Module 8: Evaluating Immune Correlates of Protection

Instructors: Peter Gilbert, Ivan Chan, Paul T. Edlefsen, Ying Huang

Talk 6: Introduction to Sieve Analysis of Pathogen Sequences

Summer Institute in Statistics and Modeling in Infectious Diseases

University of Washington, Department of Biostatistics

July 15-17, 2013

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER A LIFE OF SCIENCE

Outline Talk 6

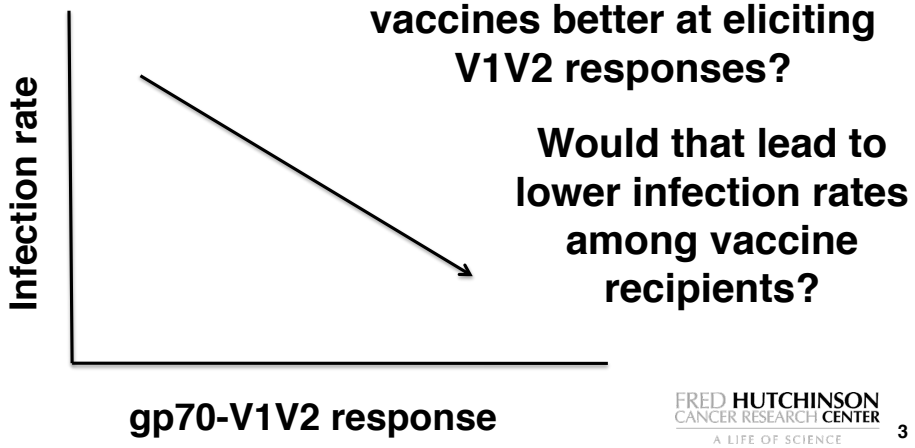
1. Introduction: Concepts and definitions of sieve effects / sieve analysis
   - Vaccine efficacy versus particular pathogen strains
   - Sieve effects and other effects
   - Some immunological considerations
   - Some sieve analysis results from HIV-1 vaccine efficacy trials

2. Some statistical approaches to sieve analysis
   - Binary endpoint (Infected yes/no)
     - Discrete pathogen types: Categorical data analysis
     - Continuous types: Distance-to-insert comparisons

3. Assumptions required for interpretation as per-exposure vaccine efficacy

RV144 Correlates Result

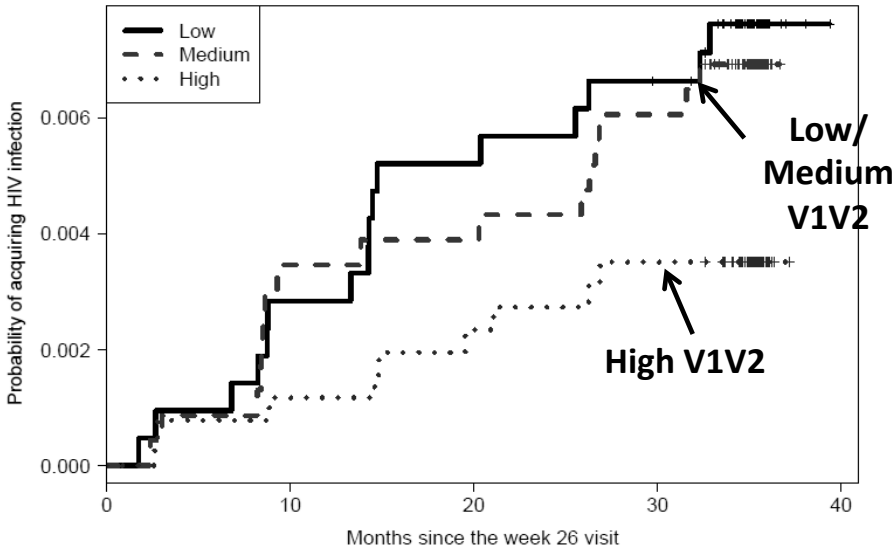
- Vaccine recipients with higher gp70-V1V2 responses tended to have lower rates of infection.



Correlation ≠ Causation

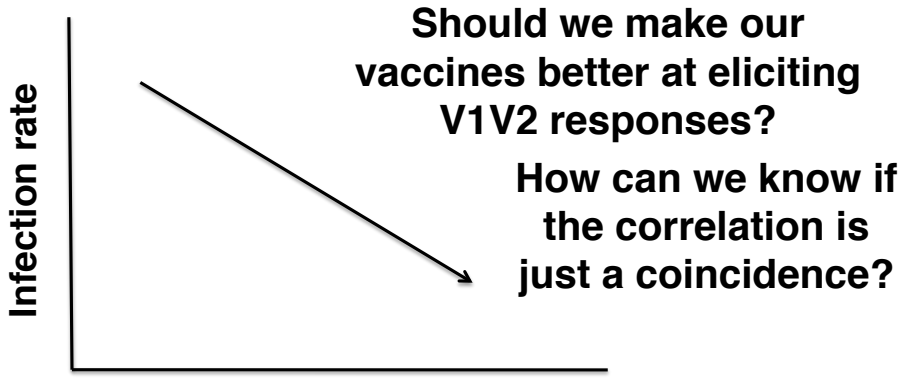
- Locations with higher sales of ice cream tend to have higher rates of drowning.

Cumulative Infection Rates With V1V2-gp70 Scaffold Assay



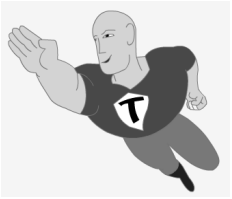
Should we make our vaccines better at eliciting V1V2 responses?

Randomized Controlled Trials (RCTs)

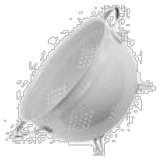


- In an RCT, treatments (vaccine or placebo) are randomly assigned.
- If you compare across treatment groups, the only explanation for a difference is the vaccine.

Towards a CoP and/or a Mechanistic CoP



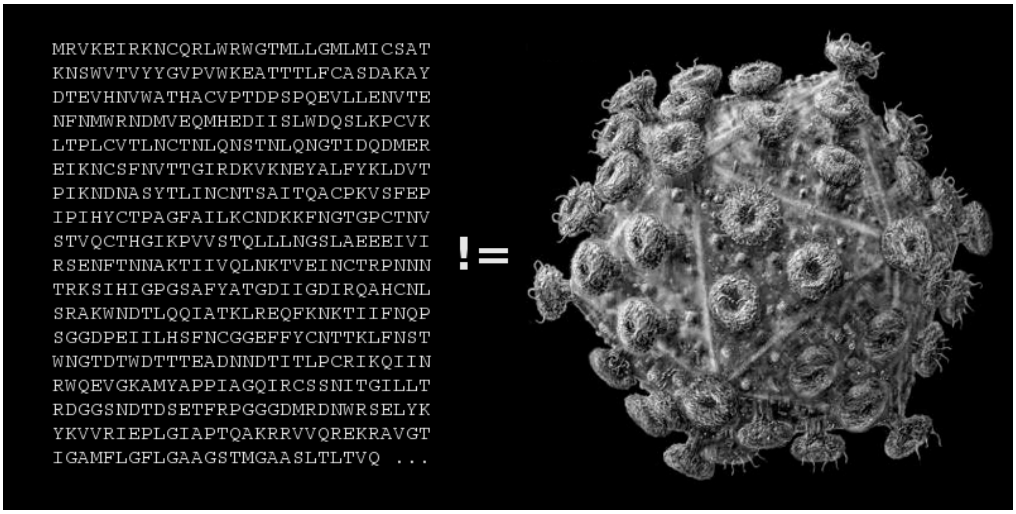
- The correlates so far are not CoPs.
  - The comparison is among vaccine recipients, not across randomized treatment arms.
- Could we randomly assign anti-V1V2 antibodies?
  - Maybe. There's other statistical ways, too.
  - We'll need to wait until future RCTs.
- Idea: use RV144 placebo vs. vaccine recipients
  - to address hypotheses implied by a causal correlate. Like: "Anti-V1V2 antibodies in vaccine recipients (partially) protected them."



Sieve Analysis

- Vaccination should induce an immune response
  that targets circulating HIV
  (at least the HIV that's similar to the vaccine HIV)
- Idea: investigate the sequence data
- If we see evidence for a difference in the sequences of viruses infecting vaccinees versus placebo recipients,
  - it must be due to the vaccine.
    - (It's a randomized trial!)
- If we see a difference in the sequences of V1V2,
  - then it supports the hypothesis of anti-V1V2 antibodies selectively filtering HIV.

Sequence data is an abstraction



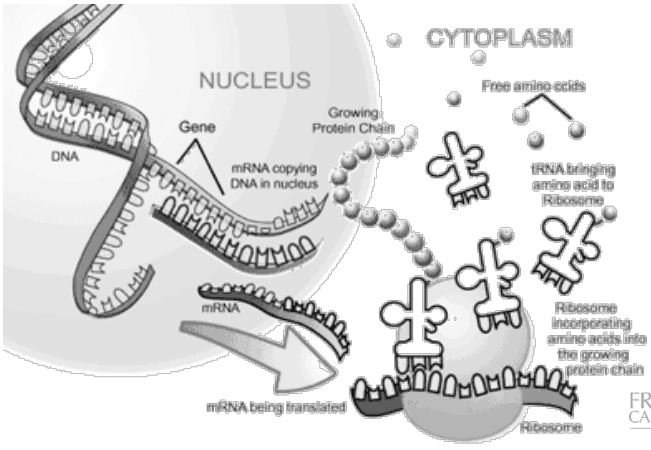
… but a useful one …

Three kinds of biosequences

- DNA: sequences of 4 nucleic acids: ACGT
- RNA: sequences of 4 nucleic acids: ACGU
- Protein: sequences of 20 amino acids

transcription

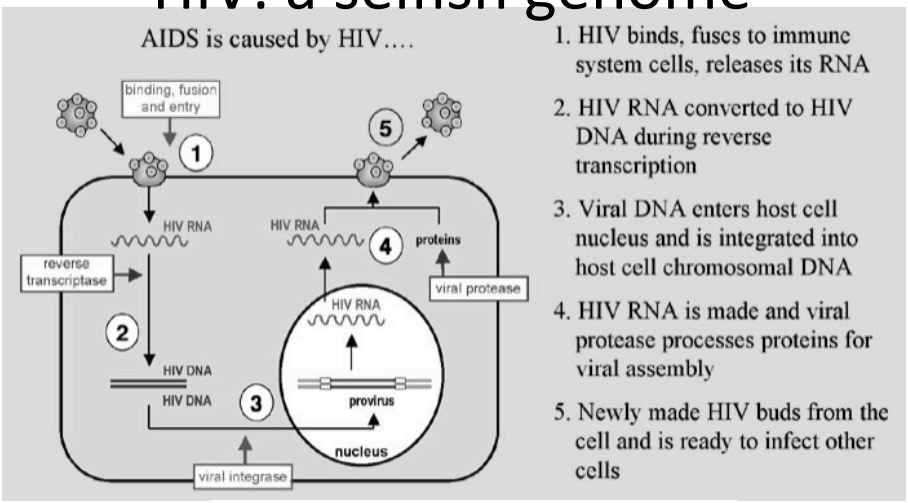
translation

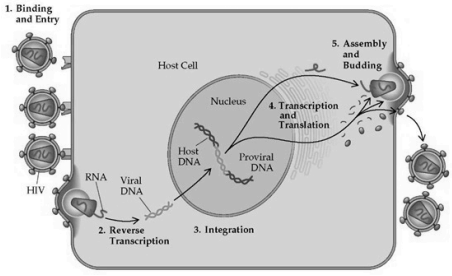


The human genome

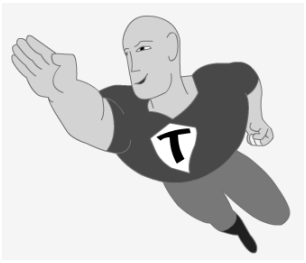
- DNA: 23 chromosomes, ~3 billion pairs of nucleic acids
- RNA: ~135,000 unique transcripts
- Protein: ~25,000 different protein products

HIV: a selfish genome

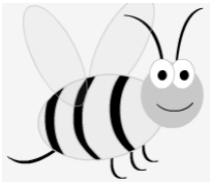




Biosequences and adaptive immunity

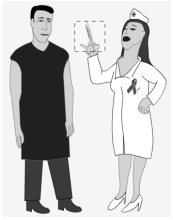


- Two major components: B cells and T cells
- Cells constantly report status: T cells monitor.
  - Fragments of protein sequences are brought to cell surface
  - Cells are destroyed when they report "bad" fragments
  - T cells adapt to learn what "bad" looks like
- B cells create antibodies,
  - which recognize proteins & flag them for destruction.
  - B cells also adapt to recognize "bad" proteins.

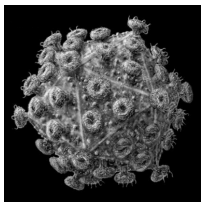


- Vaccines can train an immune system to recognize HIV earlier
  - and more effectively.

HIV Vaccines

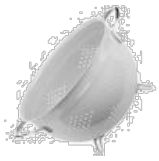


- Contain fragments of the HIV genome
  - Either proteins or DNA that will be expressed as proteins
- Recipients produce HIV-targeting T&B cells
  - No need to wait: destroy HIV before it destroys the immune system
  - Like when you become immune to a flu after infection or vax.
- What sequence(s) to include in the vaccine?
  - Want to create immune responses that protect people



Variation in the HIV genome

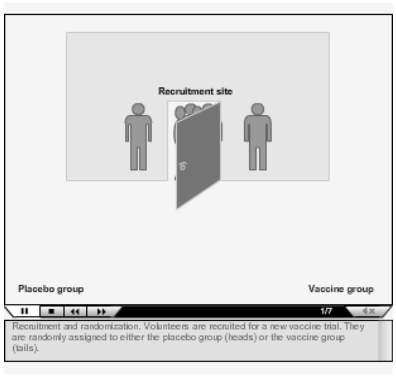
- The HIV genome is highly variable
  - due in part to a sloppy reverse-transcriptase.
- HIV evolves rapidly to evade immune systems
  - Variation and selection: Darwin's essentials for evolution
- Some adaptations hinder HIV
- Ideal vaccine: immune system targets Achilles' heel



Back to Sieve Analysis

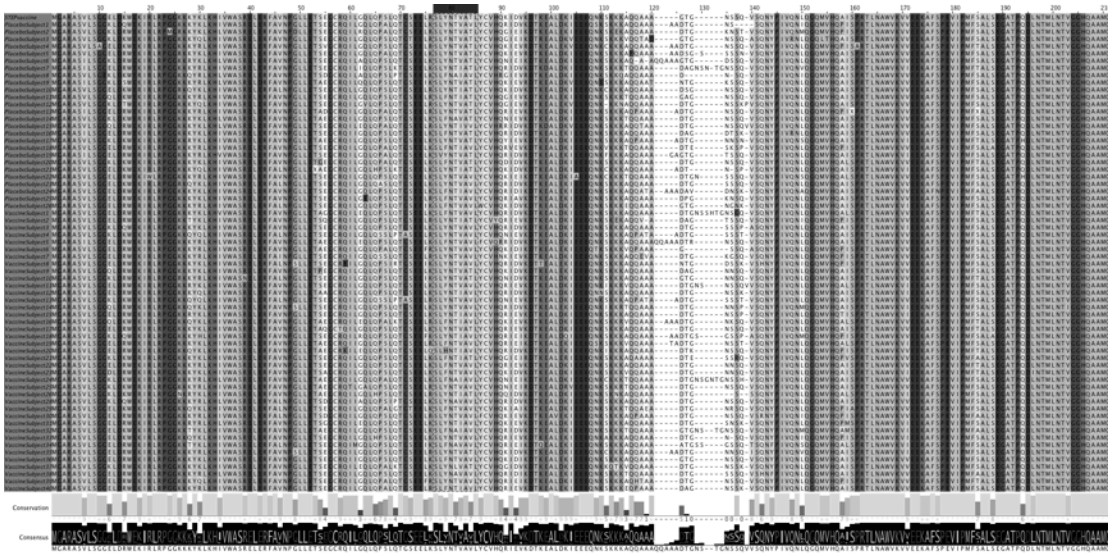
- Vaccination should induce an immune response
  that targets circulating HIV
  (at least the HIV that's similar to the vaccine HIV)
- Idea: investigate the sequence data
- If we see evidence for a difference in the sequences of viruses infecting vaccinees versus placebo recipients,
  - it must be due to the vaccine.
    - (It's a randomized trial!)
- If we see a difference in the sequences of V1V2,
  - then it supports the hypothesis of anti-V1V2 antibodies selectively filtering HIV.

The sieve effect



Click here to view SieveAnimation.swf

Looking for sequence differences



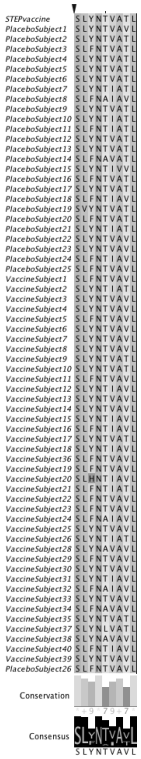
… a needle in a haystack …

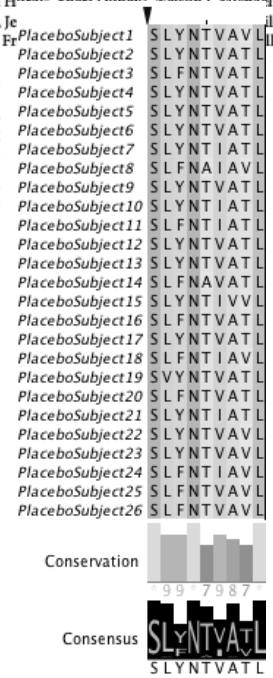
ARTICLES

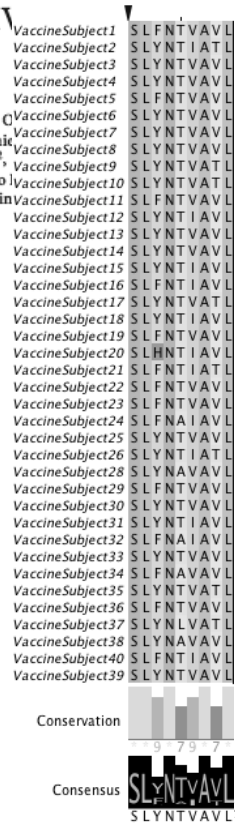
Genetic impact of vaccination on breakthrough HIV sequences from the STEP trial

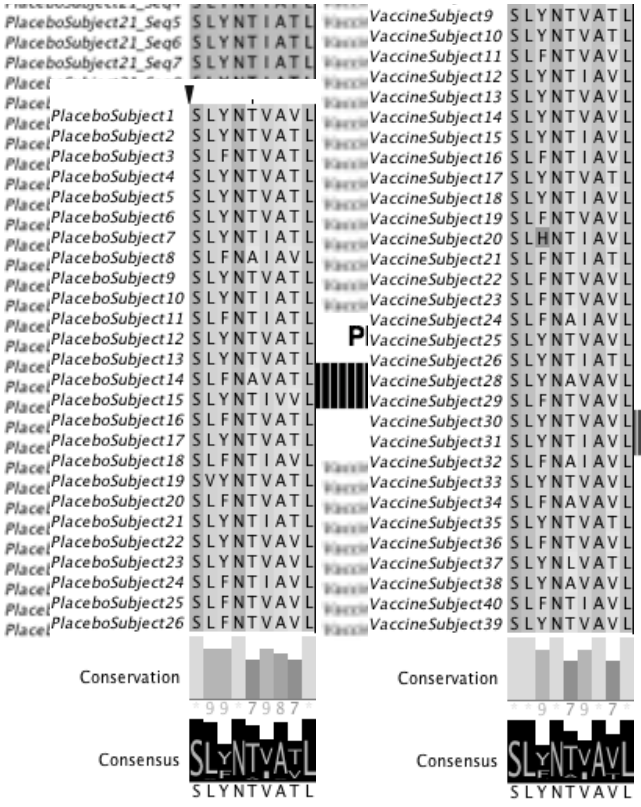
Morgane Rolland[1,7], Sodsai Tovanabutra[2,7], Allan C deCamp[3,7], Nicole Frahm[3,7], Peter B Gilbert[3], Eric Sanders-Buell[2], Laura Heath[1], Craig A Magaret[3], Meera Bose[2], Andrea Bradfield[2], Annemarie C Jacqueline Crossler[2], Teresa Jones[2], Marty Nau[2], Kim Wong[1], Hong Zhao[1], Dana N Raugi[1], Stephanie Julia N Stoddard[1], Brandon S Maust[1], Wenjie Deng[1], John H[illegible], John Shiver[4], Lawrence Corey[3], Fusheng Li[3], Steve G Self[3], Je[illegible], Michael N Robertson[4], Ann Duerr[3], M Juliana McElrath[3], Fr[illegible]

Signature site Gag84 showed the greatest distinction between vaccine and placebo recipients ($q = 0.012$): it is encompassed by several known CTL epitopes, including the well-characterized HLA-A*02 epitope SLYNTVATL (amino acids 77–85; italicized T is the site showing the distinction between vaccine and placebo recipients). Thirty-six of 64 subjects had an HLA class I allele restricting epitopes that spanned Gag84 (ref. 6). The signature at Gag84 was more pronounced among individuals with an HLA allele matching an 'A-list' epitope (epitopes fulfilling criteria intended to ensuring reliable description of the optimal length epitope and correct assignment of the restricting HLA class I alleles)[7] (79%:17% mismatch vaccine:placebo compared to 80%:46% mismatch in the 28 subjects without an A-list restricting allele), supporting that vaccine-induced T cell pressure









We begin with Sanger sequences, usually multiple per subject.

We align and translate the DNA sequences to AAs.

Some analysis methods use all of the subjects' sequences.

Others use one per subject: a representative sequence.

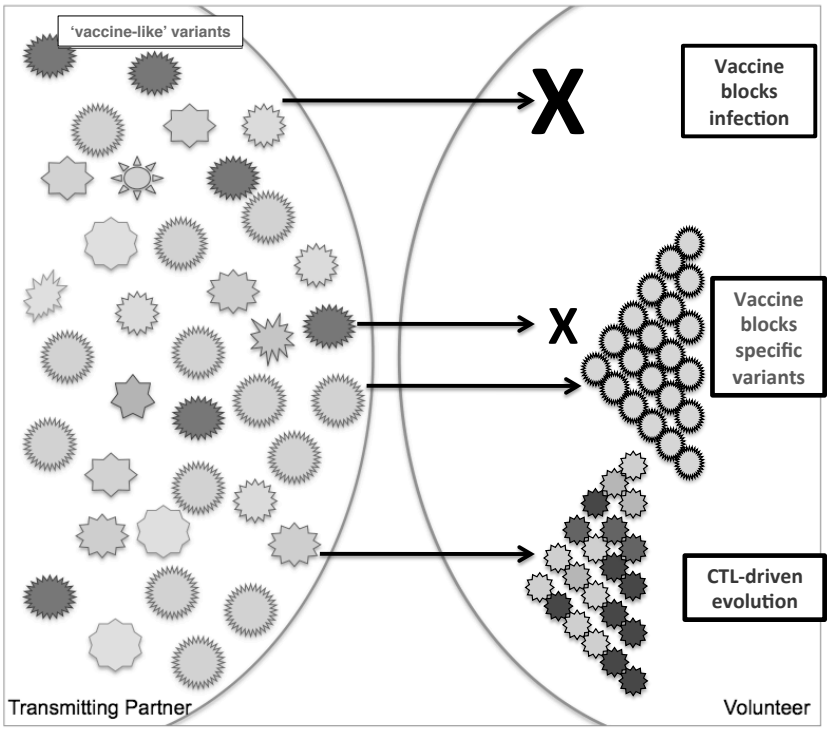
Two Types of Potential Selective Effects

1. Acquisition Sieve Effect
The vaccine selectively blocks (or enhances) acquisition with specific HIV variants

2. Post-Infection Selective Effect
The vaccine drives HIV sequence evolution

- **Longitudinal HIV sequences (and some acute-phase sequences) are needed to distinguish these two types of effects**
- **But at the moment we only have one time-point per subject**

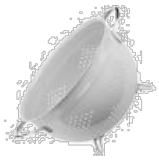
Potential selective effects of vaccines



Challenged Statistical Power

- Achieving high statistical power requires:
  - Large n of infected subjects with sequence data
  - A vaccine that induces immune responses that 'react strongly' with the infecting viruses.

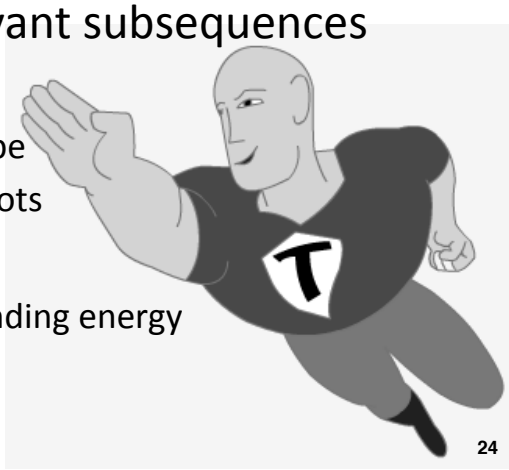
- For most HIV trials, the sieve analyses have low power
  - rv144: n = 121
    - **But for analysis, only n = 110**
      - **(44 vaccine recipients, 66 placebo)**
  - Phambili: n = 82
    - **But for analysis, only n = 43**
      - **(23 vaccine recipients, 20 placebo)**
  - STEP: n = 66
  - VaxGen: n = 336

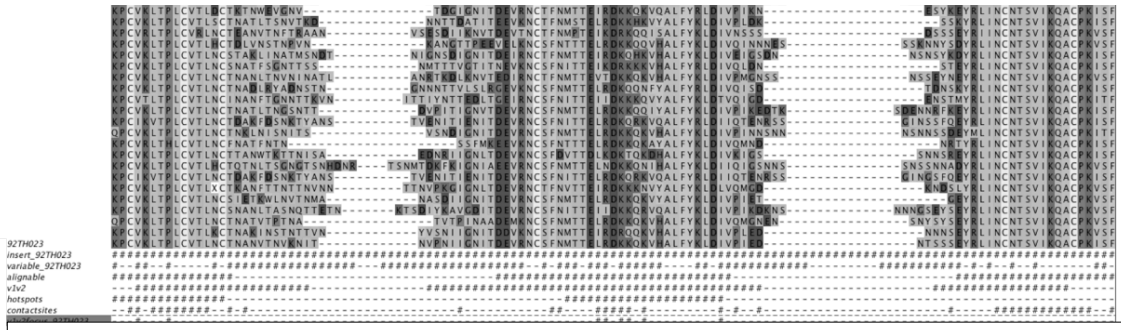


- Can only detect relatively large sieve effects

Maximizing power

- Compare sequences to the vaccine insert
- Pre-filter based on treatment-blinded data
  - Fewer analyses ⇨ greater power
- Focus analysis on relevant subsequences
  - Epitopes
    - CTL epitopes by HLA type
    - Antibody binding hotspots
  - Escape routes
    - Consider changes to binding energy
- **Plan ahead**



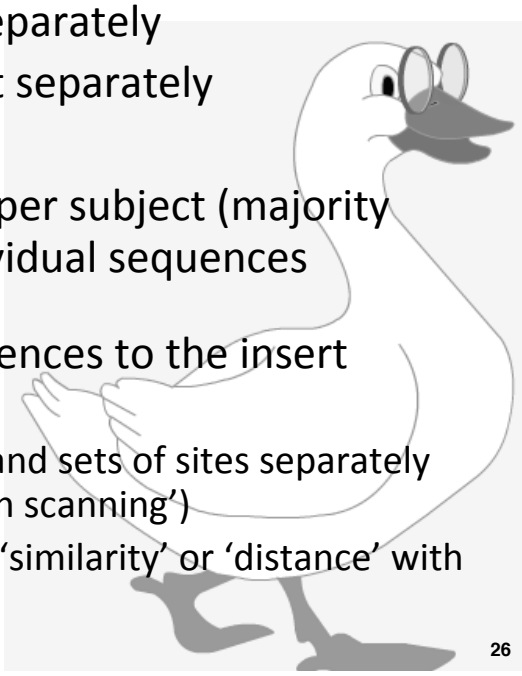


Screening to Maximize Statistical Power

- Only include sites contained in every one of these sets:
  - The 85 sites in the V1V2 region
  - Sites with sufficient variability
  - Sites for which we have confidence in the alignment
  - Sites in antibody-relevant sites
    - (we asked our expert colleagues for sites)
- All of this screening is done **before unblinding**

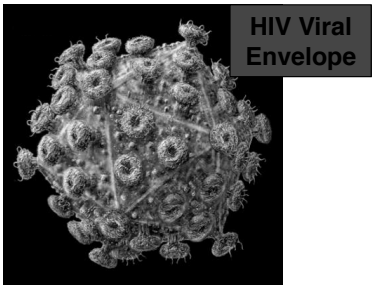
Methods for RV144 Sieve Analysis

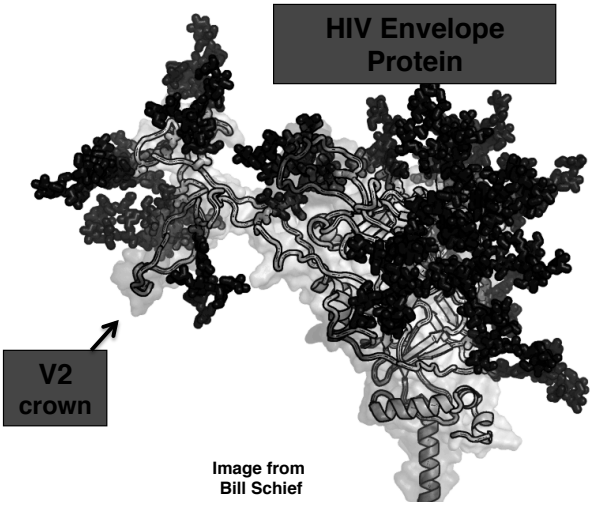
- Assess each HIV-1 gene separately
- Assess each vaccine insert separately
- Assess either 1 sequence per subject (majority consensus) or use all individual sequences
- Compare a subject's sequences to the insert sequence in 2 ways:
  - Local: Evaluate each site and sets of sites separately (eg. 'site scanning', 'antigen scanning')
  - Global: Summarize overall 'similarity' or 'distance' with a single number



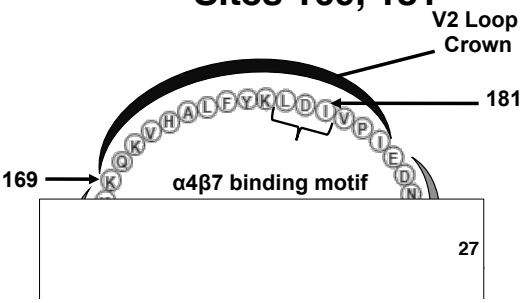
Summary of RV144 V1V2 Results

- V1V2 focused analysis.
- Analyzed only 9 sites!
- Used multiplicity correction to protect against false discoveries.





- **2 sites with evidence of a sieve effect:**
  - **Sites 169, 181**



Vaccine Efficacy by HIV-1 Genotype

(Defined by Site 169, 181)

HIV-1 Genotype	Number Infections	Estimated VE*	95% CI	P-value
169 match	87	48%	18% to 66%	0.0036
169 mismatch	23	-55%	-258% to 33%	0.30
181 match	88	17%	-26% to 45%	0.38
181 mismatch	22	78%	35% to 93%	0.0028

- VE greater against 169-matched than mismatched HIV-1: p = 0.034**
- VE greater against 181-mismatched than matched HIV-1: p = 0.024

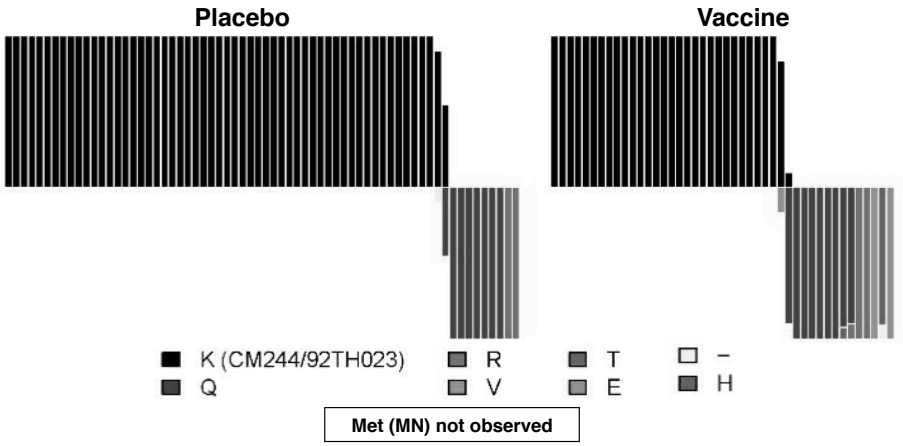
*** Estimated with a Cox model (Prentice et al., Biometrics, 1978)**

**** Estimated with a Cox model (Lunn and McNeil, Biometrics, 1995)**

Position 169

Gilbert, Wu, Jobes:
p =.018, q =.077

Model-Based Sieve:
P(sieveldata) = .334, p =.050, q = .202

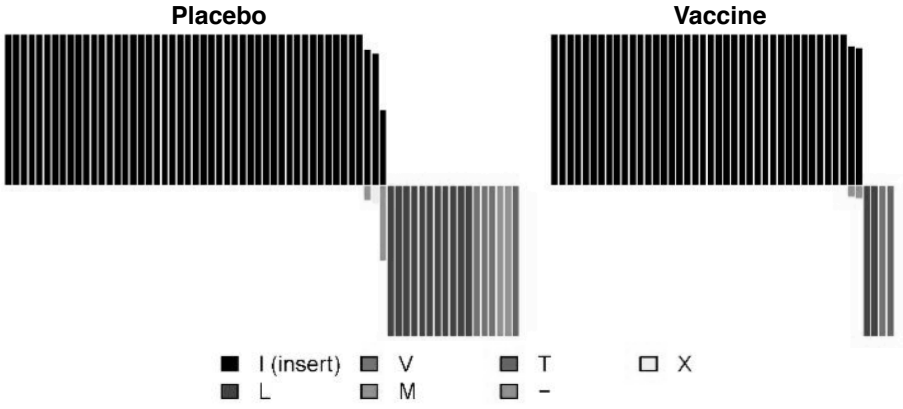


Key:
Each subject is represented by a bar
Bars all have equal height. Insert AA residue, in black, is shown above the midline
Within a bar, colors depict the fraction of the subject's sequences with that AA residue

Position 181

Gilbert, Wu, Jobes:
p=.019, q =.077

Model-Based Sieve:
P(sieveldata) = .002, p =.021, q = .065



Key:
Each subject is represented by a bar
Bars all have equal height. Insert AA residue, in black, is shown above the midline
Within a bar, colors depict the fraction of the subject's sequences with that AA residue