Promises and Pitfalls of High Dimensional Assays for Vaccine Signature Studies

Greg Finak, PhD
Vaccine and Infectious Disease Division
Fred Hutchinson Cancer Research Center
Seattle

Aging and Immunity III
Jan 11-13, 2016, Siena, Italiy,

The RGLab

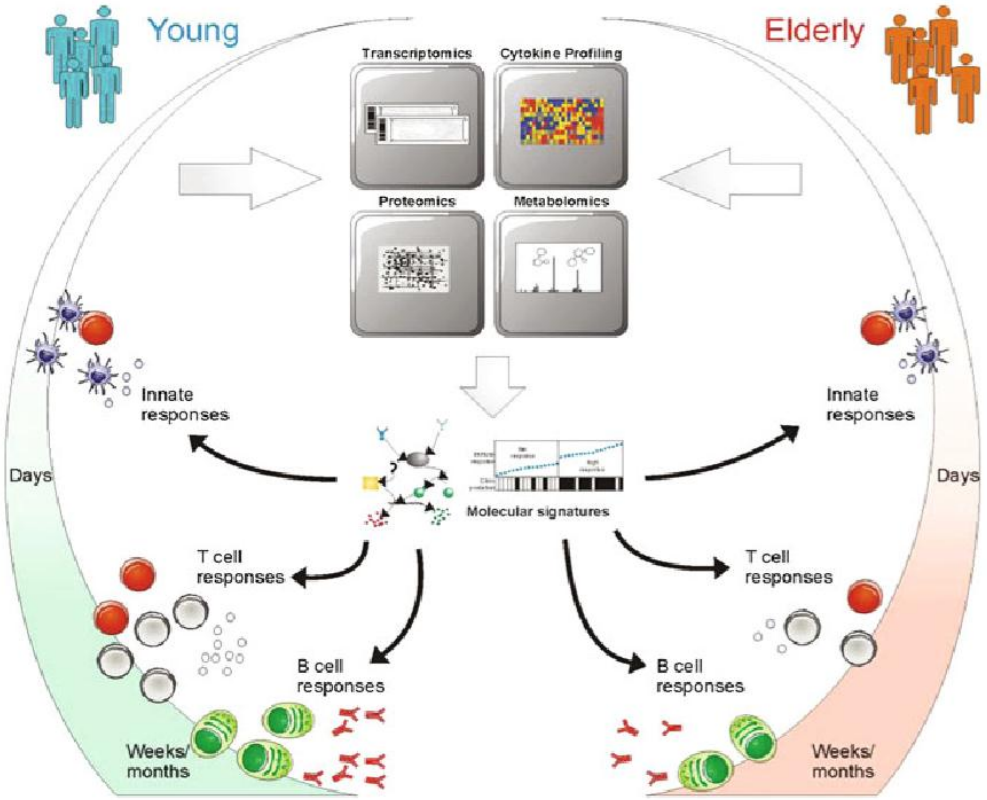
Statistical methods, computational tools, bioinformatics pipelines for high throughput immunological assays.

- Flow and mass cytometry (**flowCore, flowWorkspace, OpenCyto, flowViz, ggcyto, flowStats, flowClust, flowMerge**)
- Single-cell gene expression, RNASeq and qPCR (**MAST**)
- Data standardization and pipelines (**preprocessData, ImmuneSpaceR**)
- Modeling Ag-Specific T-cell responses (**MIMOSA, COMPASS**)

- Focus on reproducible research

http://github.com/RGLab http://rglab.org

Promises of High Throughput Systems Immunology

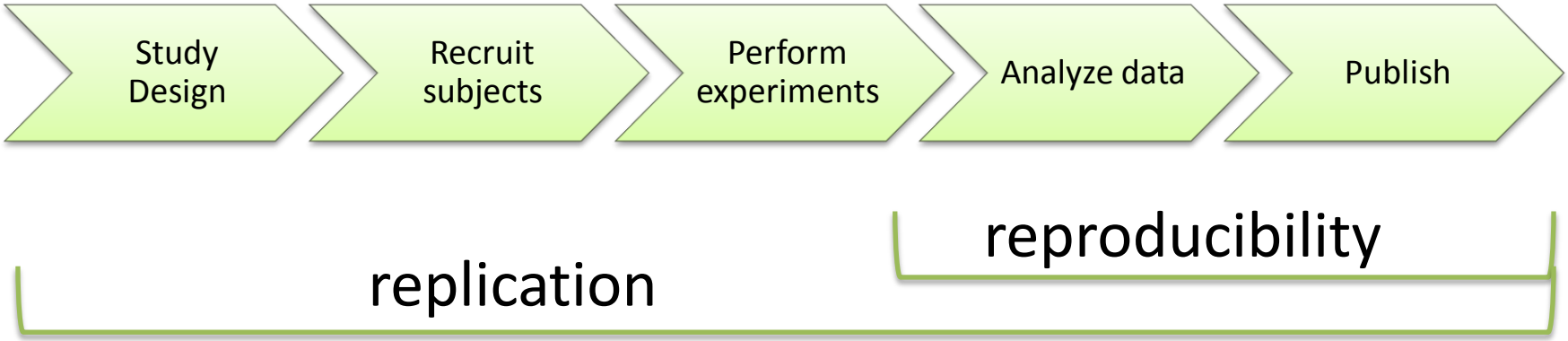


- Integrate diverse range of responses (early, intermediate, late).
- Integrate diverse biological assays (classical immunological assays and novel high throughput technologies).
- Hypothesis: early molecular signatures can predict late responses.
- Translation to new vaccine development.

Duraisingham, Rouphael, Cavanagh, Nakaya, Goronzy, Pulendran. **Systems Biology of Vaccination in the Elderly**. *Curr. Topics in Microbiol. and Immunol.* (2012)

Increasing focus on reproducibility and replicability

- Follow-on to Duke University cancer trial scandal
- Increased attention over past 5 years by scientific journals, media.



- Why does replicability and reproducibility matter?
- Increasing use of high dimensional assays.
- Journals tightening requirements for sharing data, code.
- Impacts ability to *successfully translate* findings into drug or vaccine development.

Reproducibility and Replicability in High Throughput Biology

- We want scientific findings to stand up to replication.
- How can we improve how we use high throughput data?

Common Causes of Irreproducibility

Irreproducible studies tends to fail early

- **Experimental Design** – Underpowered studies.
- **Data generation** – Batch effects, assay reproducibility, assay characteristics (sensitivity, specificity, dynamic range, etc.)
- **Data analysis** – no statistical analysis plan, ad-hoc analysis.
- **Data management** – Data annotation / mislabeling, version control
- *Need to approach data analysis more formally even in pre-clinical and discovery studies.*
- Communication, awareness, training of scientific staff (post-docs, graduate students, technical staff).

Challenges in Irreproducible Research Nature 2012

Underpowered Studies – Why?

- **Limited resources, many comparisons**
  - Small sample size, comparisons are underpowered.
  - Attempting to answer too many questions – loss of power.

Strategy

- *Engage a statistical collaborator.*
- Rank questions by order of importance.
- Design and power study to ensure primary questions are answered *unambiguously*.
- Then plan for exploratory analyses.

Power Analysis Informs Feasibility

- Can I answer the questions I'm interested given available resources (samples, funds, time)?
- Assay operating characteristics vary from lab to lab.
  - Preferably use preliminary data from **the same lab that will run the assays**.
  - Signal to noise: assay reproducibility is critical.
- Should take **study design** and **statistical analysis plan** into account.
  - Complex study designs – power often assessed by simulations .

A Statistical Analysis Plan Mitigates Against “Fishing Expeditions”

- Detailed outline of how data will be analyzed
  - Defines **primary hypotheses**.
  - Defines **secondary / exploratory objectives**
  - Defines **endpoints** and **statistical procedures.**
- Mitigates against “fishing expeditions”.
  - Control Family-Wise Error Rate
- The analysis plan facilitates power analysis.
- Helps identify oversights before resources are spent.
- **Ensure everyone is on the same page.**

Mitigating Confounding and Batch Effects

Batch effects impact the best designed studies.

Common Causes

- Timing of sample collection and preparation.
- Consistency of protocol adherence by lab, assay reproducibility
- e.g.
  - **RNASeq, single-cell RNASeq** – library preparation date, instrument, minor changes in protocol, etc..
  - **Flow cytometry** – staining panel reagents, date of assay, gating scheme, analyst.
- Batches are unavoidable in larger studies. **Anticipate and mitigate.**
- Aim to balance treatment groups across batches.
- The person performing the experiment should communicate with analyst or statistician **as they plan the experiment.**

Batch Effects in a BCG Vaccination Study

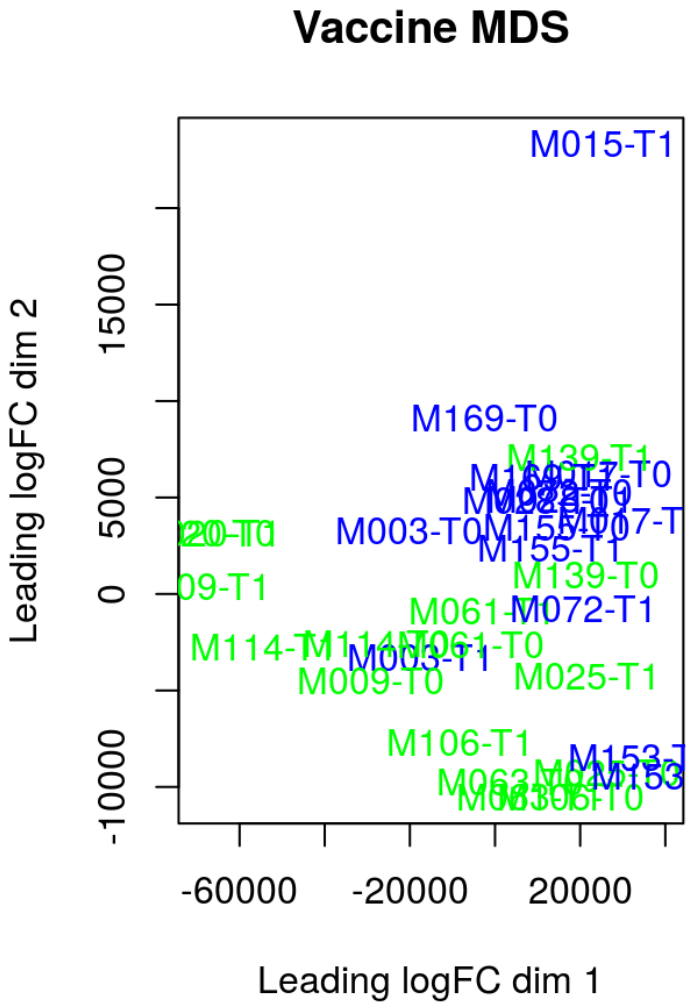
N = 16 subjects

2 replicates per subject

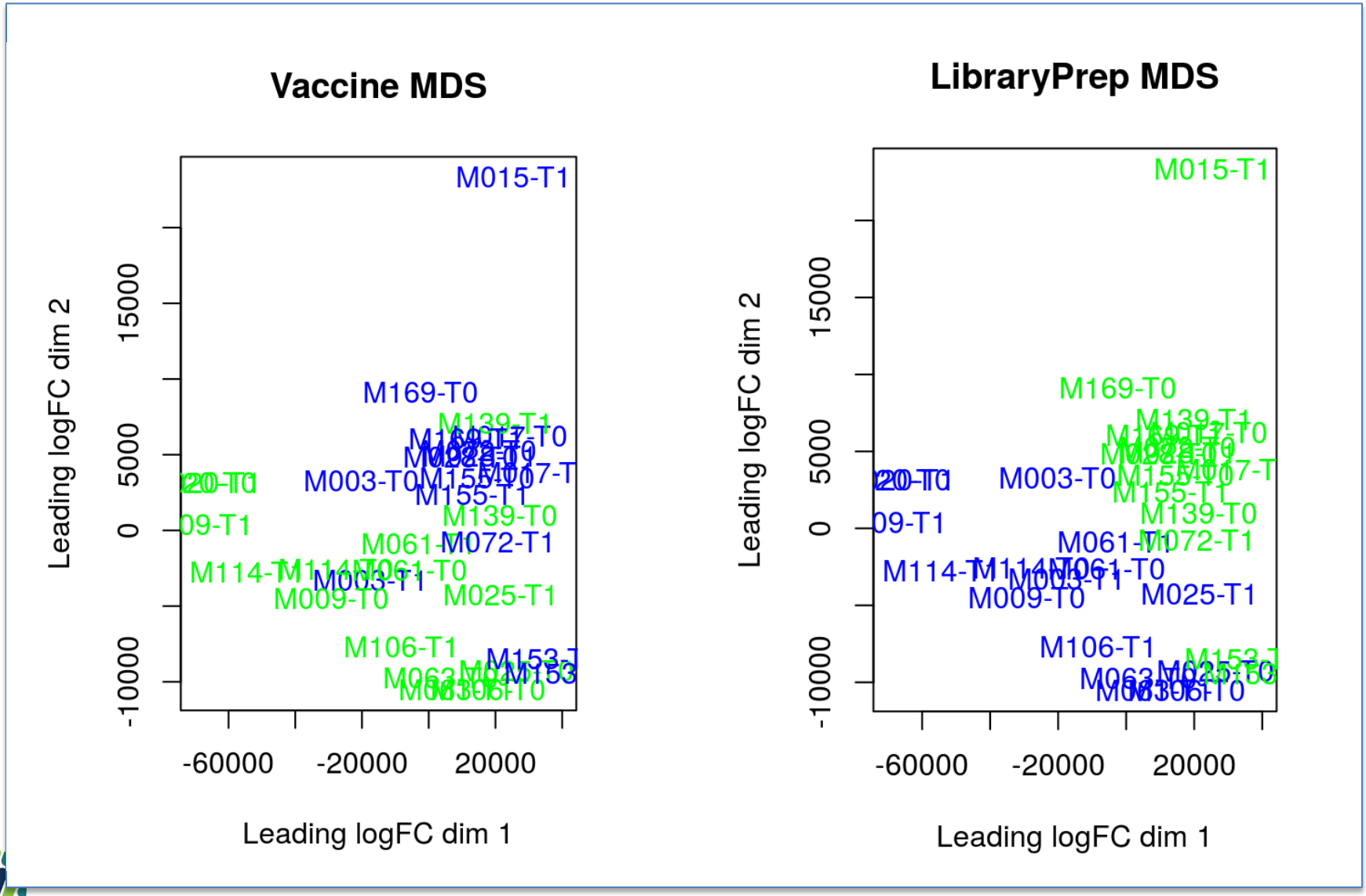
3 way design

Primary objective: Vx-specific changes pre-vs. post vaccine

Vx	TB test	Baseline	Post-Vx
BCG	TB+	2	2
	TB-	2	2
Placebo	TB-	2	2
	TB+	2	2



Vaccine Effect Confounded With Library Prep Date.

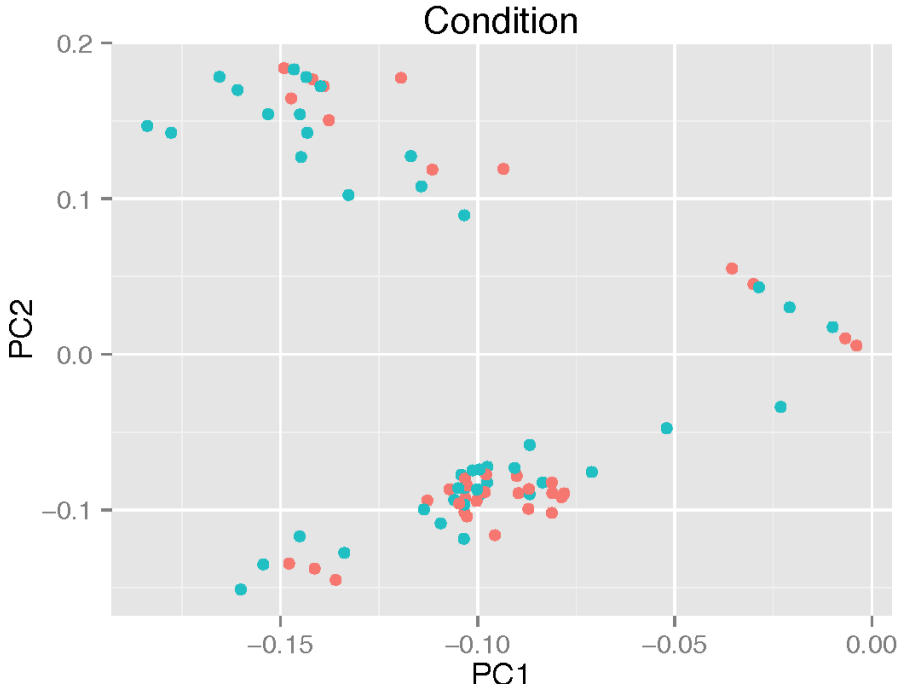


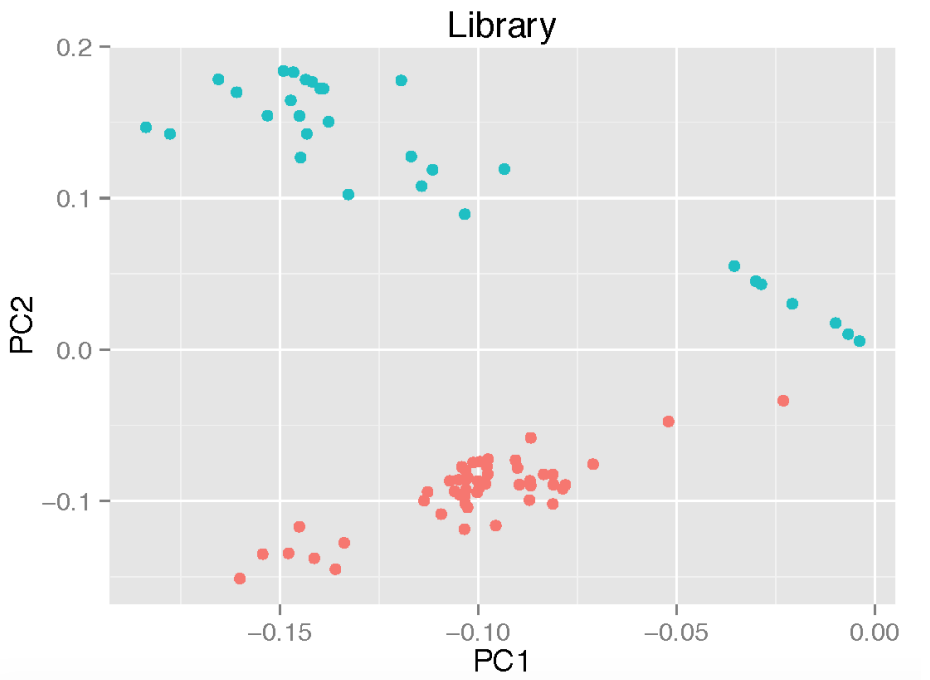
Outcome

- Primary question of interest are vaccine-specific differences pre vs. post vaccine.
- **But**
  - **Data not comparable due to batch effects.**
  - Need to analyze treatment groups separately.
- TB+ / TB - was secondary, more covariates to estimate.
  - Loss of power.
- Limited findings of interest.

Another Example

- RNASeq of human samples from two conditions
- Samples run by two different post-docs on two different sequencers, 12 months apart.





- How much should we trust gene signatures derived from such a data set?
- Should $ be spent validating them?

Very Important to Assess Assay Reproducibility

- When is an assay sufficiently reproducible for biomarker discovery studies?
- Validation: reproducibility, sensitivity, specificity, accuracy, and precision.
- At a minimum: assess discriminative power
  - Does the assay detect what you are trying to measure in an experiment (e.g. discriminate vaccinees and placebos)?
  - Does it discriminate between baseline and post-vaccine?

RV144 Thai Trial Primary Results

The NEW ENGLAND JOURNAL of MEDICINE

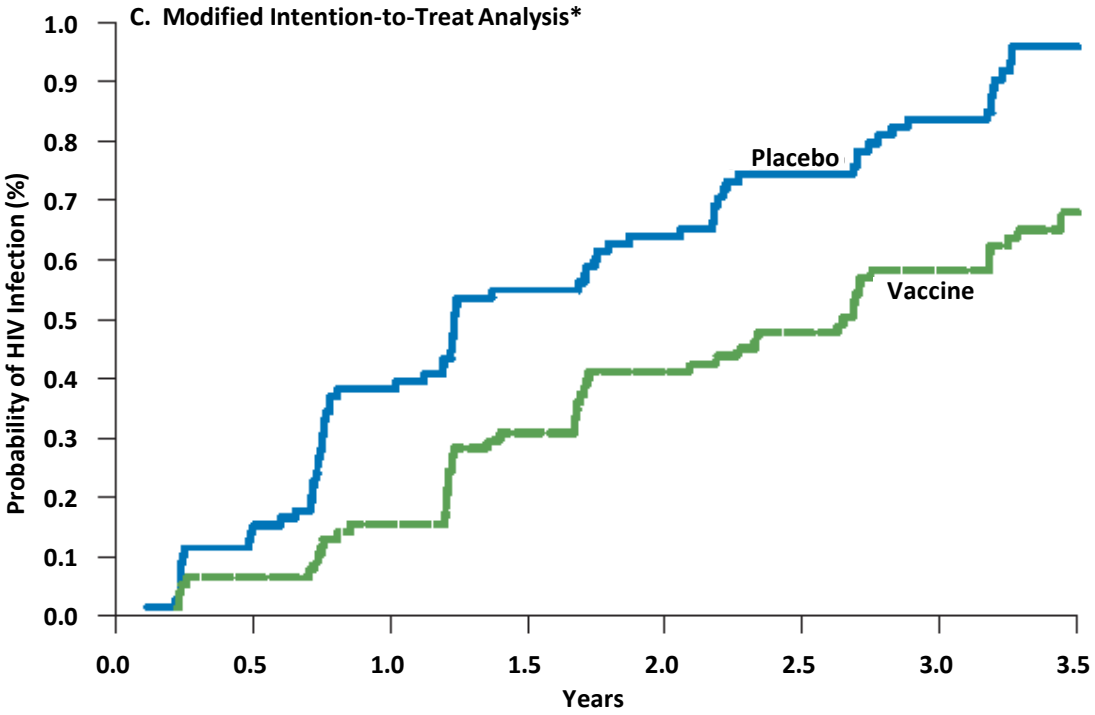
ESTABLISHED IN 1812 DECEMBER 3, 2009 VOL. 361 NO. 23

Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand

Supachai Rerks-Ngarm, M.D., Punnee Pitisuttithum, M.D., D.T.M.H., Sorachai Nitayaphan, M.D., Ph.D., Jaranit Kaewkungwal, Ph.D., Joseph Chiu, M.D., Robert Paris, M.D., Nakorn Premsri, M.D., Chawetsan Namwat, M.D., Mark de Souza, Ph.D., Elizabeth Adams, M.D., Michael Benenson, M.D., Sanjay Gurunathan, M.D., Jim Tartaglia, Ph.D., John G. McNeil, M.D., Donald P. Francis, M.D., D.Sc., Donald Stablein, Ph.D., Deborah L. Birx, M.D., Supamit Chunsuttiwat, M.D., Chirasak Khamboonruang, M.D., Prasert Thongcharoen, M.D., Ph.D., Merlin L. Robb, M.D., Nelson L. Michael, M.D., Ph.D., Prayura Kunasol, M.D., and Jerome H. Kim, M.D., for the MOPH–TAVEG Investigators*

Impetus for the Correlates Study:

Evidence for Partial Vaccine Efficacy



*N=16,395 assessed; 51 Vaccine, 74 Placebo HIV-1 infected
Estimated VE = 31% [95% CI 1–51%], p=0.04

Objective: To carry out an immune correlates analysis to begin to identify how the vaccine might work

RV144 Correlates of Risk Results

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 APRIL 5, 2012 VOL. 366 NO. 14

Immune-Correlates Analysis of an HIV-1 Vaccine Efficacy Trial

Barton F. Haynes, M.D., Peter B. Gilbert, Ph.D., M. Juliana McElrath, M.D., Ph.D., Susan Zolla-Pazner, Ph.D., Georgia D. Tomaras, Ph.D., S. Munir Alam, Ph.D., David T. Evans, Ph.D., David C. Montefiori, Ph.D., Chitraporn Karnasuta, Ph.D., Ruengpueng Sutthent, M.D., Ph.D., Hua-Xin Liao, M.D., Ph.D., Anthony L. DeVico, Ph.D., George K. Lewis, Ph.D., Constance Williams, B.S., Abraham Pinter, Ph.D., Youyi Fong, Ph.D., Holly Janes, Ph.D., Allan DeCamp, M.S., Yunda Huang, Ph.D., Mangala Rao, Ph.D., Erik Billings, Ph.D., Nicos Karasavvas, Ph.D., Merlin L. Robb, M.D., Viseth Ngauy, M.D., Mark S. de Souza, Ph.D., Robert Paris, M.D., Guido Ferrari, M.D., Robert T. Bailer, Ph.D., Kelly A. Soderberg, Ph.D., Charla Andrews, Sc.M., Phillip W. Berman, Ph.D., Nicole Frahm, Ph.D., Stephen C. De Rosa, M.D., Michael D. Alpert, Ph.D., Nicole L. Yates, Ph.D., Xiaoying Shen, Ph.D., Richard A. Koup, M.D., Punnee Pitisuttithum, M.D., D.T.M.H., Jaranit Kaewkungwal, Ph.D., Sorachai Nitayaphan, M.D., Ph.D., Supachai Rerks-Ngarm, M.D., Nelson L. Michael, M.D., Ph.D., and Jerome H. Kim, M.D.

What the Correlates Study Assessed

- The analysis sought to discover **Correlates of Risk:** Immune response variables that predict whether vaccinees become HIV-1 infected
- Generate hypotheses about **surrogates of protection** for validation in future research.

Study Design Planned for a Test and Validation Study

- **Pilot immunogenicity studies (2010-2011)**
  - Open process inviting immunology labs to perform assays on sample-sets from HIV uninfected RV144 participants
  - Standardized comparative analyses of all candidate assays,
    - Down-select the best performing assays
    - Cover immunological space
    - Optimize the immune variables to study as correlates
- **Case-control study (2011)**
  - Assess the selected immune variables as correlates of infection risk

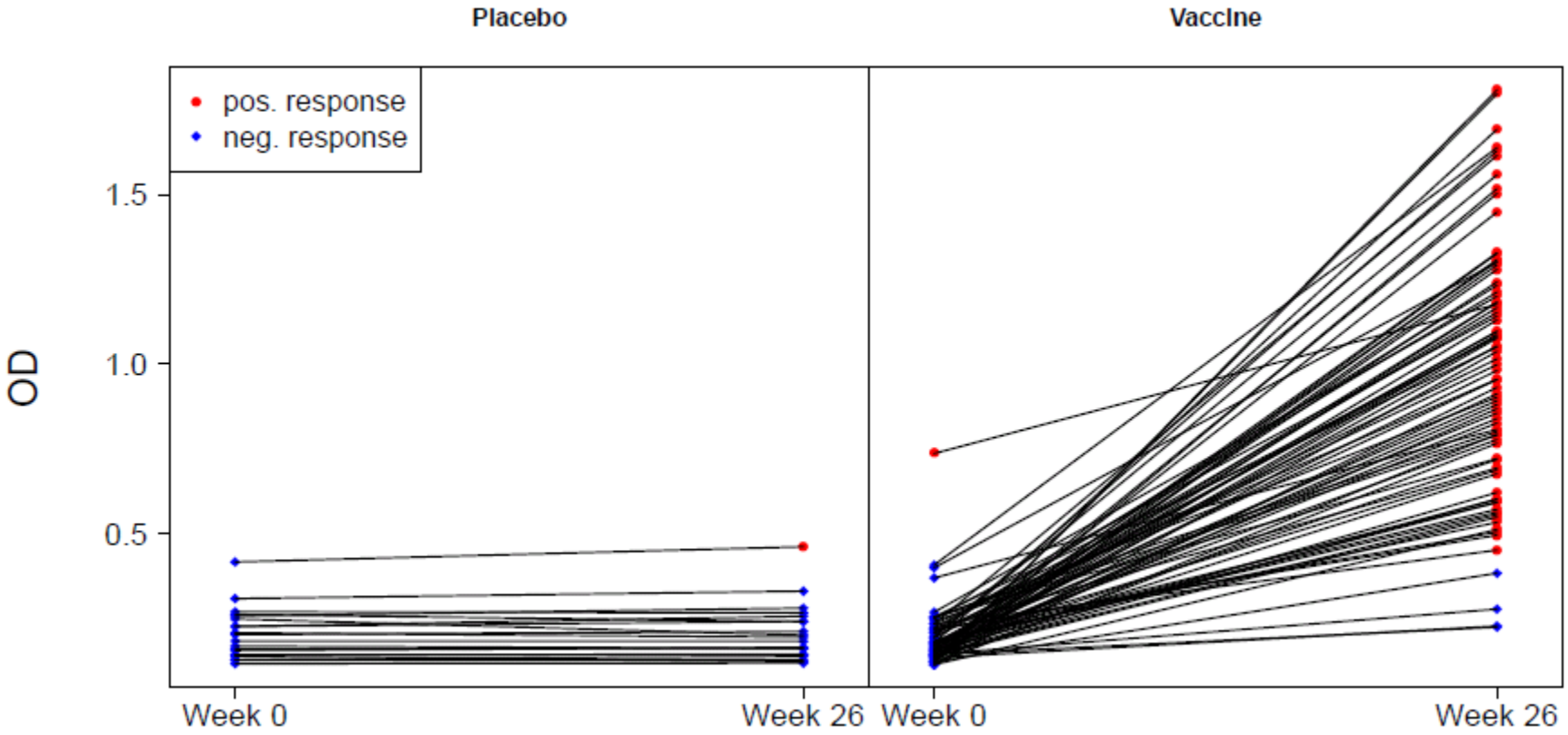
Pilot Immunogenicity Studies

- **Objective:** Comparative analysis of all candidate assays. Evidence for advancing assays to case-control study.
- ***Prototype* pilot data-set: 100 uninfected RV144 subjects**
  - 80 vaccine: 20 placebo balanced over men and women, pre-immunization and peak immunogenicity samples (Weeks 0, 26)
  - **Assess vaccine-induced responses**
    - Compare readouts Week 26 vs. Week 0, for vaccinees
    - Compare readouts Week 26 vaccine vs placebo

Pilot Studies: Criteria for Advancing Assays to the Case-Control Study

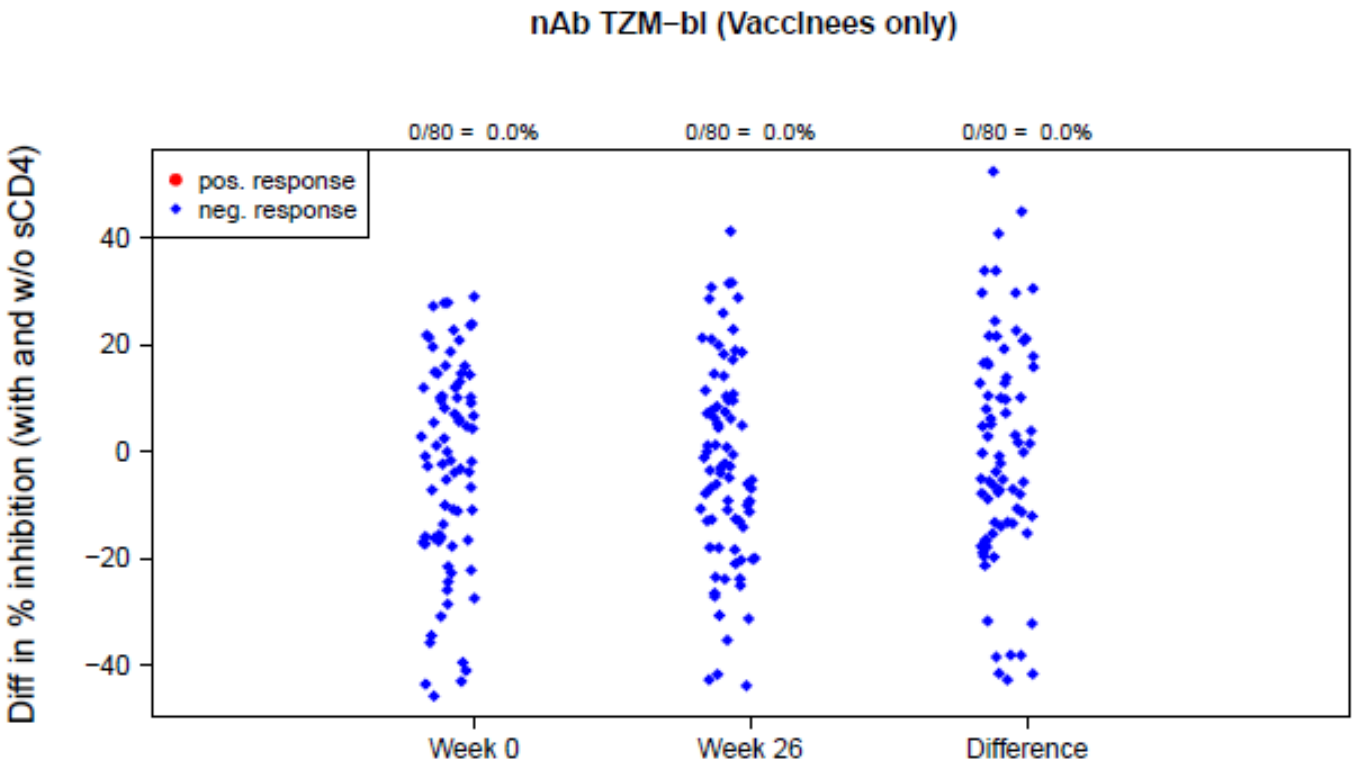
Criterion	
1. Represents a niche in immunological space (not highly correlated with other assays)	✓
2. Low false positive rate (judged in placebo recipients and Week 0 responses of vaccinees)	✓
3. Vaccine-induced responses with broad variability	✓
4. Relatively low noise (e.g., high reproducibility on replicate samples)	✓
5. Relatively low specimen volume requirement	✓
6. Previously supported as a correlate of infection in the North American VaxGen trial of AIDSVAX	✓

Example of well performing assay. Pilot Data: gp70-V1V2 Binding Antibodies (ELISA)



Screen Out Assays Failing Criteria 1 or 2 (High False Positive Rate or Lack of Vaccine-Induced Responses)

- Typical example of a screened-out assay: nAB TZM-bl assay (data on 80 vaccinees)



Summary of Outcome of Pilot Studies

- Assays from 47 proposals evaluated from 20 immunology labs
- Assays were “scored” on a scale of 1 to 5 by the leadership committee.
- 17 assay types passed pilot study criteria, and performed on the case-control samples
- **6 “best performing” immune variables** covering 6 immunological classes were selected for the **primary analysis**
- The **remaining 152 qualifying immune variables** were assessed in **secondary and exploratory analyses**

Case-Control Analysis: Primary and Exploratory Analysis

- **Primary Analysis:** 6 priority immune response variables
- **Secondary Analysis:** All other immune response variables that passed pilot study criteria for use
  - Type I error rates controlled separately
  - This division maximizes statistical power for the priority immune variables while allowing a broader exploratory analysis

Primary Analysis Accounted for Study Design and Potential Confounders

- Two regression models that accounted for the sampling design
  - Logistic regression full maximum likelihood*
  - Cox proportional hazards partial likelihood§ (yielded essentially the same results)
- Confounding control
  - Adjust for gender, baseline behavioral risk (low, medium, high)
  - Evaluate the 6 primary variables together in multivariate models, and as single variables

* Breslow and Holubkov (1997, *Biometrika*)

§ Borgan et al. estimator II (2000, *Lifetime Data Analysis*)

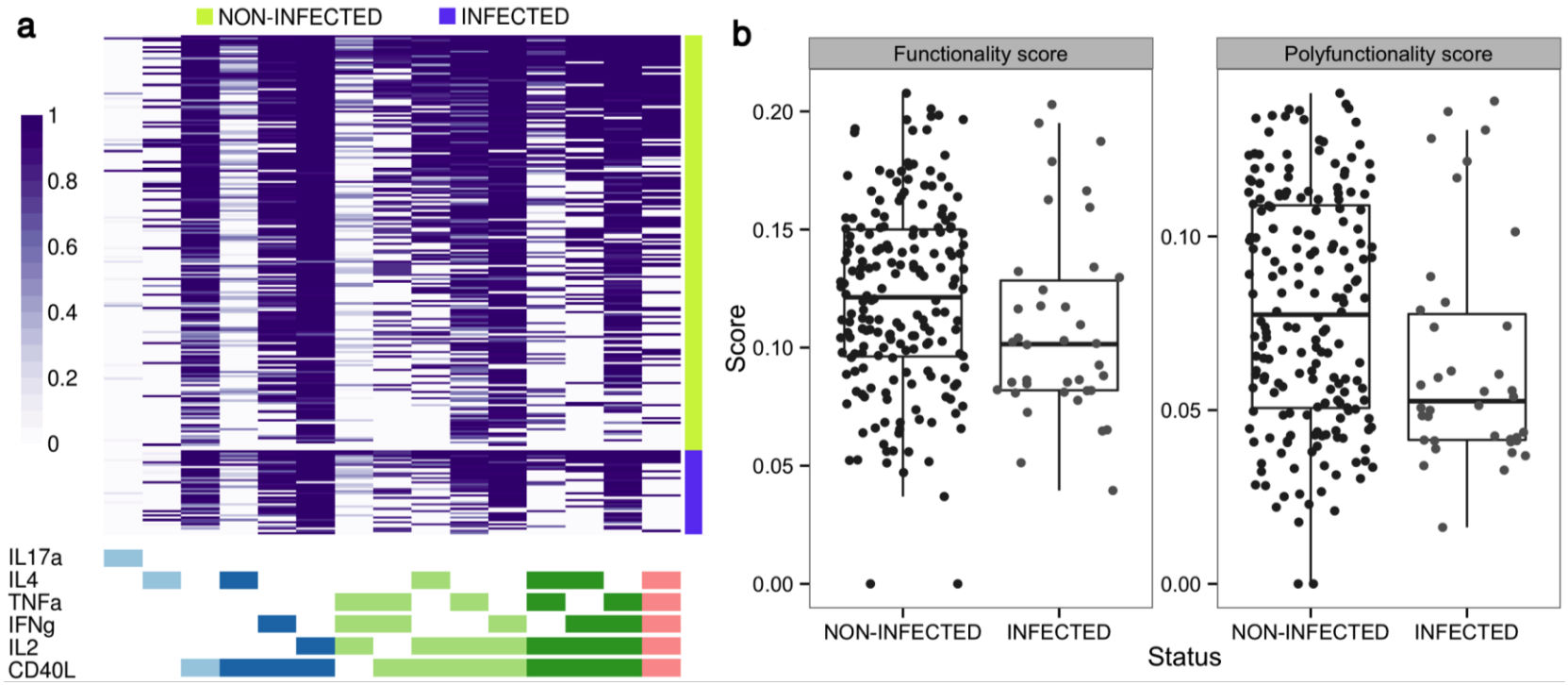
Hypothesis Generation can Allow Greater False Positives

- **Goal is not to miss potential correlates**
- **Corrections for multiple tests**
  - False discovery rate (FDR) correction
    - q-values < 0.2 deemed to provide evidence for a correlate (this means any detected correlate can have up to 20% chance of being a false positive)
  - FDR correction prioritized over Holm-Bonferroni correction because the study is hypothesis generating, and hence was designed to be sensitive for not missing correlates
- **Caution is needed in drawing conclusions that non-significant variables are unimportant for protection**

Two primary correlates identified

Variable	Relative risk	P-value	Q-value
IgA Binding to Envelope Panel	**1.54**	**0.027**	**0.08**
IgG Avidity A244 gp120	0.81	0.37	0.56
ADCC AE.HIV-1 Infected CD4 Cells	0.92	0.68	0.68
Tier 1 Neutralizing Antibodies	1.37	0.22	0.45
IgG Binding to gp70-V1V2	**0.57**	**0.015**	**0.08**
CD4+ T Cell Intracellular Cytokines	1.09	0.61	0.68

Examples of secondary correlate High-dimensional ICS analysis



(OR: 0.62; p=0.01)

(OR: 0.57; p= 0.0049)

Lin et al. *Nature Biotechnology* (2015). doi:10.1038/nbt.3187.

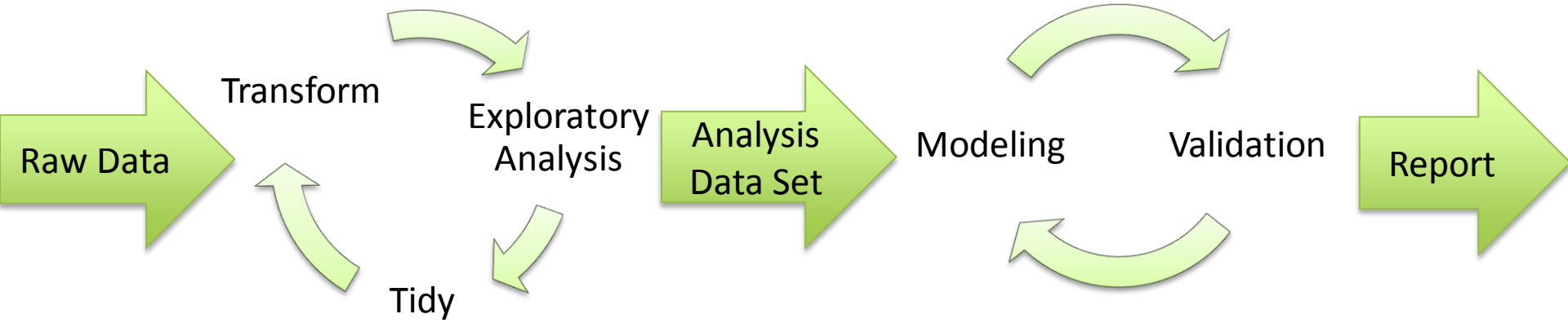
Primary Analysis Maintained Data Integrity for Case-Control

- The statistical analysis plan (SAP) was finalized before conducting the primary analysis
- Each immune variable definition finalized before unblinding the data
  - The primary data-set was set in stone and then the analysis was carried out
- Primary results validated by an independent statistical team

Correlates Conclusions

- Pilot studies play an important role in immune correlates analyses
  - Eliminate noisy assays and reduce overlap
  - Increase power
  - Improve analysis integrity
- Secondary/Exploratory analyses are important too
  - Don't want to be too stringent
  - Important for high dimensional assays
  - Revisit biomarker definition
- RV144 is a successful model that has been reproduced in several other studies (HVTN505, Dengue, Malaria, etc).
- On a smaller scale, plan for the testing / validation data set paradigm and formalize the data analysis process.

Finally: Reproducible Data Analysis



Data analysis is iterative.

- Large and complex data sets
- Many possible decisions.
- Transformation from *raw data* to an *analysis data set*, and final report needs to be documented.
- Manuscript "Methods" sections are not sufficient to capture the complexity.
- Need alternative approaches.

Modern Tools for Reproducible Research (on a smaller scale)

Avoid "File Salad" - ZIP Files by e-mail

- State of the art computational tools, standardized pipelines for high throughput biological assays (RNASeq, expression arrays, flow cytometry, multiplex qPCR and many others).

and RMarkdown

- Literate programming framework for reproducible reporting using R. Integrates analysis code and reports.

- Version control for code, reports, and data sets, collaborative environment.

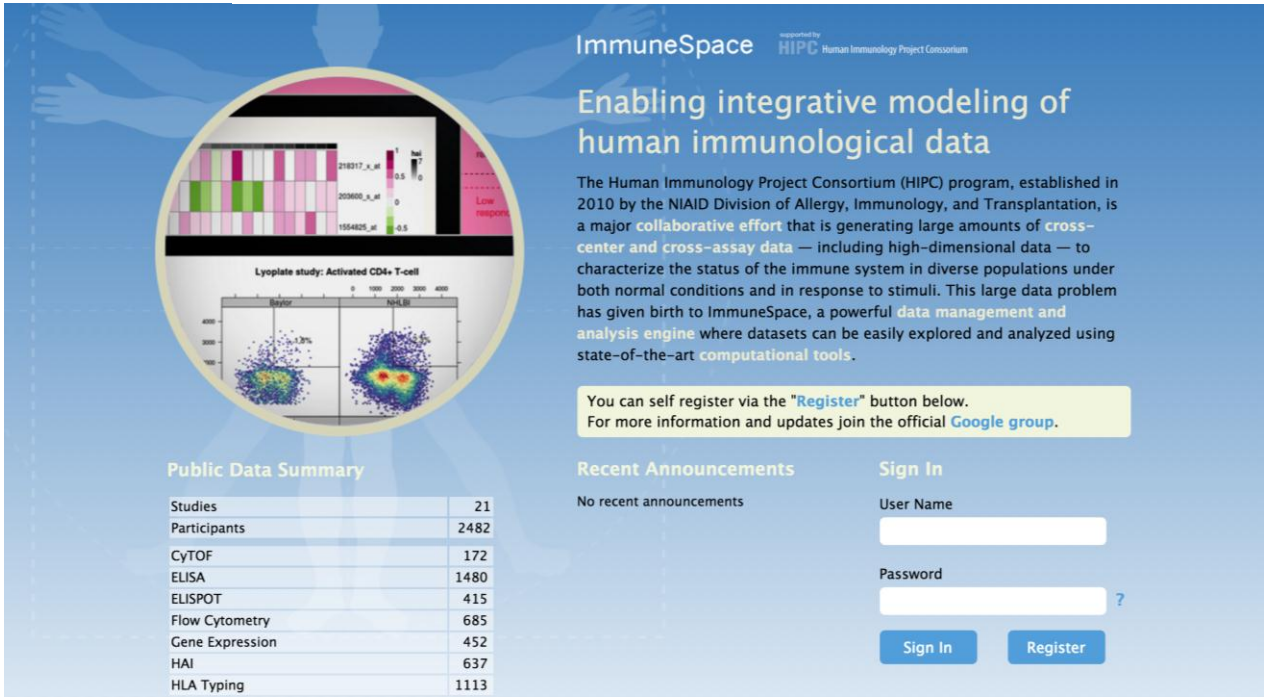
- Assign a DOI to data sets, software, reports for referencing in papers and and sharing with the public.

HIPC ImmuneSpace

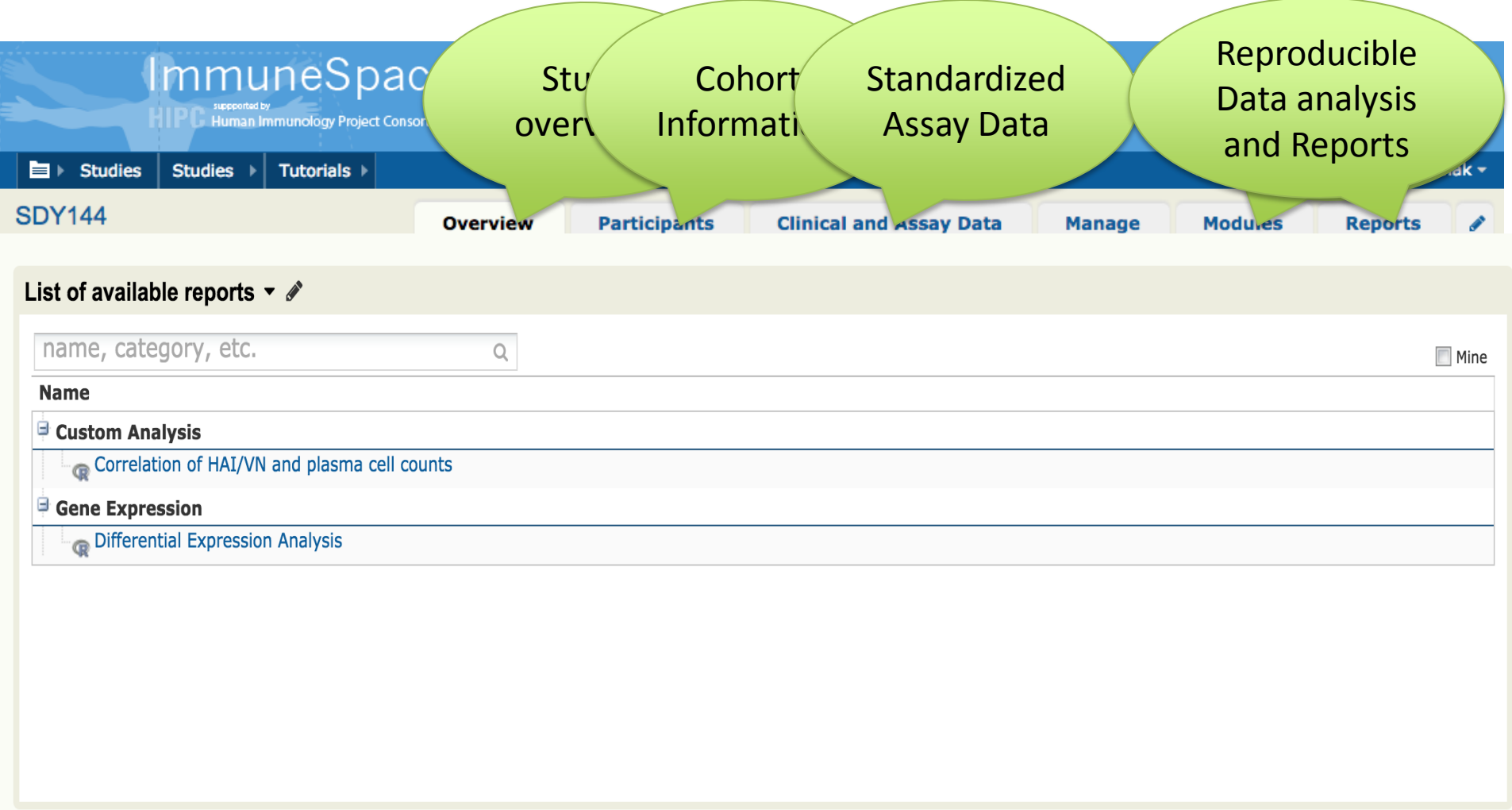
http://www.immunespace.org

Standardized and curated data base of immunological data sets from NIAID funded studies.

HIPC centers → ImmPort → ImmuneSpace



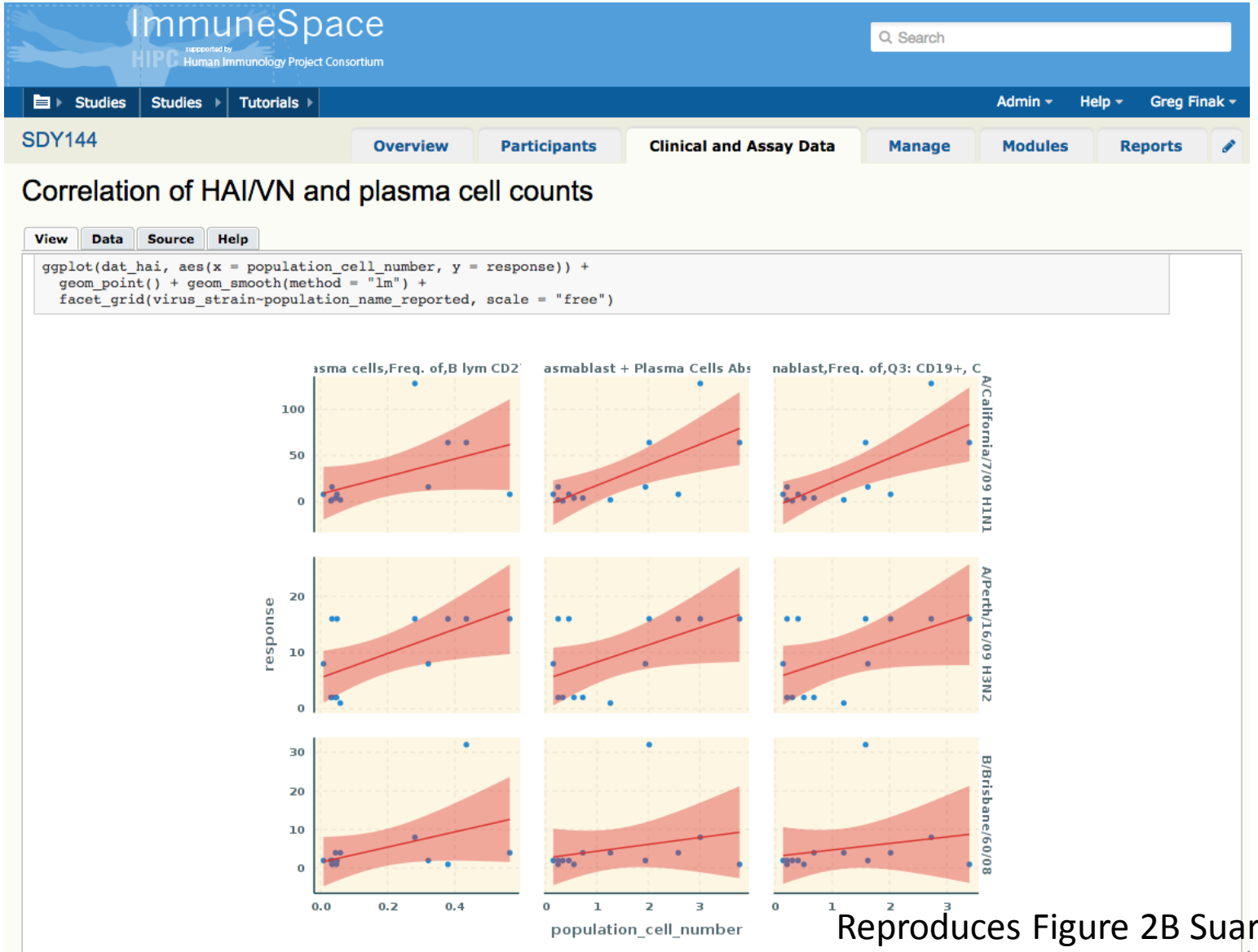
- Publicly accessible
- 26 studies
- 2787 participants
- Multiple assay types
- Demographics and metadata
- Searchable
- Standardized and computable



Study description, cohort information, standardized assay data, publication reference, and reports accessible for each study.

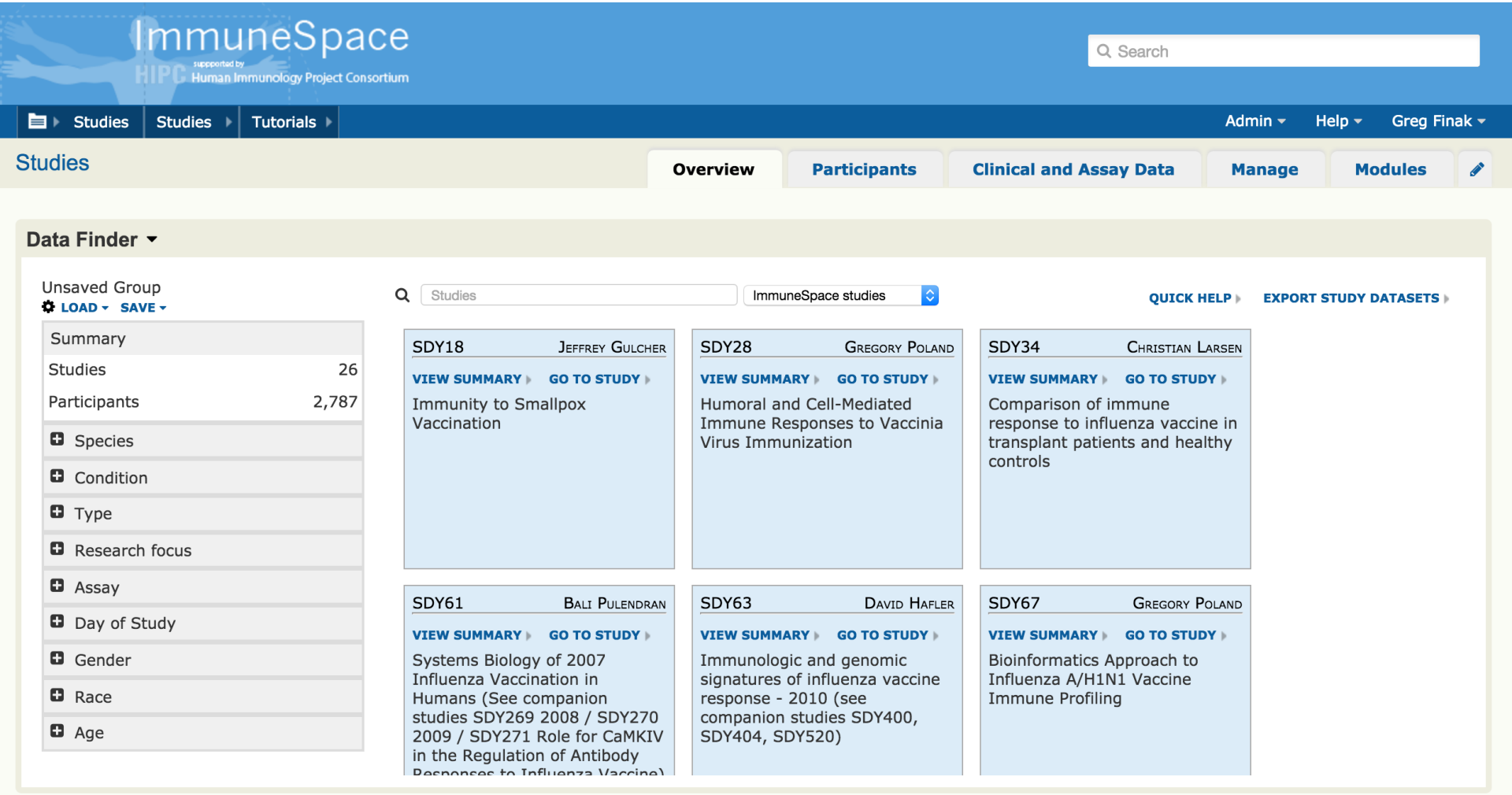
Suarez *et. al* J. Infect Dis. Jul (2014)

Rmarkdown Reports Reproduce Published Figures using “live” Data



Reproduces Figure 2B Suarez *et. al.*

Searchable Data Sets



Datasets are standardized, searchable, selectable, and downloadable for local analysis.

Leveraging Public Data to Fill Gaps

- Collaboration with HIPC Centers and Kleinstein, Tsang, Shen-Orr, Khatri, Gottardo labs and others
- **Four cohort meta-analysis evaluating flu vaccine responses in young (<35) vs. older (>60) individuals.**
- **500 subjects across 5 consecutive flu vaccine seasons from 2008 to 2009.**
- Goal to identify predictive signatures of vaccine response in young and old subjects.

Summary

- In Immunology we are fortunate – most studies are longitudinal, larger, data are more complete.
- Approach data analysis more formally to improve reproducibility..
  - Formally plan statistical analysis
  - Ensure studies are powered appropriately
  - Evaluate assay reproducibility
  - Plan for independent validation data
- Encourage scientific staff to engage early and communicate with with statisticians and data analysts.
- Leverage public data
- ImmuneSpace provides standardized immunological data.
  - Useful for validation or discovery data sets.
  - Meta-analysis and re-analysis.
  - Use public data to increasing sample size for planned studies.

Acknowledgements

Our Group

- Raphael Gottardo, Mike Jiang, Andrew McDavid, Carl Murie, Masanao Yajima, Phu Van, Leo Dashevskiy, Renan Sauterand

HIPC Centers and Collaborators

HVTN

- Steve De Rosa, Nicole Frahm, Julie McElrath, RV144 Leadership Team

Funding

- CAVD (Gates Foundation), HVTN (NIAID), Human Immunology Project Consortium (NIAID), NIH/NIBIB R01EB008400