Clinical Utility of Diagnostic Tests

David A. Eberhard MD, PhD

Director, Pre-Clinical Genomic Pathology,
Lineberger Comprehensive Cancer Center

Associate Professor, Depts. of Pathology and Pharmacology

University of North Carolina – Chapel Hill

NAS National Cancer Policy Forum Workshop on Policy Issues in the Development and Adoption of Molecularly Targeted Therapies for Cancer

Nov 10-11, 2014

UNC LINEBERGER

UNC CANCER CARE

Uses of Diagnostic Tests or Biomarkers

- **Diagnostic**: Does the pt have a condition? What is the condition? What caused the condition?
- **Prognostic**: How is the pt going to do?
- **Predictive**: How will the pt respond to an intervention?
- **Pharmacodynamic; surrogate endpoint**: Is the intervention having an effect?

Steps in Diagnostic Test Development or Biomarker Qualification

1. Analytical Validation

How well the assay measures the molecular event of interest: Range, accuracy, precision, bias, assay/operator/instrument reproducibility

2. Clinical validation

Accuracy and predictability of assay (strength of association with w condition of interest)
Sensitivity, specificity, cutoffs, PPV/NPV, ROC etc
- In the intended clinical setting,
- On the sample types that will come from the intended pt population.

3. Clinical utility

What is it useful for? ***Use - specific fitness:***
- Provide value for use in health care?
- Support regulatory filings & decision making in product development?

Does it offer more than what we have now?

Woodcock, J. Clin Pharmacol Ther, 88:765 -73, 2010.
Febbo PG et al, J Natl Compr Canc Netw 9 (Suppl 5):S1-32, 2011.

Steps in Diagnostic Test Development or Biomarker Qualification

1. **Analytical Validation**

How well the assay measures the molecular event of interest: Range, accuracy, precision, bias, assay/operator/instrument reproducibility

Accuracy and predictability of assay (strength of associat
Sensitiv
- In the
- On the
intende

Diagnostic performance/accuracy
Can be part of clinical validation or clinical utility, depending on context
Test for particular analyte *vs*
Test that directly classifies pts into prognostic/predictive subgroups (e.g., genomic signatures)

2. **Clinical validation**

cific fitness:
- Provide value for use in health care?
- Support regulatory filings & decision making in product development?

Does it offer more than what we have now?

3. **Clinical utility**

Woodcock, J. Clin Pharmacol Ther, 88:765 -73, 2010.
Febbo PG et al, J Natl Compr Canc Netw 9 (Suppl 5):S1-32, 2011.

Clinical Utility Levels of Evidence

from Febbo PG et al. NCCN Task Force Report: Evaluating the Clinical Utility of Tumor Markers in Oncology. J Natl Compr Canc Netw 9 (Suppl 5):S1-32, 2011.

Table 1 Tumor Marker Utility Grading System Levels of Evidence

Level	Definition
I	Prospective, marker primary objective Well-powered or meta-analysis
II	Prospective, marker the secondary objective
III	Retrospective, outcomes, multivariate analysis (most currently published marker studies are level of evidence III)
IV	Retrospective, outcomes, univariate analysis
V	Retrospective, correlation with other marker, no outcomes

Adapted from Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst 1996;88:1464; with permission.

Clinical Utility: NCCN Task Force Report

Table 2 Use of Archived Tissues to Determine Clinical Validity of Tumor Markers

Category Trial Design	A Prospective	B Prospective Using Archived Samples	C Prospective/ Observational	D Retrospective/ Observational
Clinical trial	PCT designed to address tumor marker	Prospective trial not designed to address tumor marker, but design accommodates tumor marker utility	Prospective observational registry, treatment and follow-up not dictated	No prospective aspect to study
Patients and patient data	Prospectively enrolled, treated, and followed in PRCT	Prospectively enrolled, treated, and followed up in clinical trial and, especially if a predictive utility is considered, a PRCT addressing the treatment of interest	Prospectively enrolled in registry, but treatment and follow-up standard of care	No prospective stipulation of treatment or follow-up; patient data collected through retrospective chart review
Specimen collection, processing, and archival	Specimens collected, processed, and assayed for specific marker in real time	Specimens collected, processed, and archived prospectively using generic SOPs; assayed after trial completion	Specimens collected, processed, and archived prospectively using generic SOPs Assayed after trial completion	Specimens collected, processed, and archived with no prospective SOPs
Statistical design and analysis	Study powered to address tumor marker question	Study powered to address therapeutic question and underpowered to address tumor marker question Focused analysis plan for marker question developed before performing assays	Study not prospectively powered at all; retrospective study design confounded by selection of specimens for study Focused analysis plan for marker question developed before performing assays	Study not prospectively powered at all; retrospective study design confounded by selection of specimens for study No focused analysis plan for marker question developed before performing assays
Validation	Result unlikely to be play of chance Although preferred, validation not required	Result more likely to be play of chance than A, but less likely than C Requires one or more validation studies	Result very likely to be play of chance Requires subsequent validation studies	Result very likely to be play of chance. Requires subsequent validation studies

Abbreviations: PCT, prospective controlled trial; PRCT, prospective, randomized controlled trial; SOP, standard operating procedure.
Adapted from Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009;101:1449; with permission.

Febbo PG et al J Natl Compr Canc Netw 9 (Suppl 5):S1-32, 2011.

Clinical Utility: NCCN Task Force Report

Table 3 Revised Determination of Levels of Evidence Using Elements of Tumor Marker Studies*

Level of Evidence	Category From Table 2	Validation Studies Available
I	A	None required
I	B	One or more with consistent results
II	B	None or inconsistent results
II	C	2 or more with consistent results
III	C	None or 1 with consistent results or inconsistent results
IV–V	D	NA†

*Levels of evidence revised from those originally proposed in Tables 1 and 2.[31]
†Not applicable (NA) because level of evidence IV and V studies will never be satisfactory for determination of medical utility.
From Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009;101:1450; with permission.

Table 4 NCCN Categories of Evidence and Consensus

Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate.

Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.

Category 2B: Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate.

Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.

Febbo PG et al. NCCN Task Force Report: Evaluating the Clinical Utility of Tumor Markers in Oncology. J Natl Compr Canc Netw 9 (Suppl 5):S1-32, 2011.

Clinical Utility vs Biomarker Qualification

from Woodcock, J. Clin Pharmacol Ther, 88:765 -73, 2010.

- "What is the test useful for?" in drug development
  - "Can the evidence from the assay be used in regulatory filings and to support decision making?" = *biomarker qualification*
- Fitness for use to generate supporting evidence
  - Re drug safety, efficacy, dosing, patient selection, etc.
- Establishes global, rather than product-specific, fitness for use
  - Information generated, for the specific use, is reliable and will be acceptable to regulators
- Companion diagnostic development
  - Biomarker assay for use with a specific drug
  - Involves evaluating its value for use in health care, i.e. "clinical utility."
- CDER Biomarker Qualification Program
  - Provides framework for scientific development and regulatory acceptance of biomarkers for use in drug development

Biomarker Assays During Drug Development and Use

Woodcock, J.
Clin Pharmacol Ther
88:765 -73, 2010.

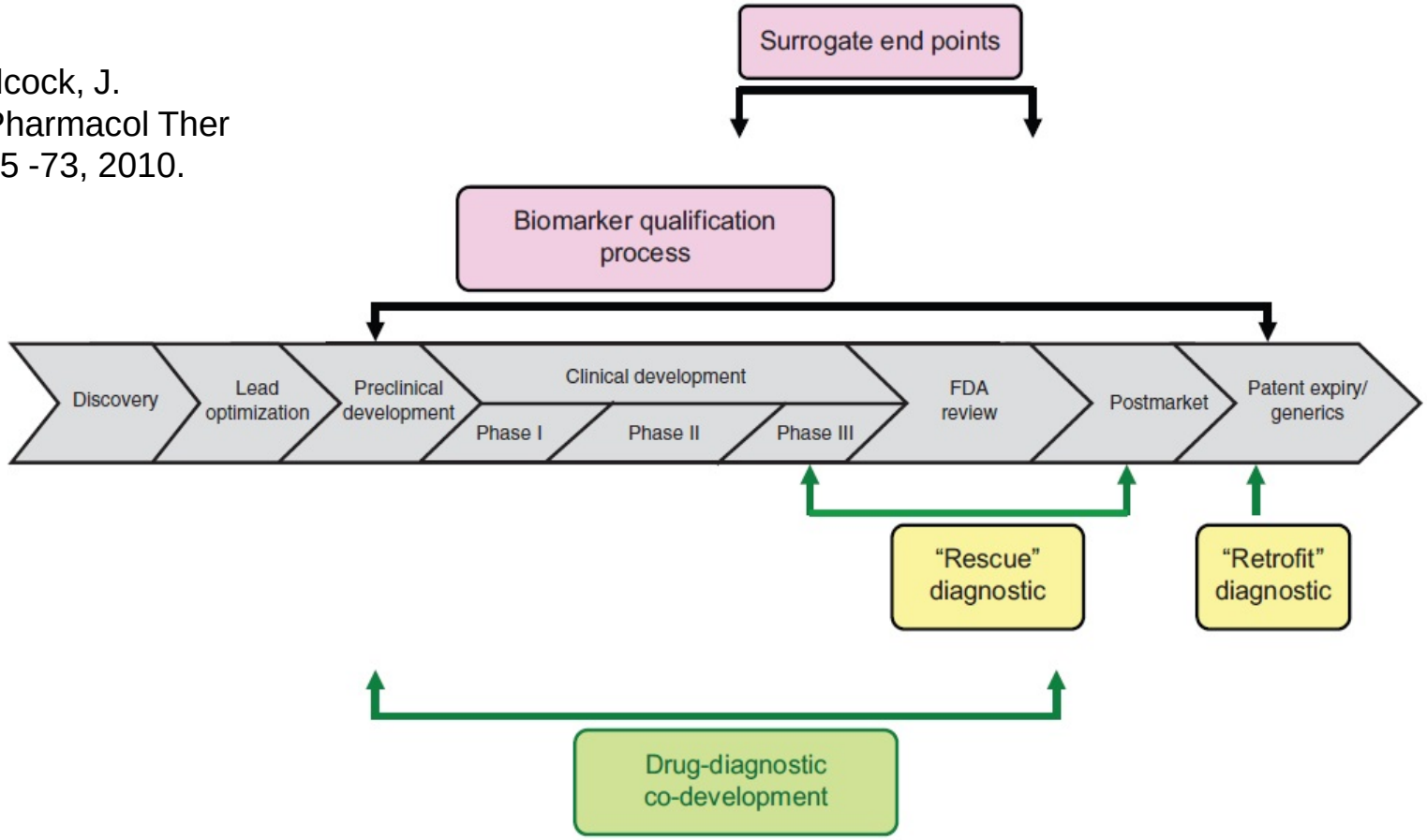


Figure 1 Introduction of new biomarker assays during drug development and use. The figure shows the timing of introduction of new diagnostics with respect to the drug development pipeline. Publicly available processes are shown above the pipeline, drug-specific processes below. In drug–diagnostic co-development, an investigational drug is intended, from the early stages, to be used with a candidate diagnostic test. "Rescue" diagnostics are introduced late in the drug development process in order to improve drug performance, whereas "retrofit" diagnostics are applied to long-marketed drugs to remedy problems related to safety, effectiveness, or dosing. "Biomarker qualification" involves regulatory acceptance of a diagnostic for a specific use during drug development. There is currently no formal regulatory process for acceptance of new surrogate end points.

Clinical Utility

- Need for test: Can a “need to be filled” be defined in terms of “problem to be solved”?
  - Use formalized approaches such as Root Cause Analysis to define and address?

- Quality of test: Diagnostic accuracy and reproducibility.
  - Clinical validation: does it do what it is supposed to….
  - Clinical utility: ….in a way that fills a clinical need?

- Fitness for use: Implementability, usability.
  - A clinically useful test must be able to be implemented in the setting where it is meant to be used

Root Cause Analysis

(adapted from Wikipedia)

Root cause analysis (**RCA**) is a method of problem solving that tries to identify the root causes of faults or problems. A root cause is a cause that once removed from the problem fault sequence, prevents the final undesirable event from recurring.

Some general principles of root cause analysis

- Identify the factors that resulted in the harmful outcomes (consequences) of past events in order to identify what needs change to prevent recurrence and lessons to be learned
- Performed systematically, with conclusions and root causes that are backed up by documented evidence.
- There may be more than one root cause for a problem.
- Solutions intend to prevent recurrence at lowest cost in the simplest way. If there are alternatives that are equally effective, then the simplest or lowest cost approach is preferred.
- Root causes identified depend on the way in which the problem or event is defined. Need effective problem statements and event descriptions.
- Analysis should establish a sequence of events to understand relationships between contributory (causal) factors, root cause(s) and the defined problem.
- Root cause analysis can help transform a reactive culture (that reacts to problems) into a forward-looking culture that solves problems before they occur or escalate.
- Root cause analysis is a threat to many cultures and environments. Threats to cultures often meet with resistance.

Example

- **SITUATION**: Drugs are approved for use in non-squamous non-small cell lung carcinoma
- **PROBLEM**: Histopath Dx of NSCLC is imprecise and inaccurate
- **SOLUTION**: Create more precise and accurate ways to diagnose NSCLC subtypes
- **INTENDED RESULT**: Better clinical treatment decisions? Change in label of drug to include test?

1. Grilley-Olson JE et al, Arch Pathol Lab Med 2012.
2. Thunnissen E et al, J Thorac Oncol 2014

Let's look at this more closely…

- **SITUATION**: Benefit or safety in clinical trials showed some association to histopath Dx subtype.
- **PROBLEM**: Histopath Dx of NSCLC can be imprecise and inaccurate. This could lead to mis-association of Dx to outcome in clinical trial or to suboptimal Tx of pt in clinic. differentiation) Biologically different tumors (e.g. well vs. poorly diff) may not have same responses to Tx.
- **CAUSE**: Accurate and reproducible subtyping can be compromised by to sampling (small size), interpretation (lack of experience) or biology (poor differentiation). [1,2]
- **SOLUTION**: New "AdenoCa vs SqCCa" diagnostics may be useful if they make the same call on small biopsies as would have been made on a larger definitive sample of the same tumor. Tests that would change the Dx of a definitive sample (e.g. from Undiff Ca to SqCCa) may not be useful for Tx decisions unless directly evaluated against clinical outcome or surrogate.

1. Grilley-Olson JE et al, Arch Pathol Lab Med 2012.
2. Thunnissen E et al, J Thorac Oncol 2014

Fitness For Use: Implementability

- Platform and assay: Suitability, robustness, complexity, expense
  - LDTs can offer flexibility, rapid deployment to serve a need
  - Does test need a special environment (central lab) to be performed properly, or can it be done in independent labs or sold as a kit?
- Sample characteristics: Define and control
  - Preanalytical: Size/quantity; processing or fixation
  - Sample presentation: e.g tissue microarrays vs single slides
- Interpretation: Process and report
  - Is there a process for robust, reliable, reproducible interpretation or analysis of data to deliver the final result to the clinician?
  - Final result is what has to have clinical utility
  - Is “how to use the result” given as part of the report, presumed to be common knowledge, or just avoided?

Fitness For Use: Gray Zones

- Gray zones can be technical (analytical validation)
  - Lack of precision
  - Continuous variable data need thresholds to convert to -/+ classification or Y/N decisions
  - Discontinous variable data can require statistical strength, e.g. mutation calling for NGS

- Black/white data can have gray zones in levels of evidence to support a decision (clinical validation and utility)

- Gray zones can be due to lack of clear definitions or incomplete situational analyses (clinical utility)
  - What do we want? What do we have? What do we do?

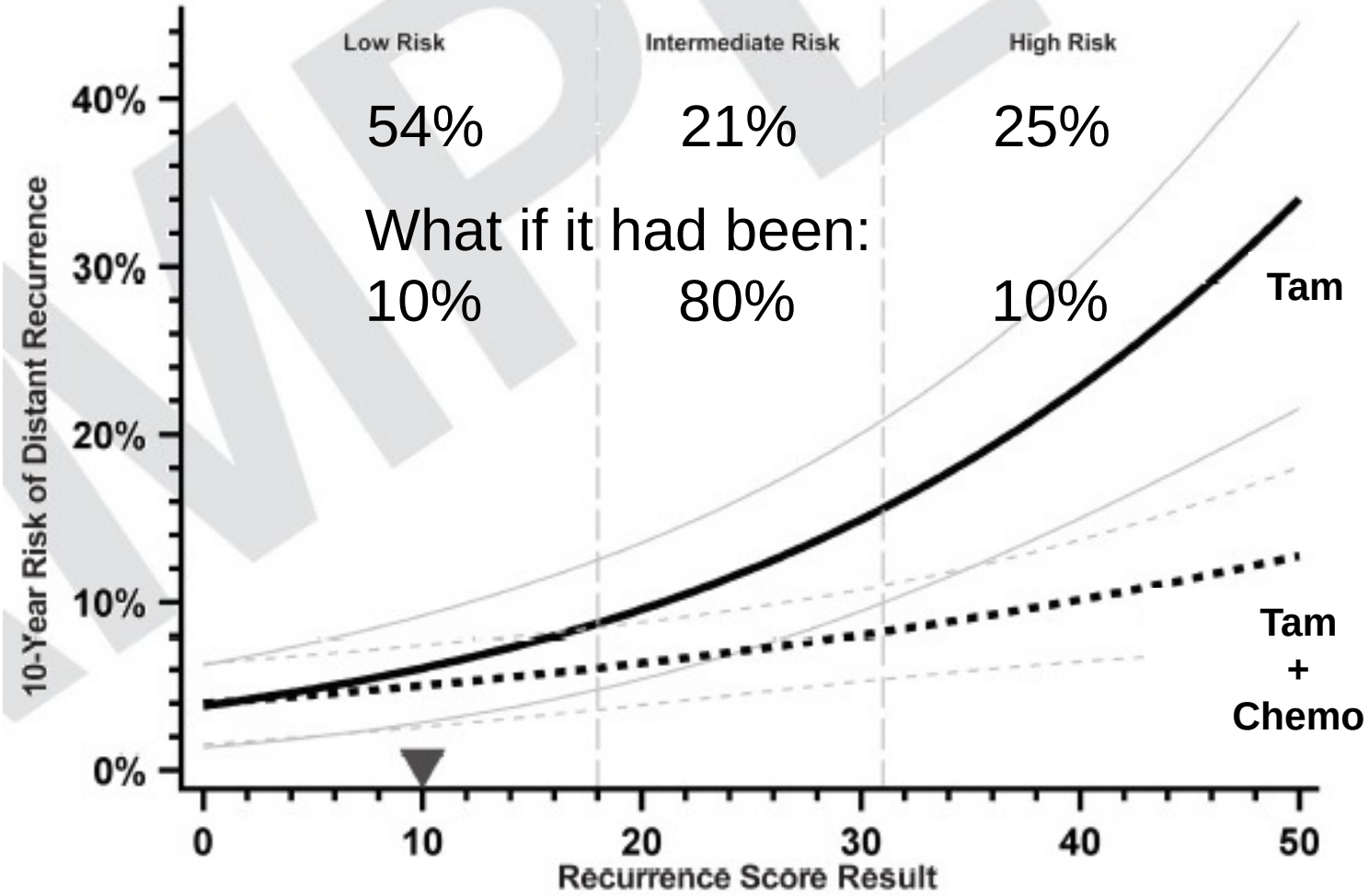
Gray Zone Examples

- Her2 IHC
  - If 2+, reflex new test (FISH) per ASCO-CAP guidelines
- Oncotype Dx “intermediate”
- VUS identified by NGS
  - EGFR TKI-sensitizing mutations (e19del, L858R) are NCCN category 1 and combined level of evidence score 1A
  - EGFR e20 insertion may predict resistance

Genomic Health® | oncotype DX® Breast Cancer Assay

Breast Cancer Report - Node Negative
Prediction of Chemotherapy Benefit

Genomic Health, Inc.
301 Penobscot Drive, Redwood City, CA 94063 USA
USA/Canada: +1.866.ONCOTYPE
International: www.oncotypedx.com/contact
www.oncotypedx.com
CLIA Number 05D1018272



NSABP B20

Absolute Benefit of Chemotherapy at 10 Years by Recurrence Score Risk Group

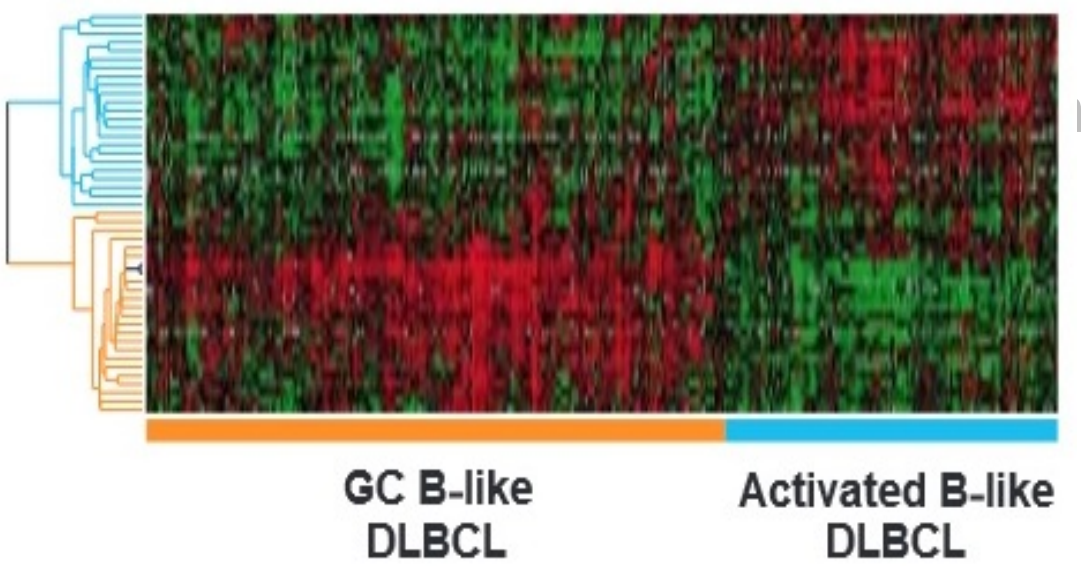
Paik S et al. J Clin Oncol 2006 Aug 10;24(23):3726-34

Different tests that are aimed at doing the same thing

- The clinical utility proposition is the same. Or is it?
  - What are the differences?
- DLBCL subtyping
  - Gene Expression profiling using arrays
  - IHC decision tree algorithms
  - Nanostring 15+5 gene FFPE GE panel (Lymph2Cx)

Different tests that are aimed at doing the same thing

- The clinical utility proposition is the same. Or is it?
  - What are the differences?
- DLBCL subtyping
  - Gene Expression profiling using arrays

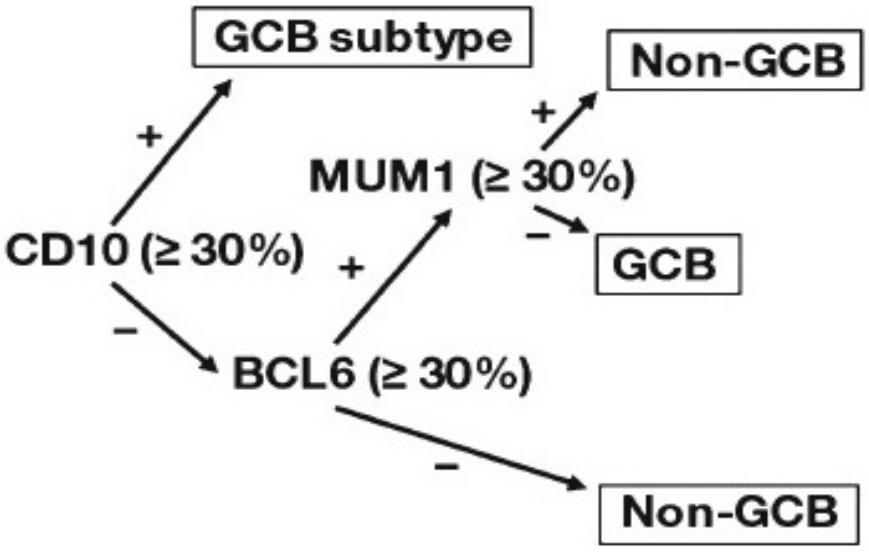


- Frozen samples.
- Complex tech
- Signature gives strength in numbers of genes.

Alizadeh AA et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000 Feb 3;403(6769):503-11

Different tests that are aimed at doing the same thing

- The clinical utility proposition is the same. Or is it?
  - What are the differences?
- DLBCL subtyping
  - Gene expression profiling using arrays
  - IHC decision tree algorithms



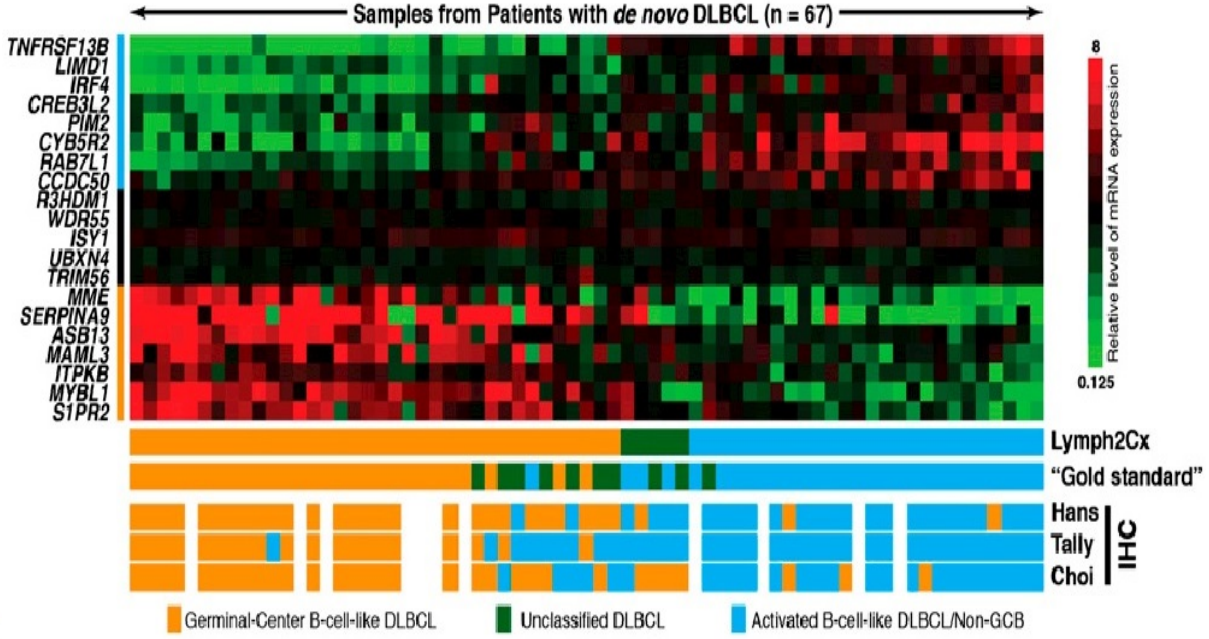
- FFPE samples.
- "Simple" tech
- Each 'gene' must stand alone – no weak links allowed.

Hans CP et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 2004 Jan 1;103(1):275-82

Different tests that are aimed at doing the same thing

- The clinical utility proposition is the same. Or is it?
  - What are the differences?
- DLBCL subtyping
  - Gene expression profiling using arrays
  - IHC decision tree algorithms
  - Nanostring 15+5 gene expression panel (Lymph2Cx)

Scott DW et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood 2014 Feb 20;123(8):1214-7



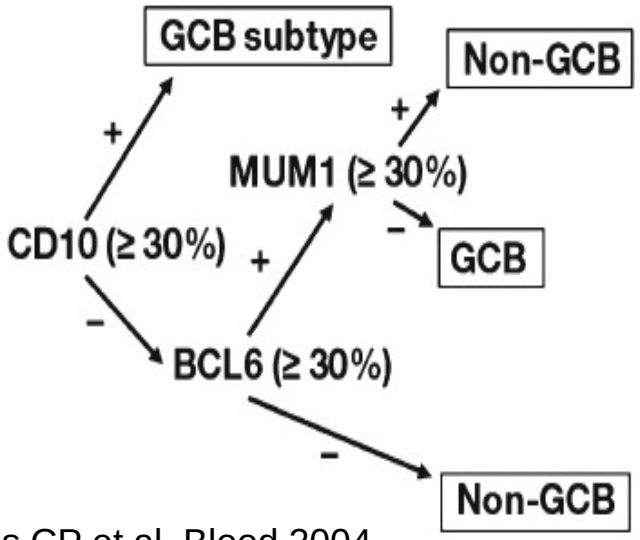
- Frozen samples.
- Complex tech
- Locked model w gene coefficients, thresholds, and quality criteria.

One Approach to DLBCL Drug Development Program with Companion Diagnostic for Subtyping

- **Testing:** Heise C et al. Implementing a Multi-analyte Immunohistochemistry Panel into a Drug Development Program. Methods in Pharmacology and Toxicology, Springer, in press.

- **Clinical:** Czuczman MS et al., A Phase 2/3 Multicenter, Randomized Study Comparing the Efficacy and Safety of Lenalidomide Versus Investigator's Choice in Relapsed/Refractory DLBCL. Submitted, ASH Annual Meeting 2014.

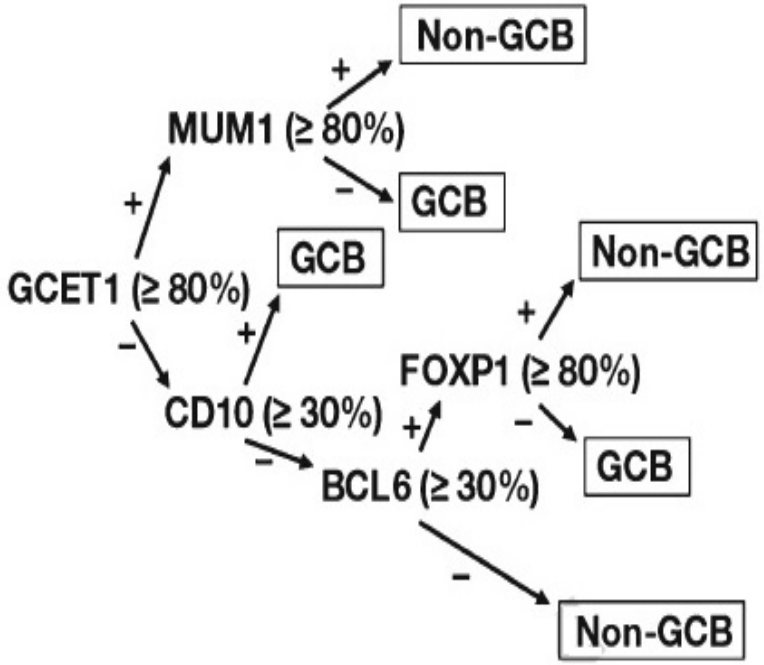
Hans Criteria



Hans CP et al. Blood 2004 Jan 1;103(1):275-82

Choi Criteria

Choi WW et al. Clin Cancer Res 2009 Sep 1;15(17):5494-502



One Approach to DLBCL Drug Development Program with Companion Diagnostic for Subtyping

Assay Optimization and Technical Validation

3 labs perform IHC panel on DLBCL TMA with GEP data:

- Share protocol for IHC assays
- Examine inter-lab reproducibility (concordance)
- Identify sources of inter-lab variation
- Optimize for comparable performance across labs

Heise C et al. Implementing a Multi-analyte Immunohistochemistry Panel into a Drug Development Program. Methods in Pharmacology and Toxicology, Springer, in press.

Clinical Evaluation

Perform assays on sections from ph II clinical trial:

- Independently in each lab, blinded to other labs
- IHC assays and interp algorithms for final test result
- Examine inter-lab concordance to decide if test is robust enough for use in registrational trials

Demonstration of Clinical Utility

Transfer locked-down test protocol to central lab:

- Prospective use in registrational study
- Use to stratify or select pts for treatment
- Discuss co-development and implementation path with FDA
- Basis for simultaneous approval of drug and CDx

Thank you!

David_eberhard@med.unc.edu