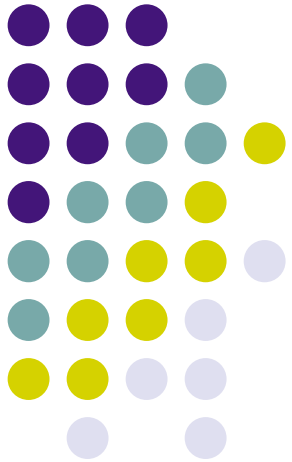
Right Drugs for the Right Patients

Or what keeps an industry statistician up at night

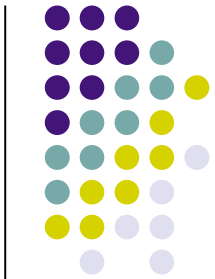


Jane Fridlyand

Genentech

SFASA Chapter Symposium
Hayward State University
May 14, 2011

Acknowledgements



- Ru-Fang Yeh
- Gracie Lieberman
- Mark Lackner
- Mika Derynck
- PI3K Program Team
- Biostatistics Predictive Biomarker Guidance 2010 Team

Future of Personalized Health Care (PHC)?

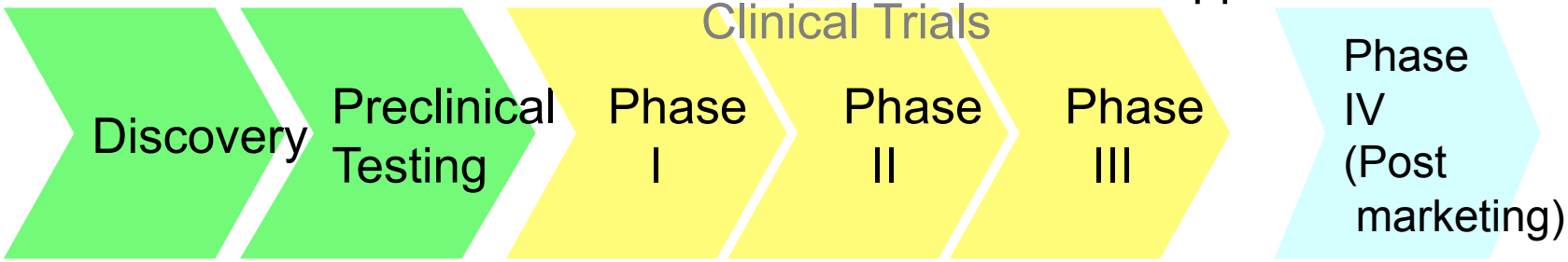
"Here's my DNA sequence."

From Garett Hampton [3]

“The goal of clinical research is not to obtain a statistically significant effect. Rather, "the primary goal should be to obtain a statistically reliable evaluation regarding whether the experimental intervention is safe and provides clinically meaningful benefit" “

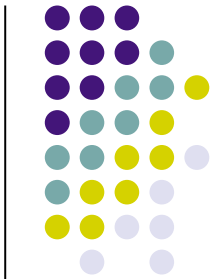
–Thomas Fleming

Drug Development



years	**6.5**	**1.5-3**	**2-3**	**3.5-5**	>12-15 years
Test population	**Laboratory and animal studies**	**<100 pts**	**~100-150**	**~500-1000**	
Purpose	**Assess safety, biological activity & formulations**	**Determine safety & dosage**	**Evaluate effectiveness & side effects**	**Confirm effectivenes; monitor adverse reactions**	Average Cost: ~$500-800M
Success rate	**5000 compounds evaluated**	**5 entered clinical trials**			1 approved

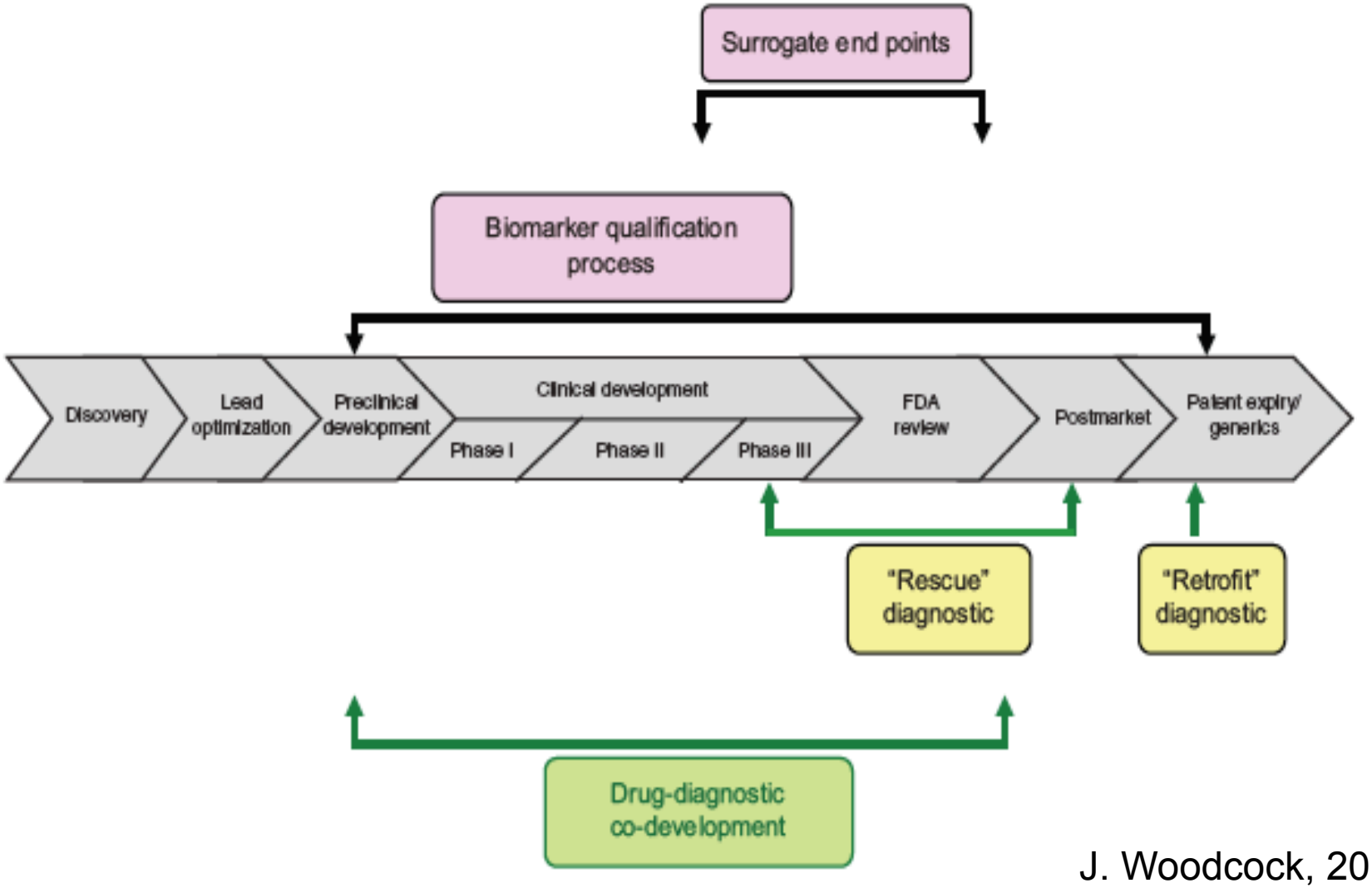
Context: Key scientific and clinical challenges



- Most of the today's focus in oncology drug development is on targeted therapies which are expected to be active only in subsets of patients

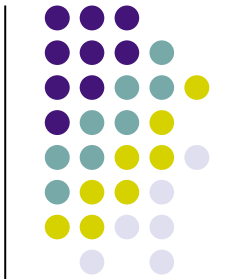
- True targeted therapy requires not just a selective agent but a means of identifying appropriate patients (*ie* a diagnostic)

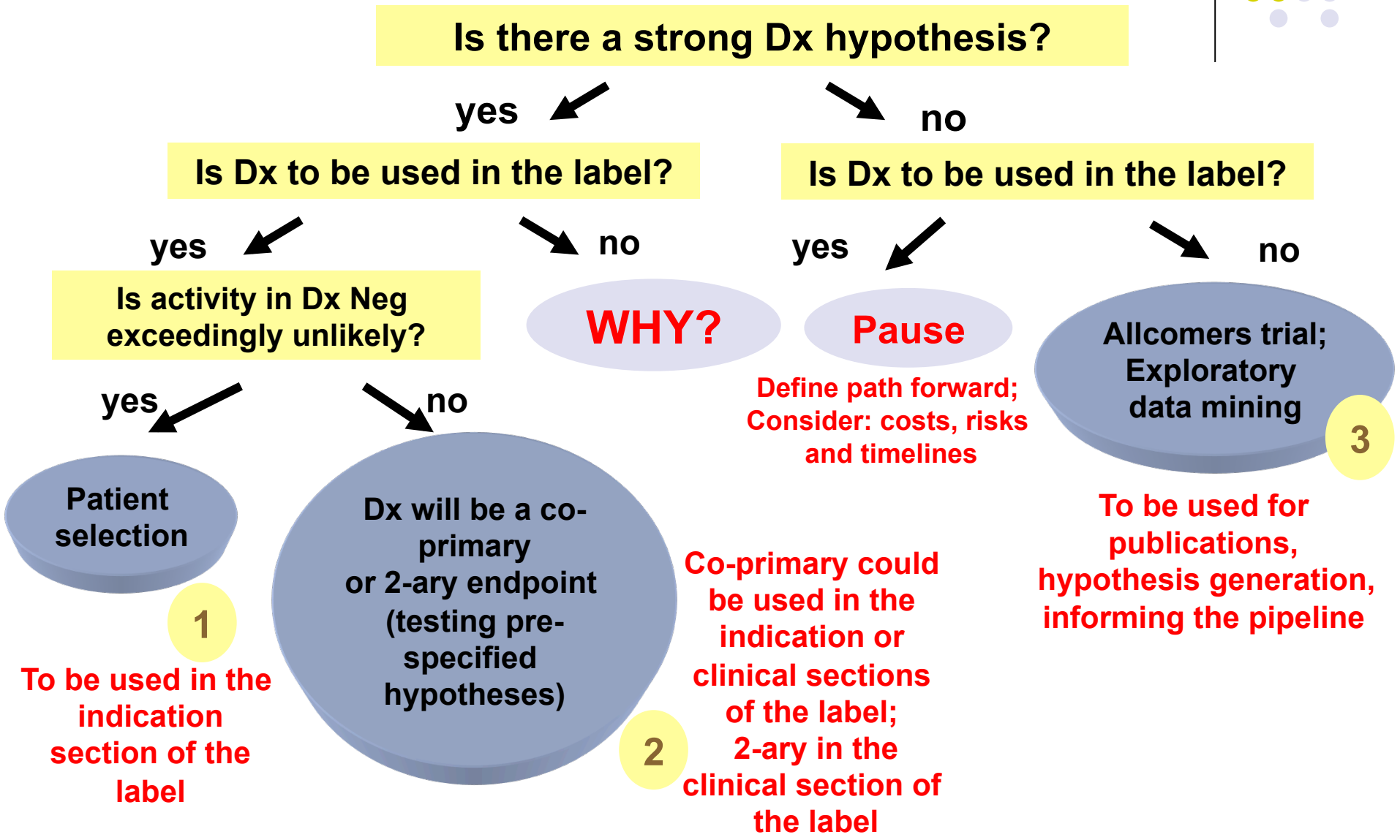
Biomarkers and drug development



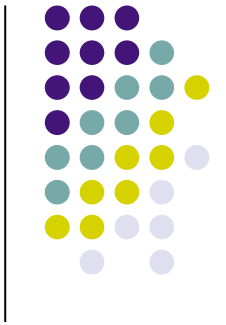
J. Woodcock, 2010

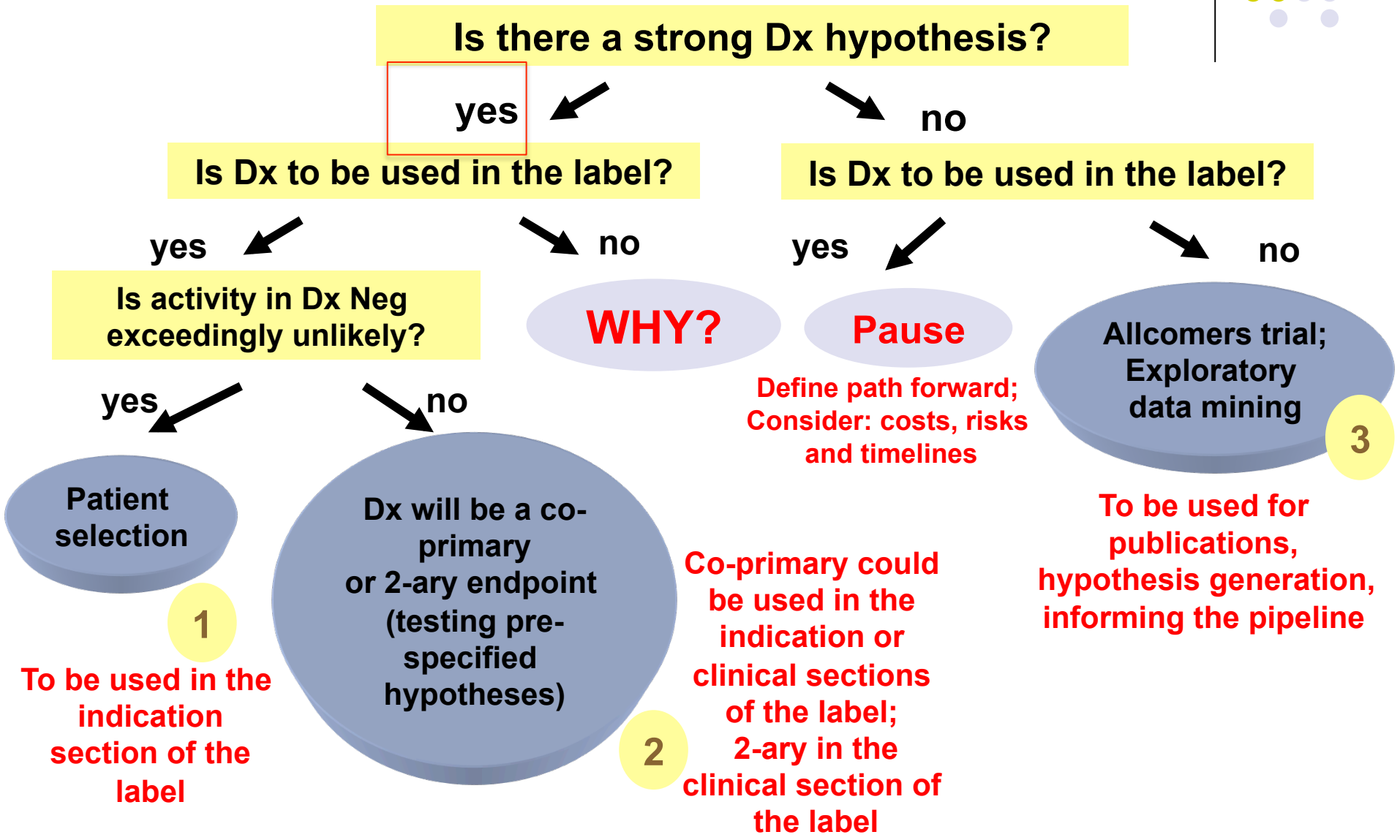
Underlying Assumption



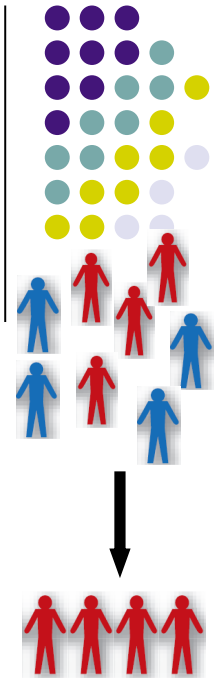


Underlying Assumption





Drug-Dx co-development

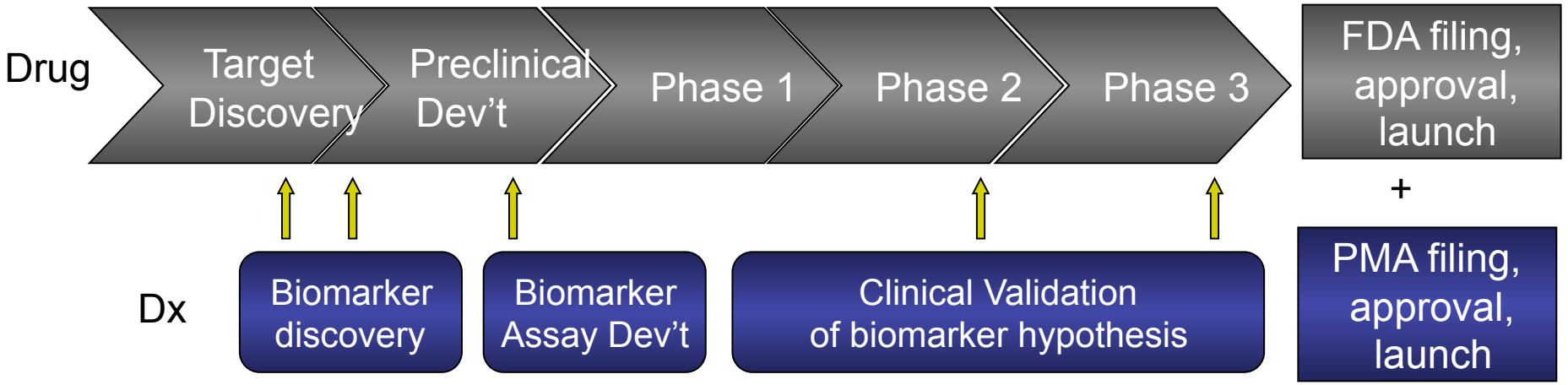


- **To maximize clinical benefit from our therapeutics:**
  - Enable patient selection
  - Informed decision making around indication choice

Predictive biomarker: a test that can be done before treatment to predict whether a particular treatment is likely to be beneficial

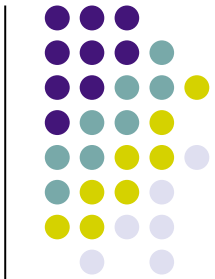
Pharmacodynamic biomarker: a test that can be done pre- and post-treatment to confirm target modulation

- **Current regulatory paradigm requires early biomarker discovery**



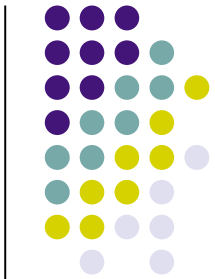
FDA draft white paper on Drug-Dx co-development

Basic Principles



- Retrospective Analyses are only considered “hypothesis” generating.
- Prospectively Defined Dx markers are required for any label enabling action.
- Dx markers must be defined prior to pivotal trials and hence planned for in a Clinical Development Plan (CDP).
- Dx. Marker evaluation often in phase II.

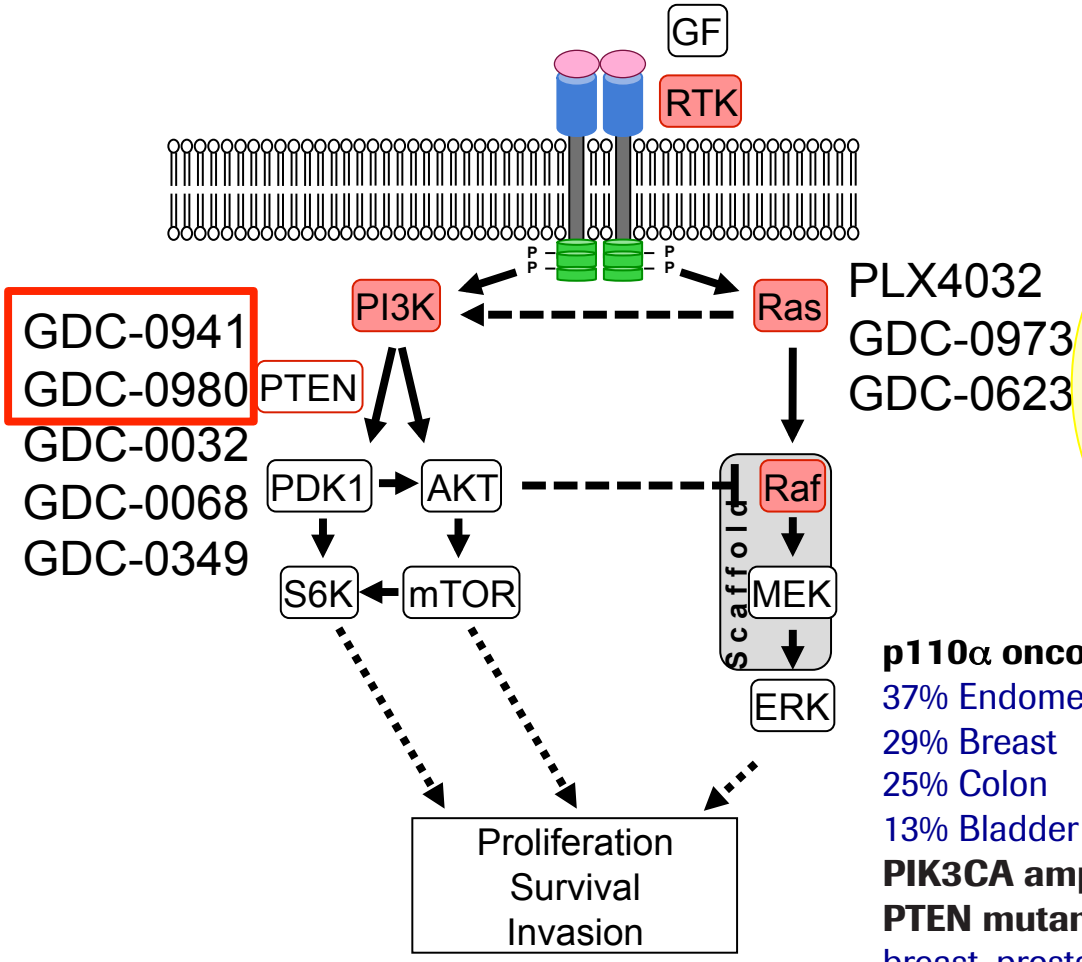
What does Dx hypothesis usually look like?

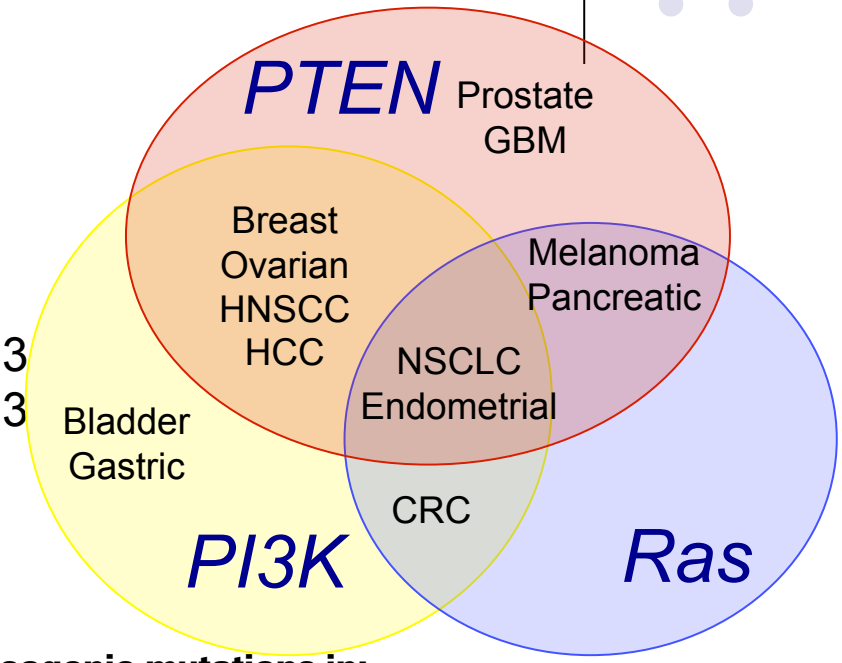


- Availability of a well defined biomarker hypothesis
  - One or at most two research-grade assays
  - Single summary measure
  - Appropriate cutoff available or to be derived for ordinal/ continuous biomarkers
- Scientific evidence based on the Mechanism of Action (MOA)
- In-vitro and xenograft assays
- Clinical evidence from our and competitor's trials *(Phase I is unlikely to yield useful efficacy information)*
- Prevalence of the proposed biomarker (description of the distribution if continuous) and known prognostic characteristics (given the line and indication)
  - Literature reports
  - Public databases
  - Should we consider a separate tumor registry to address the question?
  - If prognostic, is genomic drift a potential issue?
    - Archival vs Fresh Biopsies

Dedicated studies may be necessary to investigate assay or biomarker properties, prevalences and prognostic significance

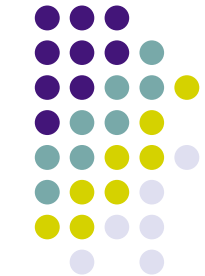
Multiple molecules targeting core oncogenic pathways: Potential broad application, combinations, with strong biomarker hypotheses





p110α oncogenic mutations in:
37% Endometrial
29% Breast
25% Colon
13% Bladder
PIK3CA amplified in 30% ovarian, lung
PTEN mutant/lost in:
breast, prostate, glioblastoma, melanoma, pancreatic, endometrial, ovarian, lung, head and neck, hepatocellular, thyroid

Developing PI3K Inhibitors: *Follow the Tumor Genetics*



–Which target: selective or multiple

–Which disease

- Reliance on different nodes in the pathway (RTKs, PI3K mt, mTOR)
- Translatability of Proof Of Concept (POC) across diseases, line, combinations

–Which patient (role of diagnostics)

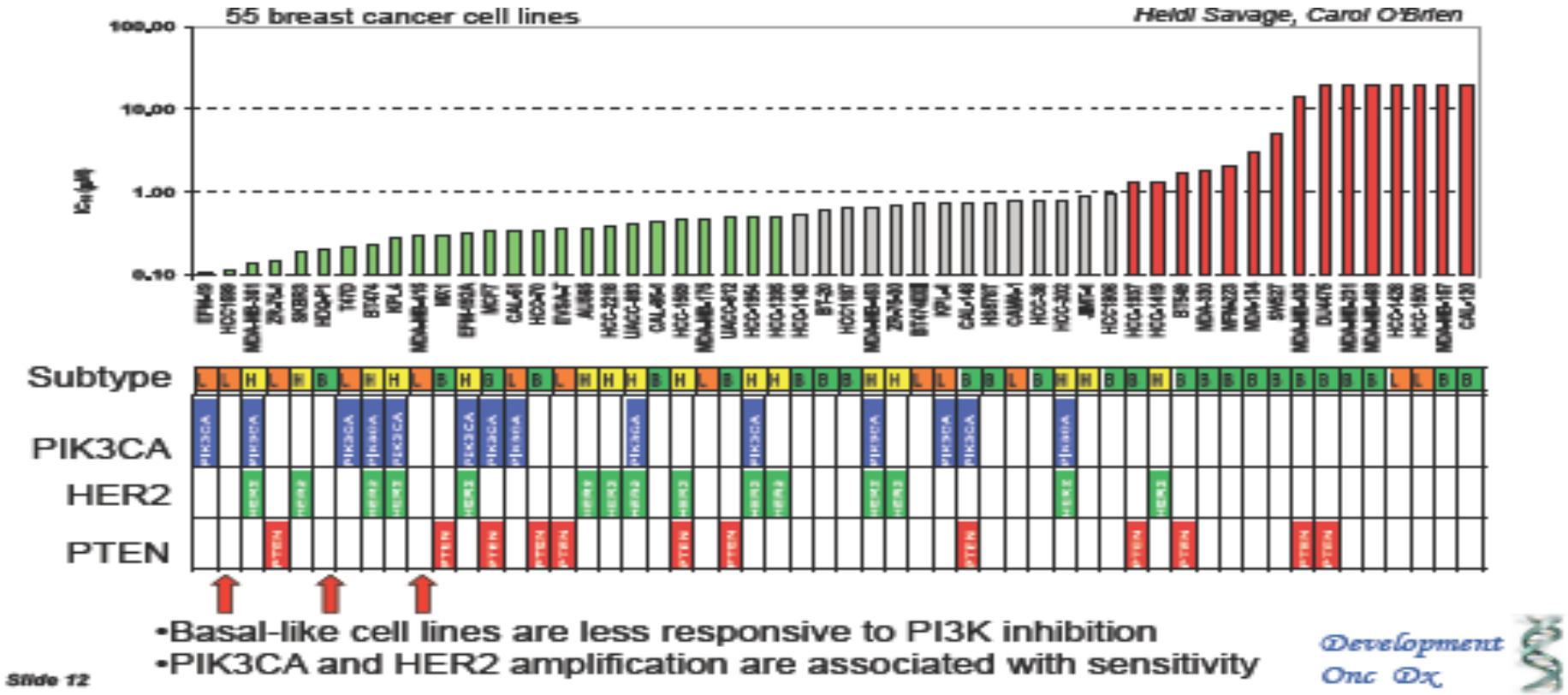
- Putative Dx markers have strong links to tumor biology
- Are pathway alterations predictive of target dependence
    - Can we select patients based on any of them?

–Single agent vs. combination

- Chemotherapy, EGFR, MEK, VEGF…..

–How hard to hit the target (dose and schedule)

PHC assessment to guide development strategy

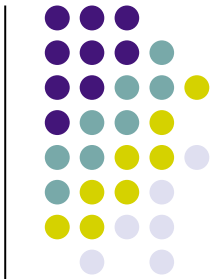


- Preclinical data in mBC suggests that cell lines harboring PI3K/PTEN alterations are highly sensitive to the pathway inhibitors

PHC Assessment ⟷ Development Strategy

PHC Assessment		
• Strong Dx hypothesis • No activity in Dx-	• Strong Dx hypothesis • Some activity in Dx-	• No strong Dx hypothesis • Exploratory Stage
Development Strategy		
• Patient selection through all phases of development	• Complex, larger phase IIs with stratification • Complex phase IIIs	• No selection or stratification • Data exploration
Selected	**Stratified**	**AllComers**

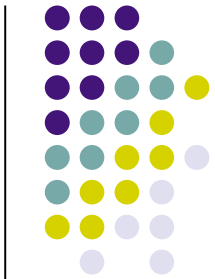
Stratified scenario - Impact on components of CDP



- Target product profile (TPP)
  - Parallel development of companion diagnostic
- Phase I trials
  - Selection for quick signal seeking
- Phase II trials
  - Complex issues become more complex
  - More unknowns, more questions to answer
- Phase III trials
  - Clinical Validation of Dx
  - Design depends on Phase II outcome
    - Selection, stratification or all-comers

Longer and Costlier but this is reality!

Phase II Considerations



- Objective: simultaneous Rx/Dx evaluation
- Scientific rationale and pre-clinical data - main determinants of the scenario prior to Phase II
- Statistical considerations
  - Co-primary endpoints
  - Value added and feasibility of stratification
  - Defining cut-offs for continuous biomarker
  - Go/No Go decision algorithm
- Dedicated studies to investigate assay or biomarker properties
  - Reproducibility, prevalence, prognostic value

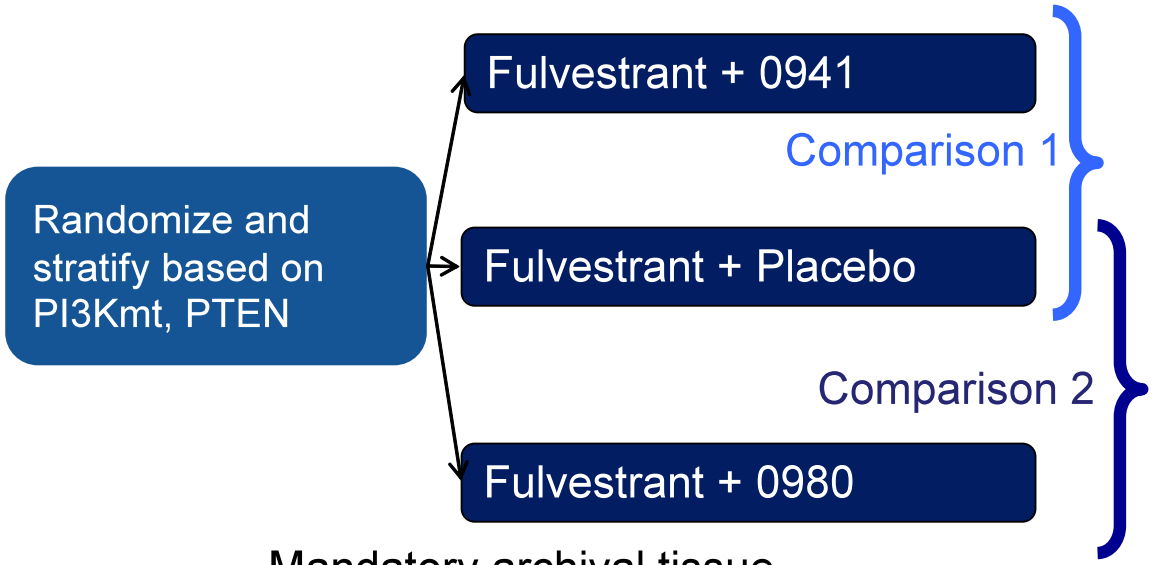
Possible POC Plans for GDC-0941 and GDC-0980

- Selected expansion cohorts in ongoing phase Ia studies
  - Expansion in patients with PIK3CA mutations
  - Separate breast cancer-only and multi-indication expansions
  - Both GDC-0941 and GDC-0980

- Randomized 3-arm Phase II Trial in ER+ Breast Cancer

Key Elig	•prior AI in adjuvant (PD<6mo) or metastatic breast CA •HR+ and HER2- •Measurable and evaluable disease

Primary endpoints:
- PFS in all comers
- PFS in Dx selected
- 350 pts



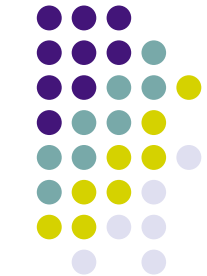
Mandatory archival tissue
Multiple exploratory Dx assays

Proof of Concept Designs

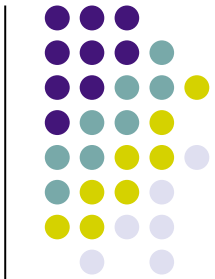
Overall Considerations

- **Rank speed, PTS and cost** according to their relative importance taking into account company portfolio and competitive landscape
- In Stratified scenario, need to define and enable **joint Rx/Dx** GO decision prior to Phase III
- **Single Arm versus Randomized Controlled trial**: in PHC setting the value of randomized trial is even higher in general than in all-comers development, although the same general drug development considerations apply

- **How does the proposed POC trial fit into the clinical development plan (CDP) for the molecule and overall PHC plan?**
  - **Overall strategic context and competitive landscape** (*what's Novartis doing?* ☺)
  - **The TPP with the Dx component** (clinical and commercial considerations)
    - The baseline scenario influenced by the available scientific/pre-clinical and possibly (minimal) clinical evidence
  - **The molecule CDP timeline and other ongoing/planned POC trials** where
    - The same Dx hypothesis may be addressed
    - Need to be informed with regards to the Dx by the proposed POC trial
  - **The follow up molecules** which need to be informed by this trial

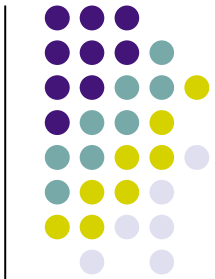


Operational Considerations



- Enrollment rate
- Tissue testing
  - Turn-around time
  - Assay-failure proportion
- Number of qualified sites and the ramp-up curve
- Population Dx prevalence
- Evaluation of sensitivity to the assumptions
- Protocol nuances…

Statistical Considerations



The design of the POC trial needs to enable a decision on the population and co-primary endpoints in Phase III, i.e. all the Dx subsets of primary interest need to be sufficient populated

- Propose and evaluate a decision algorithm operational characteristics and its robustness to the departures from the assumptions under a variety of the underlying treatment/biomarker scenarios, including the decisions to be made on multiple markers

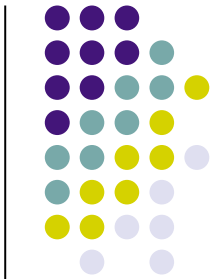
- For ordinal markers with pre-defined cutoffs, the proposed sample size in each biomarker sub-group of primary interest should be roughly similar to the size of the AllComers trial without the Dx (i.e. 30-50 events per biomarker subgroup)

- Significantly larger sample sizes may be required for a continuous biomarker, possibly on the order of 100's or 1000's of events. *As of today, the question of identifying an appropriate cutoff for a truly continuous biomarker remains an open problem from all, statistical, regulatory and clinical perspectives.*

Sample Size Considerations

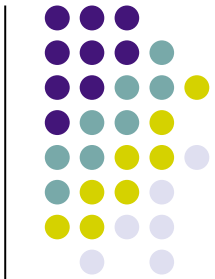
Num of events / True HR (GO Rule)	20	30	40	50	60	70
0.30 (< 0.65)	0.96 *(0.17, 0.53)*	0.98 *(0.19, 0.48)*	0.99 *(0.20, 0.45)*	1.00 *(0.21, 0.43)*	1.00 *(0.22, 0.42)*	1.00 *(0.22, 0.41)*
0.50 (< 0.65)	0.72 *(0.28, 0.89)*	0.76 *(0.31, 0.80)*	0.80 *(0.33, 0.75)*	0.82 *(0.35, 0.72)*	0.85 *(0.36, 0.70)*	0.86 *(0.37, 0.68)*
0.60 (<0.75)	0.69 *(0.34, 1.06)*	0.73 *(0.38, 0.96)*	0.76 *(0.40, 0.90)*	0.78 *(0.42, 0.86)*	0.81 *(0.43, 0.84)*	0.82 *(0.44, 0.82)*
0.70 (<0.75)	0.56 *(0.39, 1.24)*	0.57 *(0.44, 1.12)*	0.59 *(0.47, 1.05)*	0.60 *(0.49, 1.01)*	0.61 *(0.50, 0.97)*	0.61 *(0.52, 0.95)*
0.90 (<0.75)	0.34 *(0.51, 1.60)*	0.31 *(0.56, 1.44)*	0.28 *(0.60, 1.35)*	0.26 *(0.63, 1.29)*	0.24 *(0.65, 1.25)*	0.22 *(0.66, 1.22)*
1.00 (<0.75)	0.26 *(0.56, 1.77)*	0.22 *(0.63, 1.60)*	0.18 *(0.67, 1.50)*	0.15 *(0.70, 1.44)*	0.13 *(0.72, 1.39)*	0.11 *(0.74, 1.36)*

Decision Making for Phase III based on Phase II Result



Phase II Results	***HR is large in AC***	***HR is small in AC***
HR is small in Dx+	*Selected*	*AC if HR in Dx+ and Dx- is similar*
		Co-primary(Dx+, AC) if HR in Dx+ << HR in Dx-
		Selected if HR is large in Dx-
HR is large in Dx+	*Stop*	*AC*

Selected Trial?

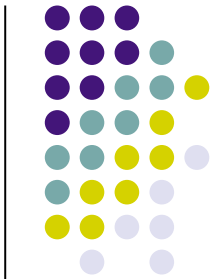


Issues to consider:

- Strength of available scientific evidence (what is the PHC assessment?)
- Timelines (selected vs stratified) *(operational considerations)*
  - Buzz
  - Number of sites.
- Regulatory considerations
  - Would we have to propose Stratified Phase III?
- What do we **not** learn by restricting POC to selected population and when may it be justifiable?

Not recommended unless PHC selected scenario is assumed.

Simultaneous PHC evaluation/Fast To Market (FTM) strategies



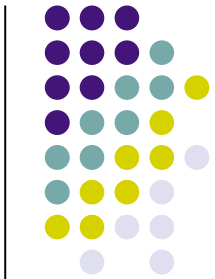
Maximizing Speed (and possibly letting cost get bigger): *How can we design a POC trial so to be able to initiate an All-Comers Phase III if warranted while preserving ability to evaluate Dx hypotheses?*

- In general, it is not possible to pursue simultaneous FTM strategy and PHC evaluation in the same trial, when Stratified scenario is pursued.

You don't want to be the horse that posts the highest speed in the middle of the course but you want to be the one that reaches the finish line first.

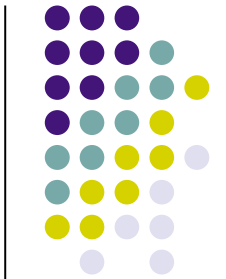
Tom Fleming, personal communications

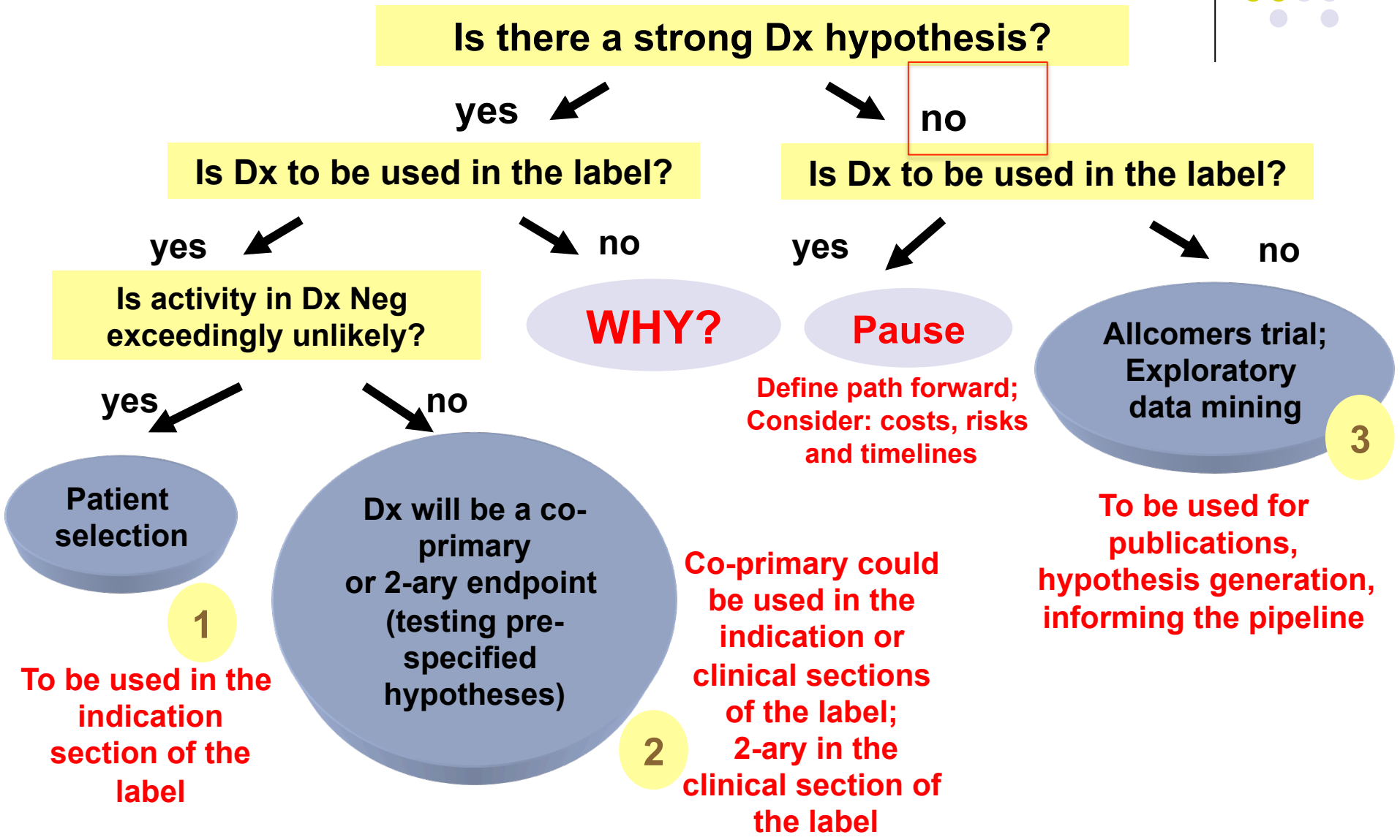
Adaptive Phase II Design considerations



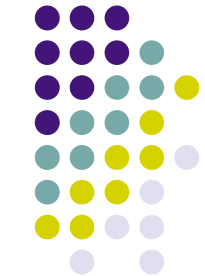
- One-arm response driven trials: modification of Simon's two stage design to include evaluation of Dx subsets
- Randomized time-to-event trials: dynamic patient allocation ratio adjustment based on early readout
  - In theory allows for arbitrary number of Dx subgroups and active treatments
  - Generally involves a Bayesian approach with continuous updating
  - Relies on the relevance of the early end point to the clinical outcome of interest
  - Increased operational complexity
  - In practice, sample size requirements or operational complexity can make complex adaptive designs intractable for the initial proof of concept evaluation.

Underlying Assumption



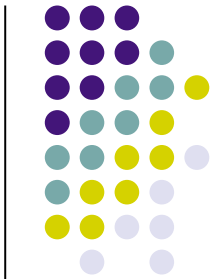


Defining Exploratory Analyses Scope



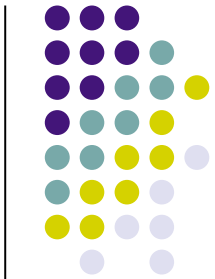
Key questions	Topics
What are the objectives of the study? How does the analysis meet the objectives?	Planning; Statistical criteria
How are the biomarkers measured? How accurate, precise, and reproducible are the measures?	Data generation Statistical criteria
How is the candidate biomarker derived?	Documentation; Statistical criteria; Cutoff selection; Complex predictor
How strong is the statistical evidence supporting a biomarker? How likely is it that a candidate biomarker will succeed in a prospective Phase III?	Communications; Statistical evidence; Validation strategy

I. Planning, Documentation & Communication



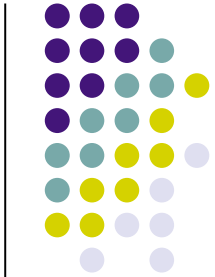
- **Biomarker analysis plan:**
  - Exploratory analysis ≠ lack of planning!
  - Require early communications to align key stakeholders
  - Prevent data-mining traps & aid resource planning
  - Flexibility to accommodate changes is expected
  - **Must have:** analysis objectives, **scope** and **ranked priorities** of analysis, definition of key variables
  - **Nice to have:** statistical methods, data format requirement, description of outputs, study background, biomarker rationale
- **Documentation:** detailed documentation & version control of programming code and all data manipulation steps
- **Communications and interpretation of results:** Context and caveats: convenience sample vs. trial population; differences in population, assays, etc.

II. Data Generation



- Sample requirement, handling, processing
  - Mandatory vs optional samples; archival vs fresh frozen; RNA processing etc.
- Sample & assay QC procedures
- Assay performance
- Data processing steps:
  - 'raw' assay data → analysis-ready biomarker datasets
  - including data acquisition, preprocessing, transformation and normalization
- Missing data handling and preventive measures:
  - out of range, assay failure, sample availability, etc.
- Comparison/concordance of multiple assays

Example: Retrofit Diagnostic



Objective: Are there biomarkers which predict benefit from approved drug Y?

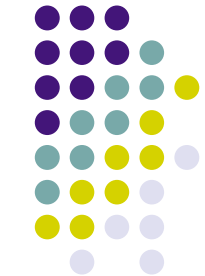
- Hypothesis generation for drug Y MOA and resistance
- To inform early development molecules targeting the same pathway
- Publications

Data: n=500 samples from a Phase III trial

Assays	#probes	# features	# relevant features
Affymetrix WT ExonST 1.0	~5M	~1.4M exons ~300,000 transcripts	278162 core exons; 17982 RefSeq genes
Affy U133plus2	~1M	54675 probesets	54675 probesets/genes
DiscovArray	~28000	14000 probesets	582 known human miRs
Illumina 1M HapMap	~2M beadtypes	~1M SNPs +38000 CN probes	~1M genotype calls; Unknown # Inferred CNVs

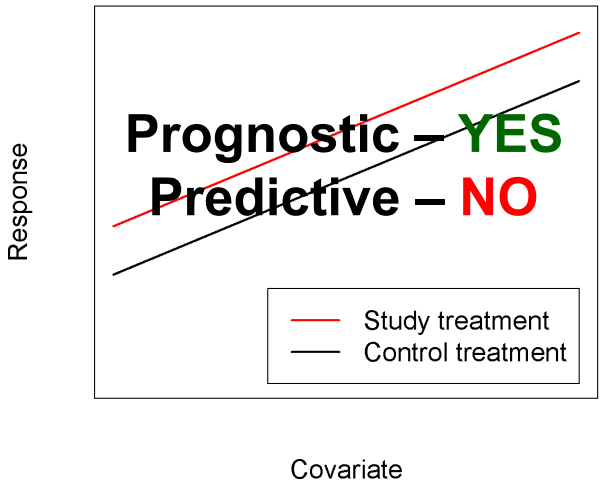
Very involved, platform-specific processes for data preprocessing, QC, annotation

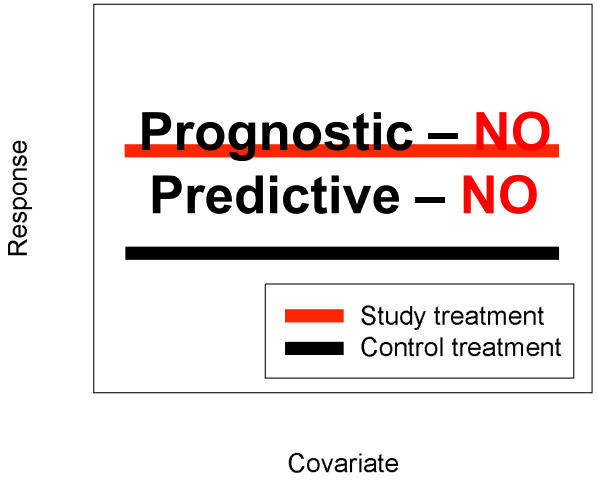
III. Fit-for-purpose Statistical Analysis & Criteria

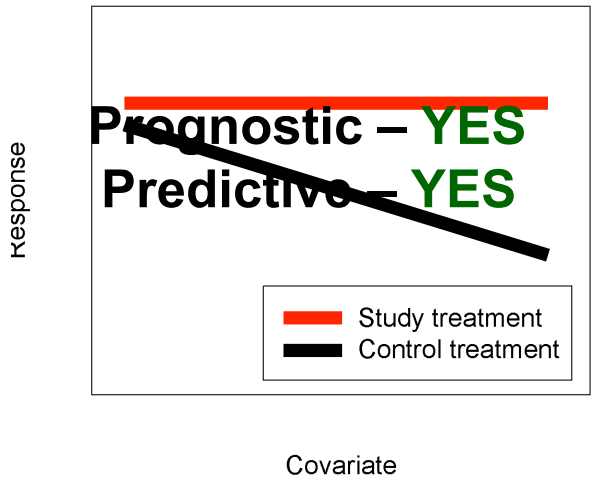


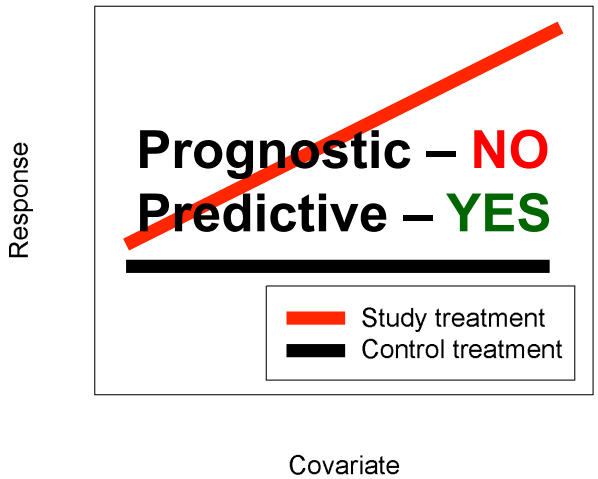
- Should be guided by project objectives, clinical significance and the TPP of the development program
- Pre-specify as much as possible
- Consider effect-size + statistical significance criteria:
  1. **Predictive** in the context of disease setting and study design:
     - Prognostic?
     - Treatment effect in Marker+ vs. Marker-
     - "*Treatment:Marker interaction*" → typically under-powered!
  2. Provide **independent predictive values:**
     - multivariate analysis adjusting for known prognostic/predictive factors
  3. **"Clinically significant" estimated effects**
     - Estimated effects for candidate biomarkers selected from discovery processes are typically optimistically biased

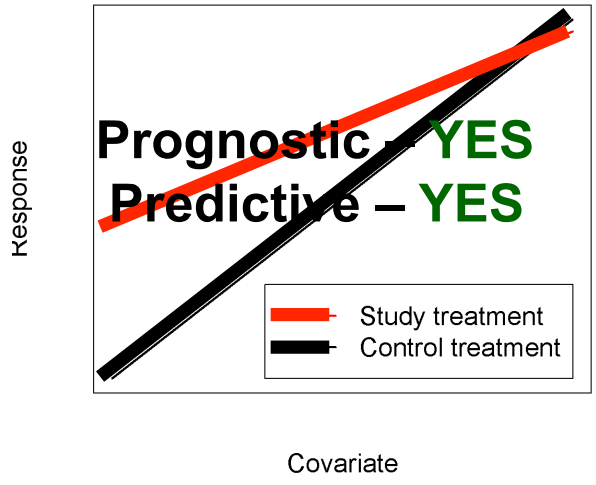
Prognostic?
Predictive?



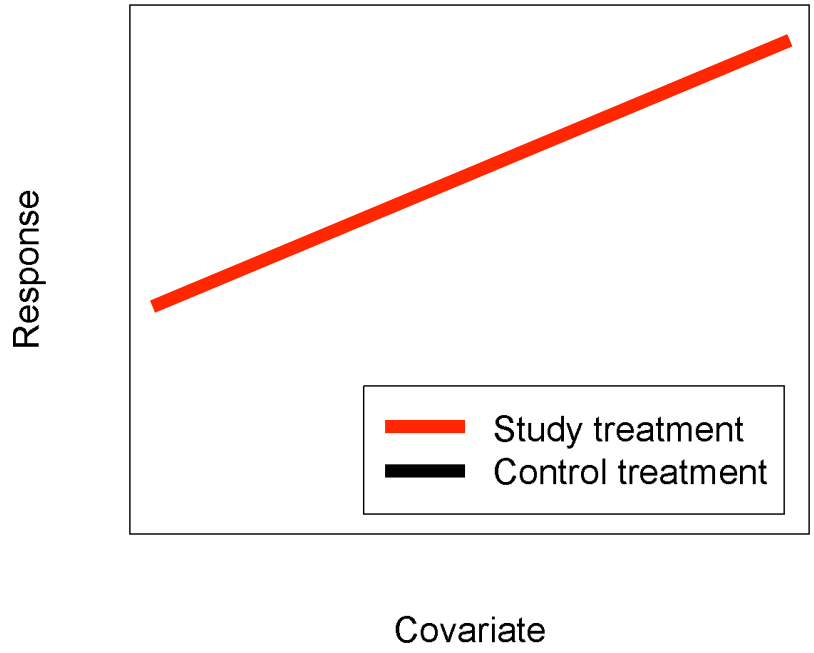






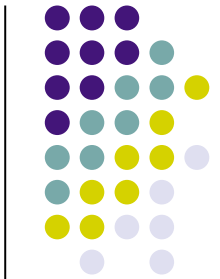


Prognostic?
Predictive?



Unable to tell, no control group
unless… assuming not prognostic (in control group)!

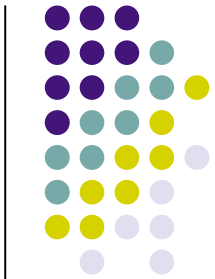
IV. Assessing & Interpreting Statistical Evidence



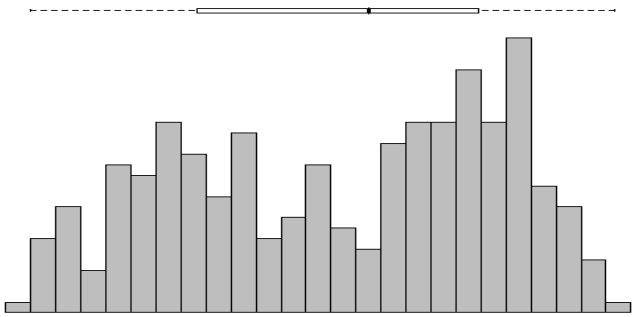
- Per-marker p-values: difficult to interpret
  - Expect 0.05*100 markers tested = 5 markers with $p \leq 0.05$ by random!

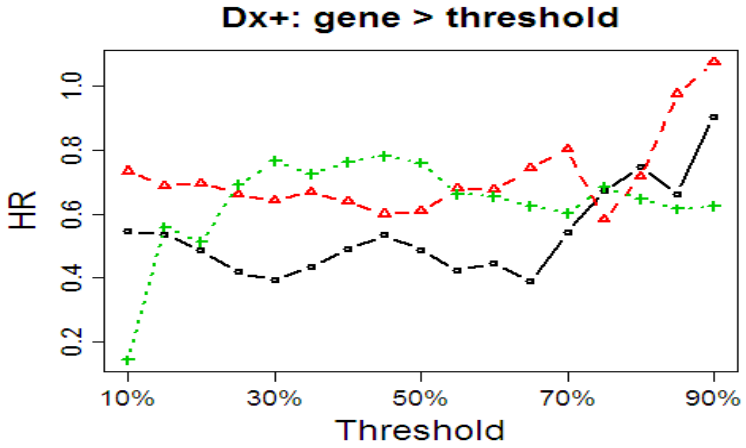
- Consider "overall type I errors":
  - Generalization of p-values, **false discovery rates**, etc.
  - Choice of statistical procedures & algorithms depends on the assumptions and utility
  - Preserve ranking of markers, but differ in nominal levels
  - Interpret accordingly
  - Main challenge: clearly define the family of hypotheses tested

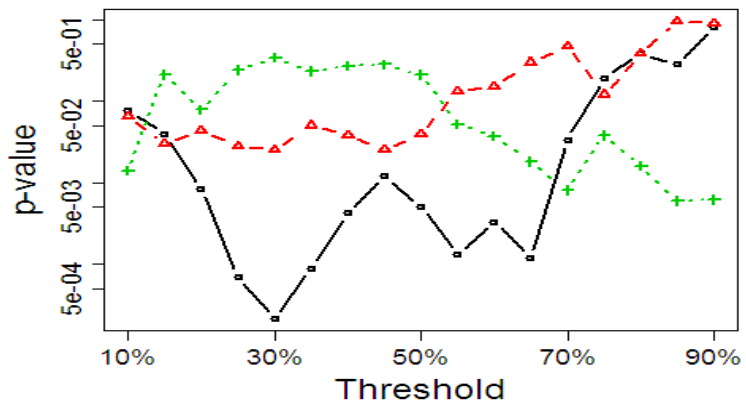
V. Cutoff Selection for Continuous Biomarkers



- Sparse literature and examples; no FDA guidance
- Ideal approach: biologically or clinically meaningful cutoff; multi-modal distribution due to underlying biology.
- Common approaches (data driven): good for exploration, but not good enough for implementation in subsequent confirmatory trials
  - Percentiles: use medians, quartiles, etc.
  - Optimization: e.g. find cutoffs that maximize
    - treatment effect differences in marker+/-;
    - marker+ subset that meets a pre-specified treatment effect.
  - multiplicity issues
- Useful to explore effect vs cutoff profiles



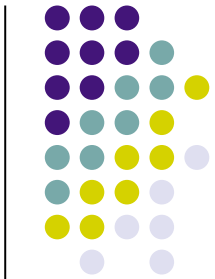




Some considerations:

- Biologically Meaningful Cutoffs
- Effect Maximization
- Subset size

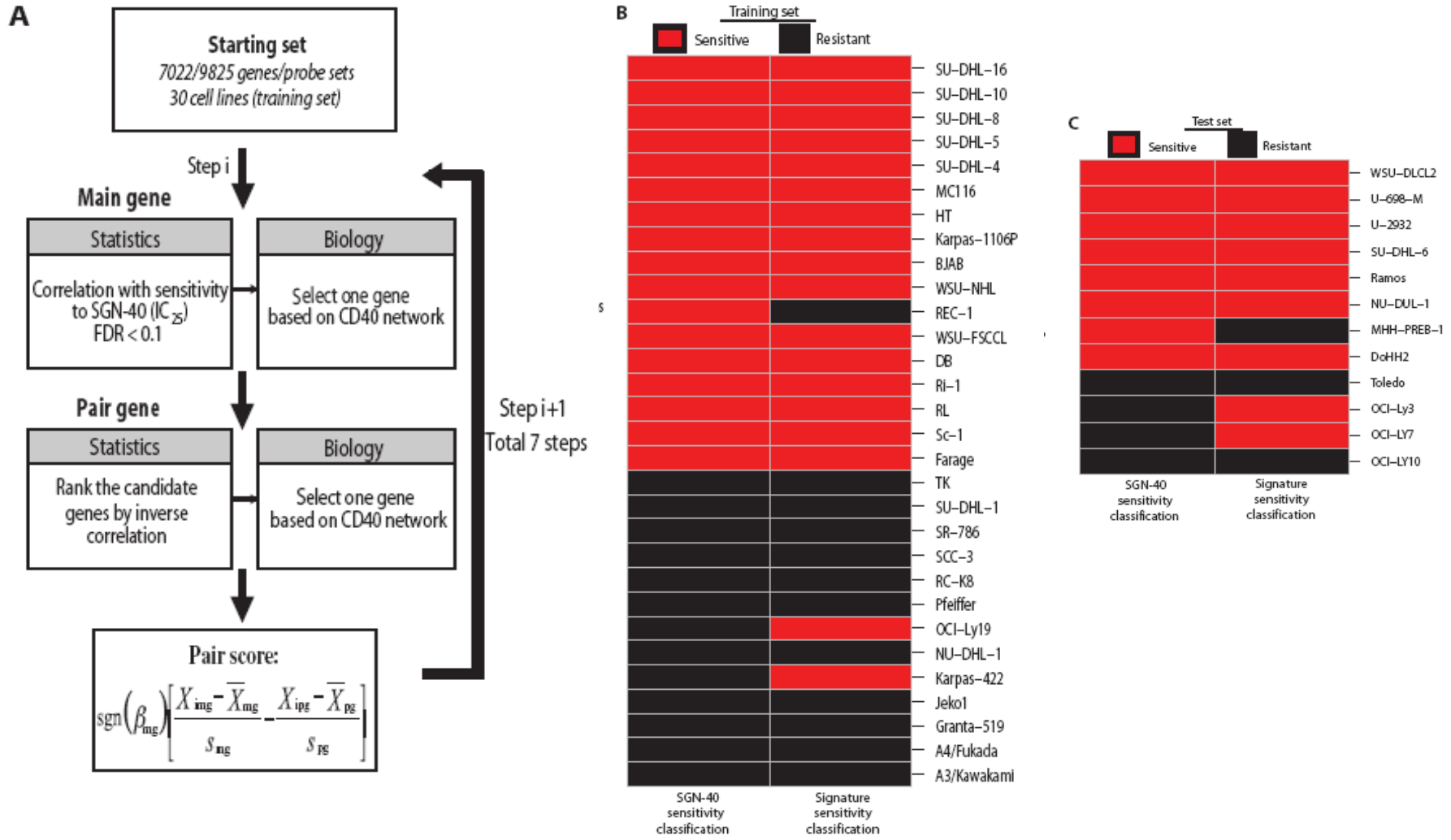
VI. Developing a Complex Predictor



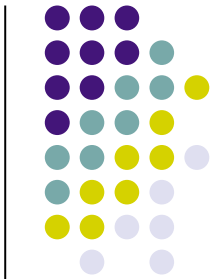
- Uncharted area for predictive application:
  - No HA-approved examples of complex predictor-based companion diagnostics predictive of a specific drug
  - Only related examples: FDA-cleared tests for prognosis of breast tumor recurrence: Oncotype DX (21-gene qPCR), MammaPrint (70-gene array)
- Multi-step processes involving many decisions & tuning:
  - Feature selection, model building, performance evaluation
  - Choice of options & algorithms, especially the performance metric, should be guided by the clinical objective and the nature of dataset
  - Major concern: high variance & overfitting due to large p (features), small n (sample size)
  - Validation strategy important
- Reality check: Set realistic expectation on the feasibility given the sample size, data quality and potential impact

Example: SGN-40

Burington et al. Sci Transl Med 16 March 2011. 3(74), p.74ra22

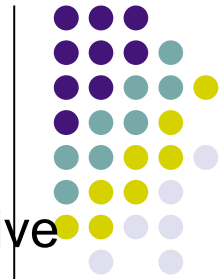


VII. Validation Strategy



- Validation & resource allocation strategy at the molecule program level to increase overall program PTS:
  - one bigger biomarker study with more power for discovery vs. two (or more) small studies, one for discovery and the other for independent validation.
- “External” validation with independent datasets
- “Internal” validation: cross-validation or bootstrapping
  - Very easy to make mistake by ignoring some aspects of predictor training or the discovery process inside the CV loop, leading to incorrect estimates (usually optimistically biased) of prediction errors.

Summary



- Developing drugs is hard (long, expensive and risky) – co-developing them with the Dx is long (slower speed) and expensive (higher cost) but more likely to succeed (PTS)
  - Making clinical decisions early based on pre-clinical data
  - Regulatory path is complex and not always clear at the present time
  - Operational issues are difficult
- Exploratory strategies for either retrofit or rescue require
  - Planning
  - Communication
  - Standardization
  - And most importantly, clear stopping rules

True strategic partnership of all functions is paramount to program success.