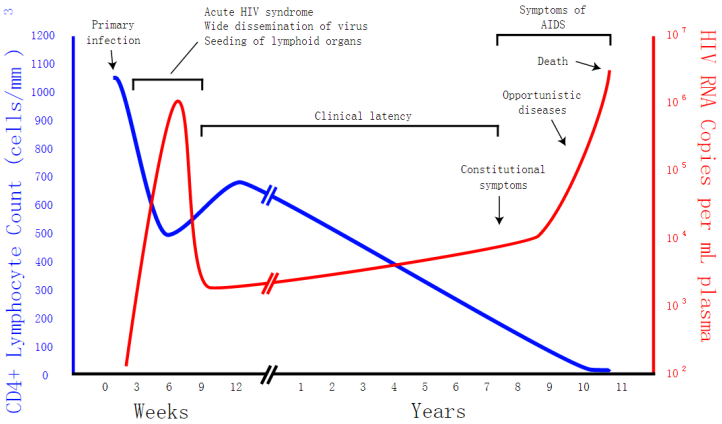
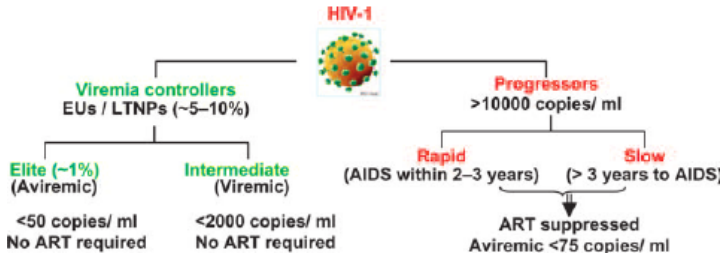
# HIV infection




# HIV Progressor vs Controllers




# HLA-HIV associations

## The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation

The International HIV Controllers Study*†

Infectious and inflammatory diseases have repeatedly shown strong genetic associations within the major histocompatibility complex (MHC); however, the basis for these associations remains elusive. To define host genetic effects on the outcome of a chronic viral infection, we performed genome-wide association analysis in a multiethnic cohort of HIV-1 controllers and progressors, and we analyzed the effects of individual amino acids within the classical human leukocyte antigen (HLA) proteins. We identified >300 genome-wide significant single-nucleotide polymorphisms (SNPs) within the MHC and none elsewhere. Specific amino acids in the *HLA-B* peptide binding groove, as well as an independent *HLA-C* effect, explain the SNP associations and reconcile both protective and risk *HLA* alleles. These results implicate the nature of the HLA–viral peptide interaction as the major factor modulating durable control of HIV infection.

HIV infection is characterized by acute viremia, often in excess of 5 million viral particles per milliliter of plasma, followed by an average 100-fold or greater decline to a relatively stable plasma virus load set point (*1*). In the absence of antiretroviral therapy, the level of viremia is associated with the rate of $CD4^+$ T cell decline and progression to AIDS. There is substantial interperson variability in the virus load set point, with most individuals having stable levels exceeding 10,000 RNA copies/ml. Yet a small number of people demonstrate sustained ability to control HIV replication without therapy. Such individuals, referred to as HIV controllers, typically maintain stable $CD4^+$ cell counts, do not develop clinical disease, and are less likely to transmit HIV to others (*2*).

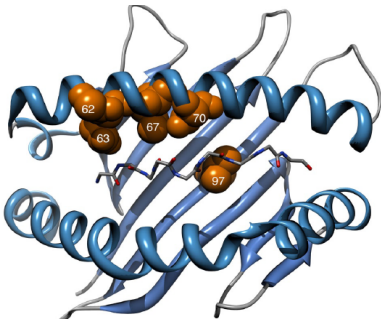
To determine the genetic basis for this rare phenomenon, we established a multinational consortium (www.hivcontrollers.org) to recruit HIV-1 controllers, who are defined by at least three measurements of plasma virus load (VL) < 2000 RNA copies/ml over at least a 12-month period in the absence of antiviral therapy. We performed a genome-wide association study (GWAS) in the HIV controllers (median VL, CD4 count, and disease duration of 241 copies/ml, 699 cells/mm$^3$, and 10 years, respectively) and treatment-naïve chronically infected individuals with advanced disease (median VL and CD4 count of 61,698 copies/ml and 224 cells/mm$^3$, respectively) enrolled in antiviral treatment studies led by the AIDS Clinical Trials Group. After quality control and imputation on the basis of HapMap

*All authors with their contributions and affiliations appear at the end of this paper.
†To whom correspondence should be addressed. E-mail: bwalker@partners.org (B.D.W.); pdebakker@rics.bwh.harvard.edu (P.I.W.d.B.)



# Best determinant of HIV-1 disease progression: HLA-B

- >300 significant SNPs within the MHC region
- Specific amino acids in the HLA-B binding groove best determinants of HIV-1 control.
- Independent HLA-C effect



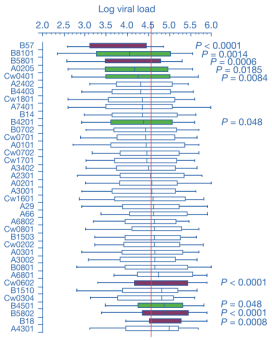

The international HIV-1 Controllers Study Science Express, 2010

## Influence of HLA on disease progression

- Rate of HIV disease progression strongly associated with expression of particular HLA-B alleles.

**low VL**: HLA-B*5703, B*8101, B*5801, B*4201, B*0702, B*1516, B*1517, B*1302, B*1301, B*3901 and B*3902

**high VL**: HLA-B*4501, B*5802, B*1801, B*1503 and B*5301






Kiepiela et al., Nature. 2004 Dec 9;432(7018):769-75

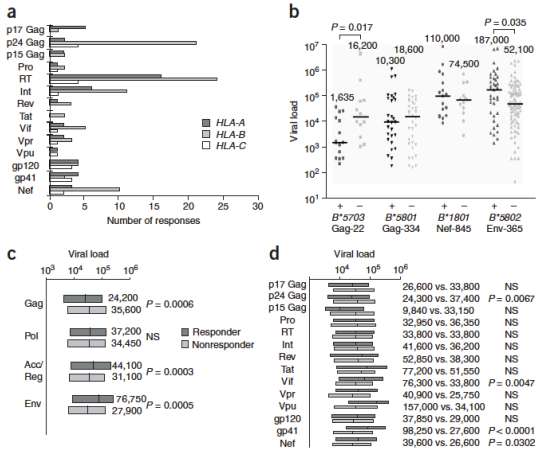
# CD8+ T-cell responses to different HIV proteins have discordant associations with viral load

Photini Kiepiela[1], Kholiswa Ngumbela[1], Christina Thobakgale[1], Dhanwanthie Ramduth[1], Isobella Honeyborne[2], Eshia Moodley[1], Shabashini Reddy[1], Chantal de Pierres[1], Zenele Mncube[1], Nompumelelo Mkhwanazi[1], Karen Bishop[1], Mary van der Stok[1], Kriebashnie Nair[1], Nasreen Khan[1], Hayley Crawford[2], Rebecca Payne[2], Alasdair Leslie[2], Julia Prado[2], Andrew Prendergast[2], John Frater[2], Noel McCarthy[3], Christian Brander[4], Gerald H Learn[5], David Nickle[5], Christine Rousseau[5], Hoosen Coovadia[1], James I Mullins[5], David Heckerman[6], Bruce D Walker[1,4,7] & Philip Goulder[1,2,4]

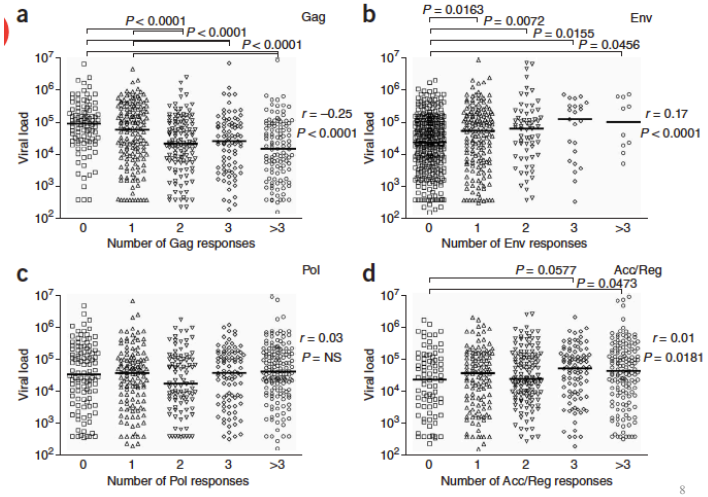
07 Nature Publishing Group http://www.nature.com/naturemedicine

Selection of T-cell vaccine antigens for chronic persistent viral infections has been largely empirical. To define the relationship, at the population level, between the specificity of the cellular immune response and viral control for a relevant human pathogen, we performed a comprehensive analysis of the 160 dominant CD8+ T-cell responses in 578 untreated HIV-infected individuals from KwaZulu-Natal, South Africa. Of the HIV proteins targeted, only Gag-specific responses were associated with lowering viremia. Env-specific and Accessory/Regulatory protein-specific responses were associated with higher viremia. Increasing breadth of Gag-specific responses was associated with decreasing viremia and increasing Env breadth with increasing viremia. Association of the specific CD8+ T-cell response with low viremia was independent of *HLA* type and unrelated to epitope sequence conservation. These population-based data, suggesting the existence of both effective immune responses and responses lacking demonstrable biological impact in chronic HIV infection, are of relevance to HIV vaccine design and evaluation.







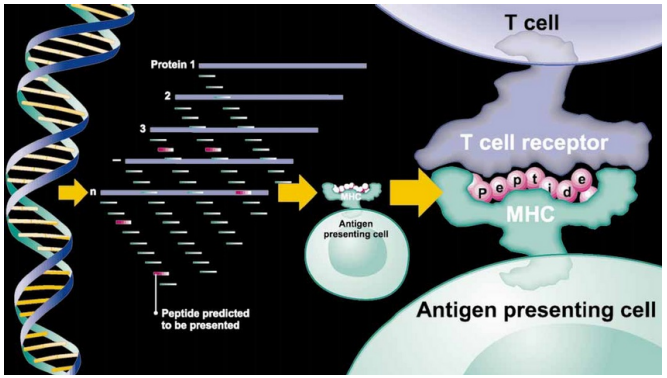



## Are B57 and B18 presenting epitopes from different HIV-1 proteins??

A computational approach!

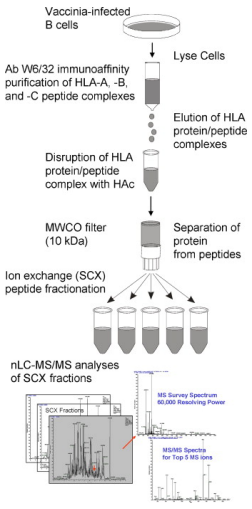


## Genomes to Epitopes




## How do we know which peptides bind to an MHC molecule?

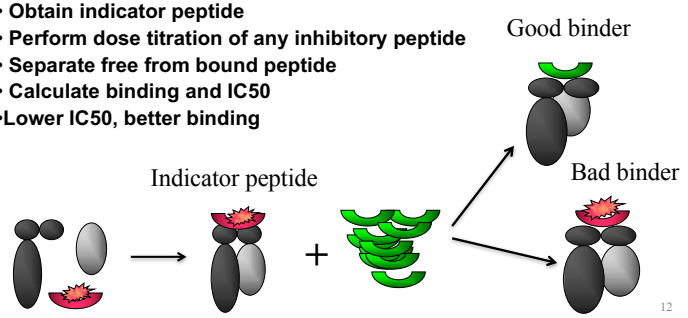
- Via peptide elution from MHC molecules
- Mass spec to identify the peptide
- False positive rate: 5%
- Only few large studies available:
  - Schellens, et al. Frontiers Immunol. 2015



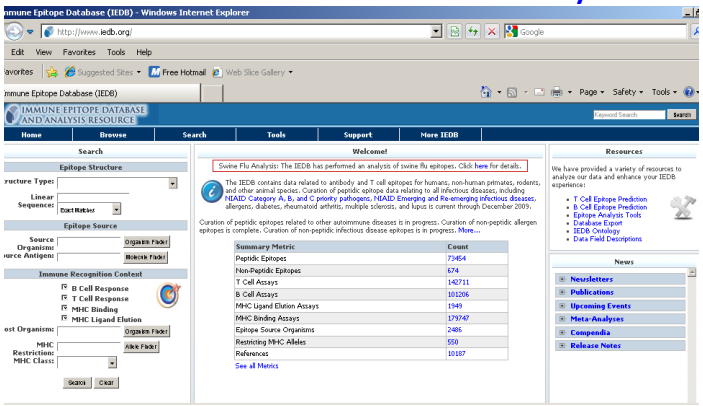



## How do we know which peptides bind to an MHC molecule? Via **Quantitative ELISA**

- **Obtain purified HLA**
- **Or recombinant heavy chain &$\beta_2$m**
- **Obtain indicator peptide**
- **Perform dose titration of any inhibitory peptide**
- **Separate free from bound peptide**
- **Calculate binding and IC50**
- **Lower IC50, better binding**



# IEDB provides quantitative data and results of elution assays






# Using the available data we can develop methods to predict MHC binding

- Simple/Extended Motifs
  - Allowed/non allowed amino acids

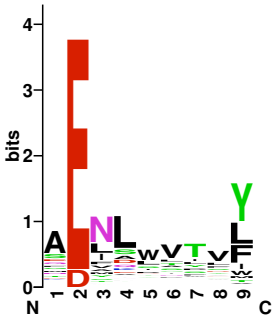
**HLA-A*0201**

| Position | 1 | **2** | 3 | 4 | 5 | 6 | 7 | 8 | **9** |
|---|---|---|---|---|---|---|---|---|---|
| **anchors** or auxiliary anchors | | **L** | | | | V | | | **V** |
| | | **M** | | | | | | | **L** |
| preferred residues | | | | | | | | | |
| | | | | | | **E** | | | **K** |
| | | | | | | **K** | | | |

- Matrix models
  - Peptide statistics from known peptides
- Neural networks
  - A machine learning system that can capture non-linear relations between residues in a peptide



# Weight Matrix approach




- How similar the peptide in question is to the known binders?
- Per position we can calculate the similarity, and then multiple all positions to find out how likely it is that our peptide is a binder.



# Matrix Methods

- Estimate amino acid frequencies from alignment of all binders $p_{ij}$
  - Here i is a position in the motif, and j an amino acid.
- In nature not all amino acids are found equally often
  - $q_A = 0.07$, $q_W = 0.013$
  - Finding 6% A is hence not significant, but 6% W highly significant
  - Better use $p_{ij}/q_{j,}$ where $q_j$ is the background frequency for amino acid j.
- To prevent overflow a weight matrix is given as

  $W_{ij} = \log(p_{ij}/q_j)$
- W is a L x 20 matrix, L is motif length

## Scoring sequences to a weight matrix

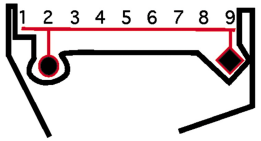
| | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.6 | 0.4 | -3.5 | -2.4 | -0.4 | -1.9 | -2.7 | 0.3 | -1.1 | 1.0 | 0.3 | 0.0 | 1.4 | 1.2 | -2.7 | 1.4 | -1.2 | -2.0 | 1.1 | 0.7 |
| 2 | -1.6 | -6.6 | -6.5 | -5.4 | -2.5 | -4.0 | -4.7 | -3.7 | -6.3 | 1.0 | 5.1 | 3.7 | 3.1 | -4.2 | -4.3 | -4.2 | -0.2 | -5.9 | -3.8 | 0.4 |
| 3 | 0.2 | -1.3 | 0.1 | 1.5 | 0.0 | -1.8 | -3.3 | 0.4 | 0.5 | -1.0 | 0.3 | -2.5 | 1.2 | 1.0 | -0.1 | -0.3 | -0.5 | 3.4 | 1.6 | 0.0 |
| 4 | -0.1 | -0.1 | -2.0 | 2.0 | -1.6 | 0.5 | 0.8 | 2.0 | -3.3 | 0.1 | -1.7 | -1.0 | -2.2 | -1.6 | 1.7 | -0.6 | -0.2 | 1.3 | -6.8 | -0.7 |
| 5 | -1.6 | -0.1 | 0.1 | -2.2 | -1.2 | 0.4 | -0.5 | 1.9 | 1.2 | -2.2 | -0.5 | -1.3 | -2.2 | 1.7 | 1.2 | -2.5 | -0.1 | 1.7 | 1.5 | 1.0 |
| 6 | -0.7 | -1.4 | -1.0 | -2.3 | 1.1 | -1.3 | -1.4 | -0.2 | -1.0 | 1.8 | 0.8 | -1.9 | 0.2 | 1.0 | 0.4 | 0.6 | 0.4 | -0.5 | -0.0 | 2.1 |
| 7 | 1.1 | -3.8 | -0.2 | -1.3 | 1.3 | -0.3 | -1.3 | -1.4 | 2.1 | 0.6 | 0.7 | -5.0 | 1.1 | 0.9 | 1.3 | -0.5 | -0.9 | 2.9 | -0.4 | 0.5 |
| 8 | -2.2 | 1.0 | -0.8 | -2.9 | -1.4 | 0.4 | 0.1 | -0.4 | 0.2 | -0.0 | 1.1 | -0.5 | -0.5 | 0.7 | -0.3 | 0.8 | 0.8 | -0.7 | 1.3 | -1.1 |
| 9 | -0.2 | -3.5 | -6.1 | -4.5 | 0.7 | -0.8 | -2.5 | -4.0 | -2.6 | 0.9 | 2.8 | -3.0 | -1.8 | -1.4 | -6.2 | -1.9 | -1.6 | -4.9 | -1.6 | 4.5 |

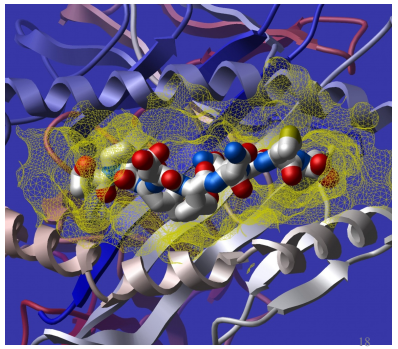
| | |
|---|---|
| ILYQVPFSV | 15.0 |
| ALPYWNFAT | -3.4 |
| MTAQWWLDA | 0.8 |



## Do the peptide positions contribute independently to binding?

Two adjacent amino acids may compete for the space in a pocket in the MHC molecule.

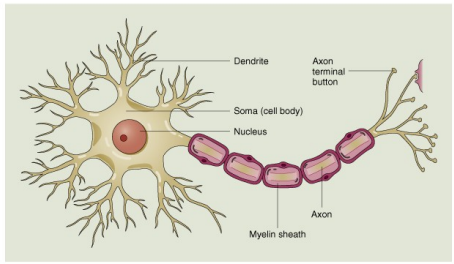


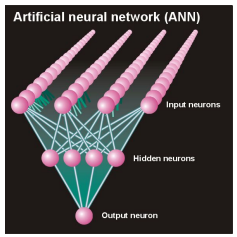


## If the interactions between positions is a major component …..

- Artificial neural networks (ANN) are ideally suited to take such correlations into account




© 2000 John Wiley & Sons, Inc.

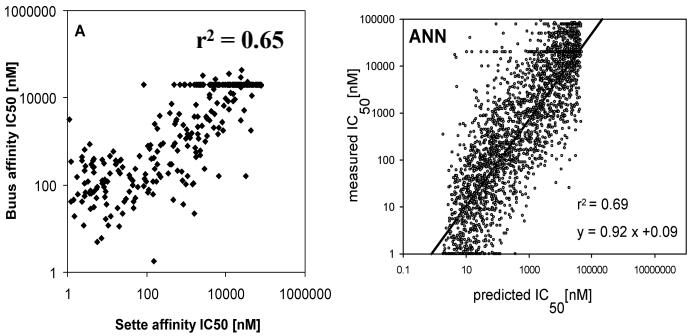





## Choice of method

- Simple and extended motif method when little or no data is available
- Neural networks when there is a lot of data available
- Weight matrices otherwise
  - Increase weight on anchor positions

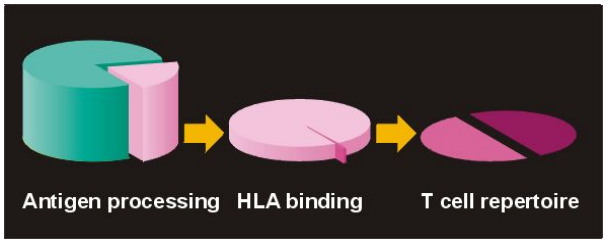
## How good are the predictions?




*Peters et al, PLOS Comp. Biol. 2006*

## From proteins to T cell epitopes

20% processed 0.5-5% bind MHC 50% CTL response




Lauemøller et al., 2000

=> 1/2000 peptide are immunogenic



## Which peptides are the "binders" or predicted epitopes?

The output of prediction methods are IC50 values (a reflection of competition experiments *in vitro*). **HOW do we then define possible epitopes?????**

- Use a fix threshold, say 500 nM, and decide that all the peptides binding better than this threshold are likely epitopes.
- Take top 1-5% of all possible 9mers as possible epitopes.
- When comparing peptide binding to different HLA molecules: fix thresholds do NOT work! Use ranks instead…



## pMHC predictions: NetMHC

cbs.dtu.dk/services/NetMHC