recommend the AIM method (paper is easy to follow, R library available in CRAN).

Method for evaluating *Predictive Significance*

- Predictive Significance of cut-point signatures can be evaluated via 5-fold cross-validation (CV).
  - Stratification for patients in each fold were predicted by applying the algorithm on the remainder folds.
  - CV p-value was estimated after aggregating the predicted stratifications of all the left-out folds.
- Variability of the CV p-values is estimated by doing multiple iterations (50-100) of the above CV procedure.
- Following performance metrics are reported:
  - Median CV p-value (this is usually adequate)
  - Upper 95% empirical limit of the distribution of CV p-values, and
  - Percentage of CV p-values that are less than 0.05.

Illustration 1:
Biomarker Cut-Point for Efficacy & Safety





BATTing estimates of Threshold on Mk-X for Time to OS & AE are ~ 39 and 101 respectively.

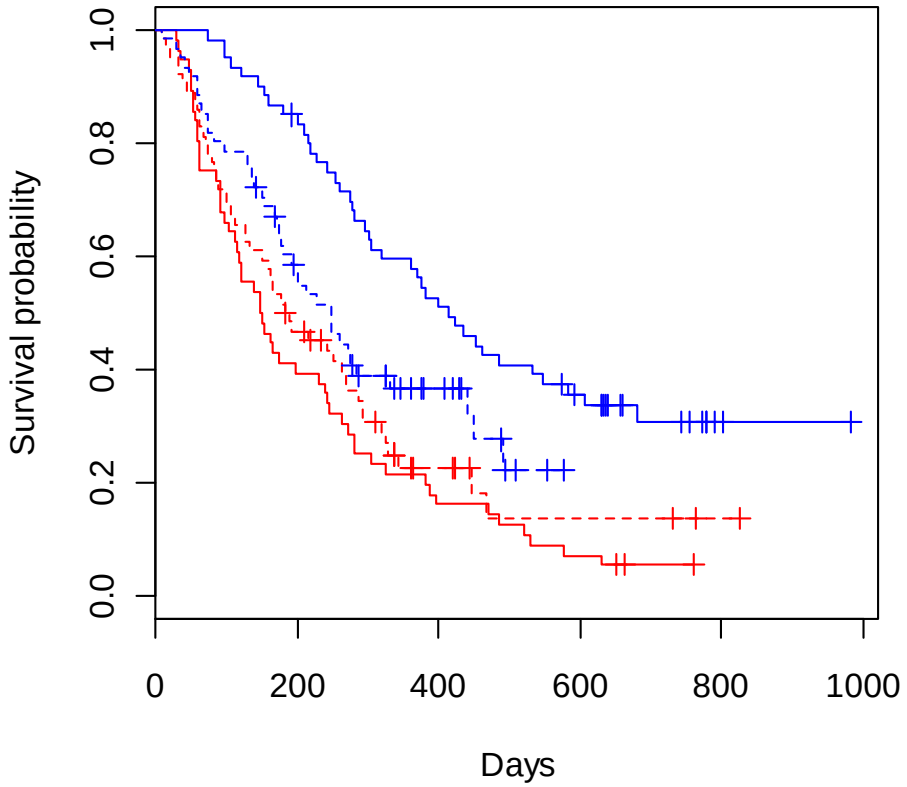
Suggests potential value of Mk-X as a marker for both efficacy & safety (therapeutic window: 39 to 101).

Bias-adjusted p-values show evidence of trend towards significance.

Simulations have shown that BATTing performance is stable for n > 50 per stratification arm.

Follow-up evaluations needed.

Illustration 2: Multivariate Biomarker Cut-Points for Prognostic & Predictive Signatures



Drug A, Sig +ve, N=60, MST=406

Other Drug, Sig +ve, N=65, MST=236

Drug A, Sig –ve, N=52, MST= 138

Other Drug, Sig –ve, N=55, MST=167

Biomarker	AIM BATTing
Mk-1	< 115
Mk-2	
Mk-3	
Mk-4	< 24.2
Mk-5	< 87.5
Mk-6	
Mk-7	
Mk-8	> 124.2

Drug A vs. Other Drug: **Signature +ve: p=0.09; Signature –ve: p=0.79**
Signature +ve vs. -ve: **Drug-A: p=0.008; Other Drug: p=0.34**

(1) Drug A works better for Signature + group.

(2) Signature + has better prognosis when receiving drug A.

Summary

Biomarker Signature Development

Predictive inference vs. Statistical inference

Statistical significance is often inadequate, and sometimes irrelevant.

A marker useless on it's own may be useful in a composite.

Numerous machine-learning methods are available, but *Simpler models (e.g., Logistic-Lasso) are often adequate.*

Fully embedded cross-validation/resampling procedure needed to evaluate the performance.

Signatures from targeted disease-pathway markers can yield similar performance as those from the whole-genome (hypothesis-free).

Methods such as AIM can yield readily usable prognostic/predictive signatures based on biomarker thresholds (attractive to Clinicians).