Advances in Autoimmune Informatics

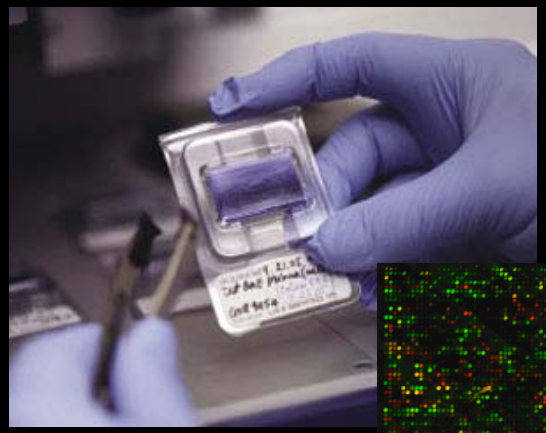
Marina Sirota
Stanford University

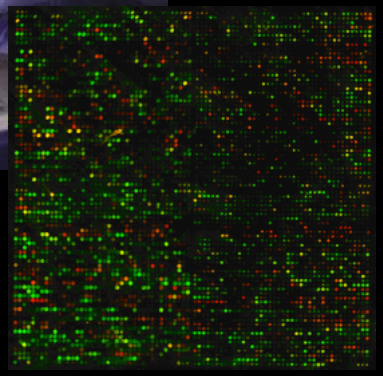
Feb 19, 2010

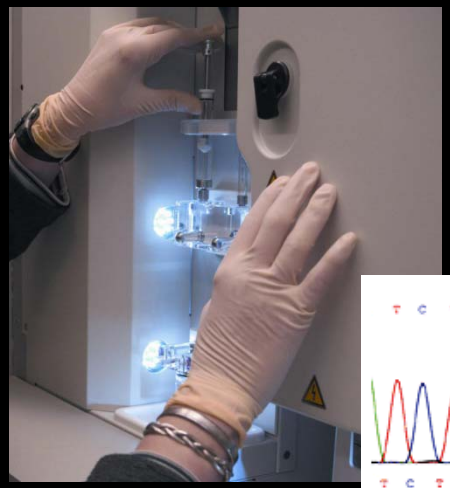
Butte Lab

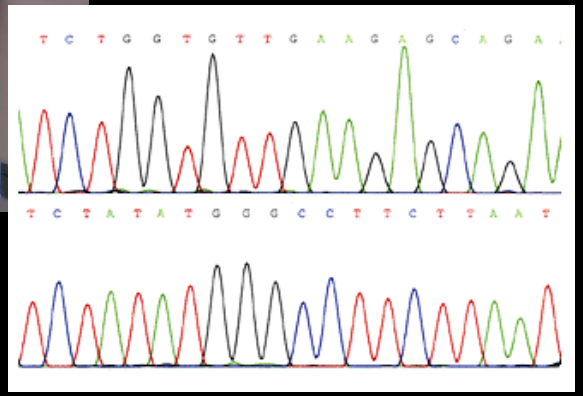
Data Driven Research

Lots of molecular data is publicly available



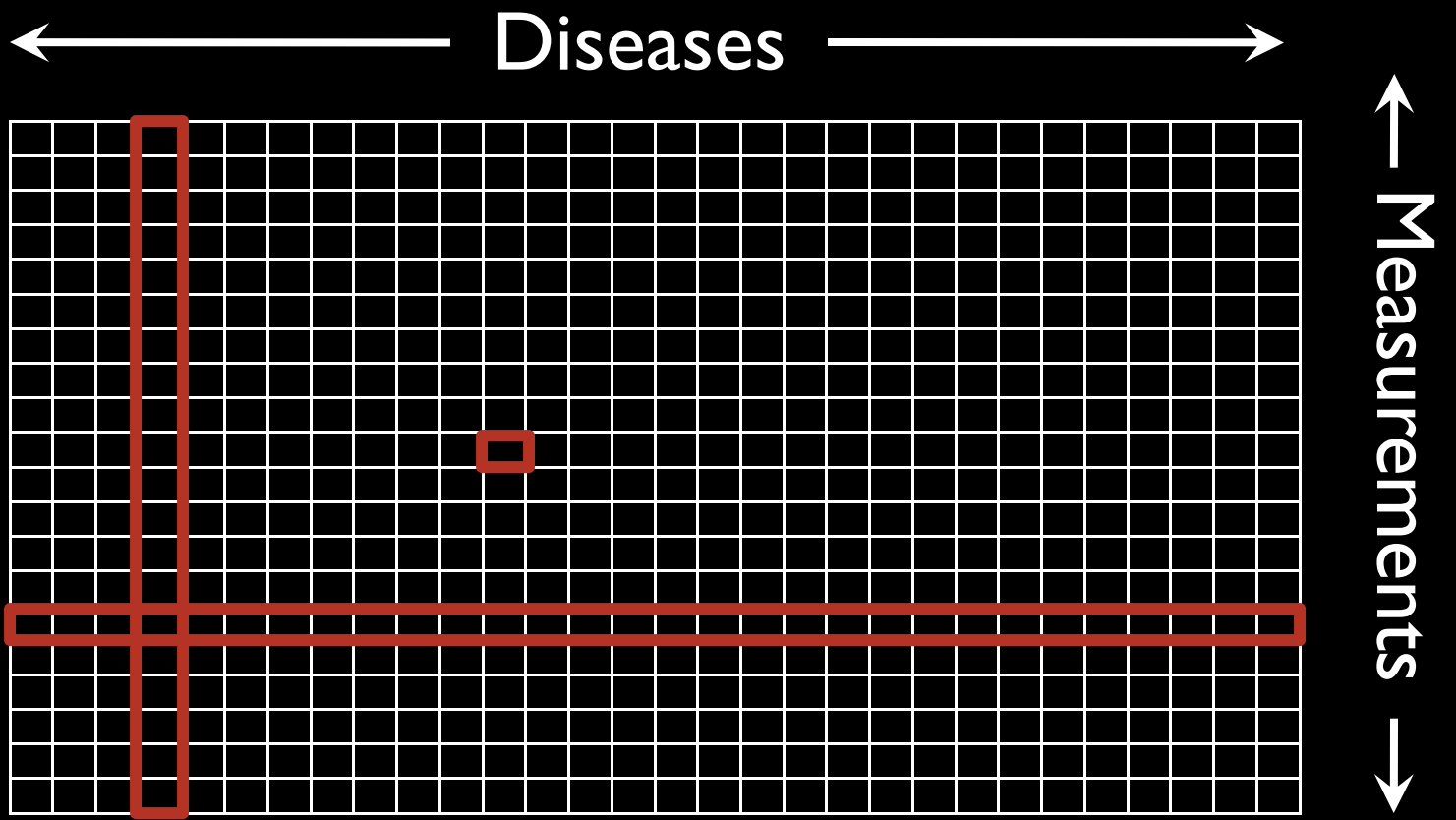






Integrative approaches enable novel questions

Translational Bioinformatics

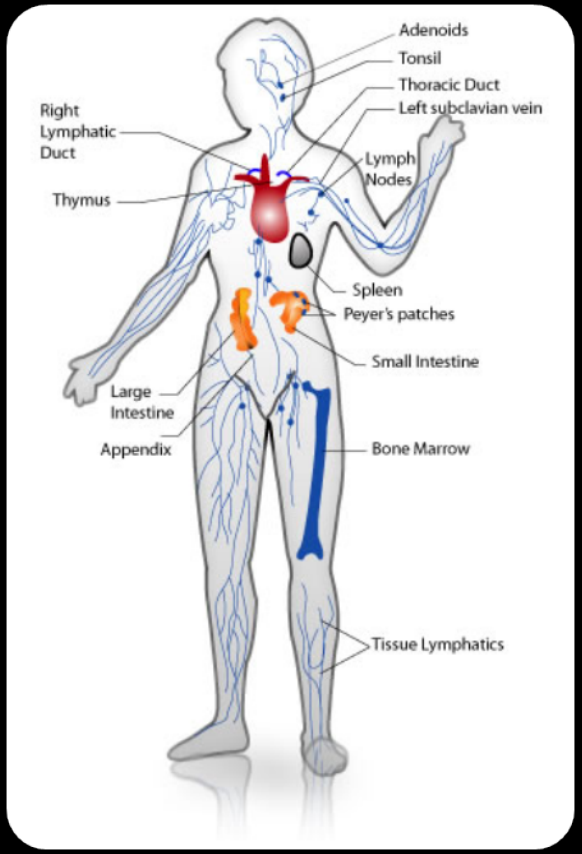


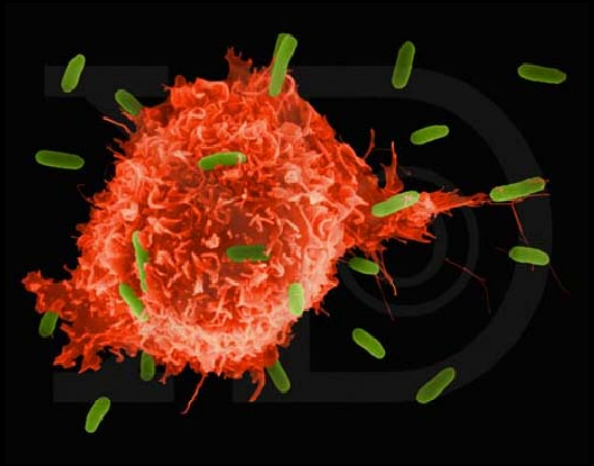
English SB, Butte AJ. Evaluation and integration of 49 genome-wide experiments and the prediction of previously unknown obesity-related genes. Bioinformatics. 2007

Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ. Autoimmune disease classification by inverse association with SNP alleles. PLoS Genet. 2009.

Autoimmune Disease

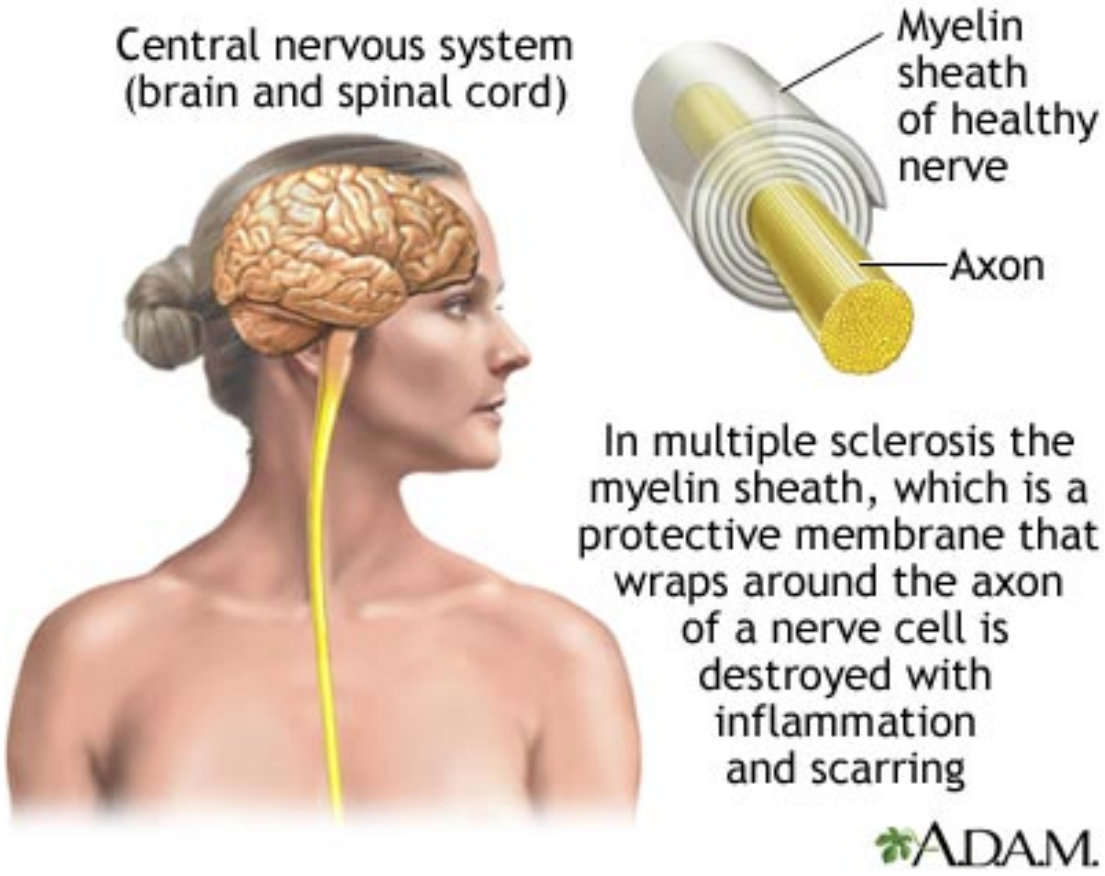
- The immune system mistakes some part of the body as a pathogen and attacks it
- Affects millions of people in the United States and around the world





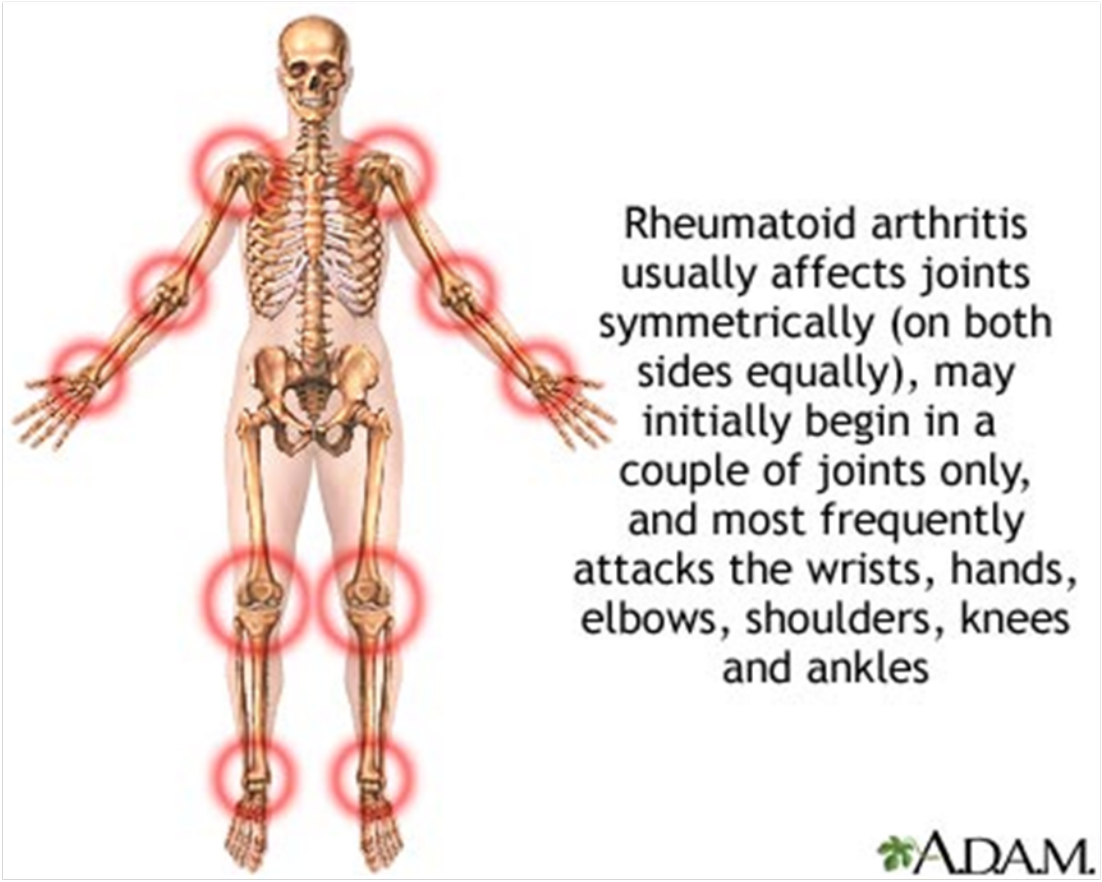
Examples of Autoimmune Disease

Multiple Sclerosis - MS



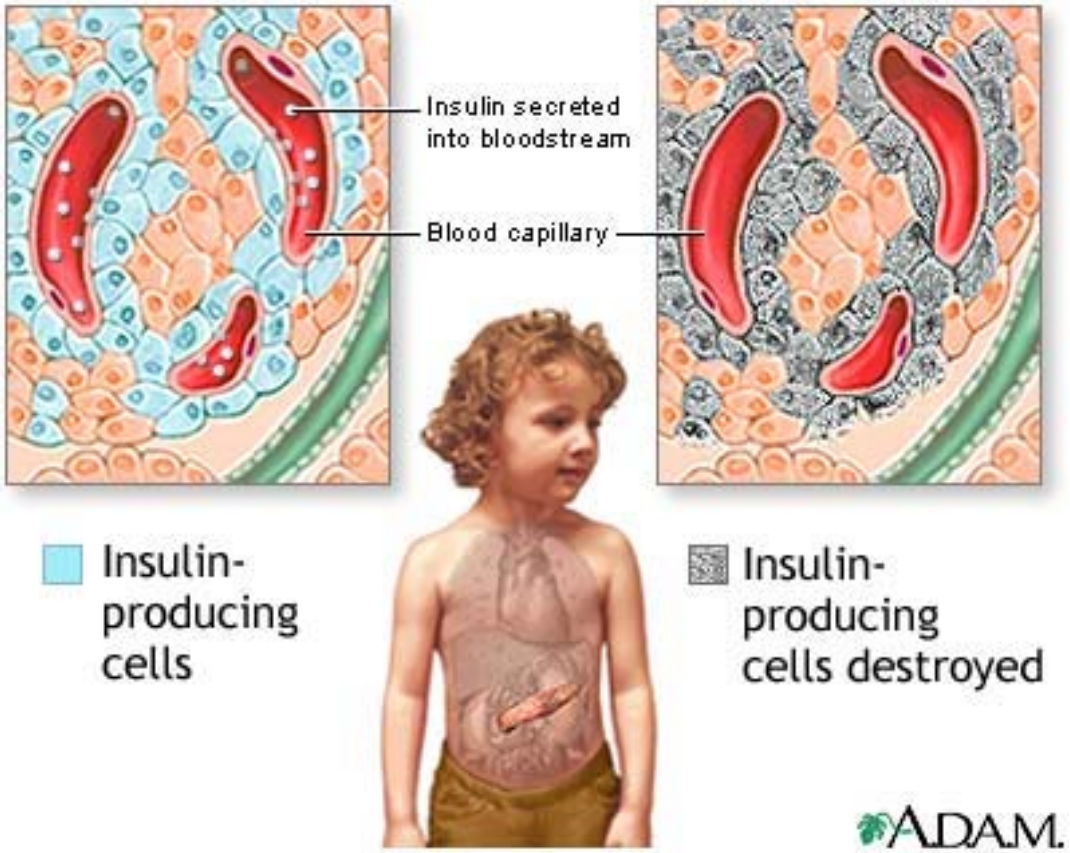
Examples of Autoimmune Disease

Rheumatoid Arthritis - RA



Examples of Autoimmune Disease

Type 1 Diabetes – T1D



Diagnostics and Treatment

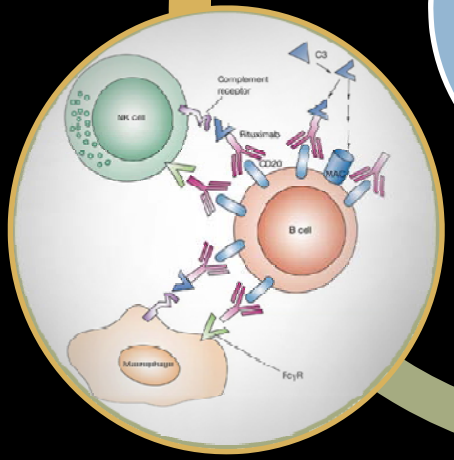
- What triggers the immune system to attack the body?
- What causes one individual to be more susceptible to autoimmune disease than another?
- Are immunosupressants the best solution?

Diagnostics

Autoimmune Disease Classification by Inverse Association with SNP Alleles

Discovery and Validation of Novel Drug Indications Using Gene Expression Data

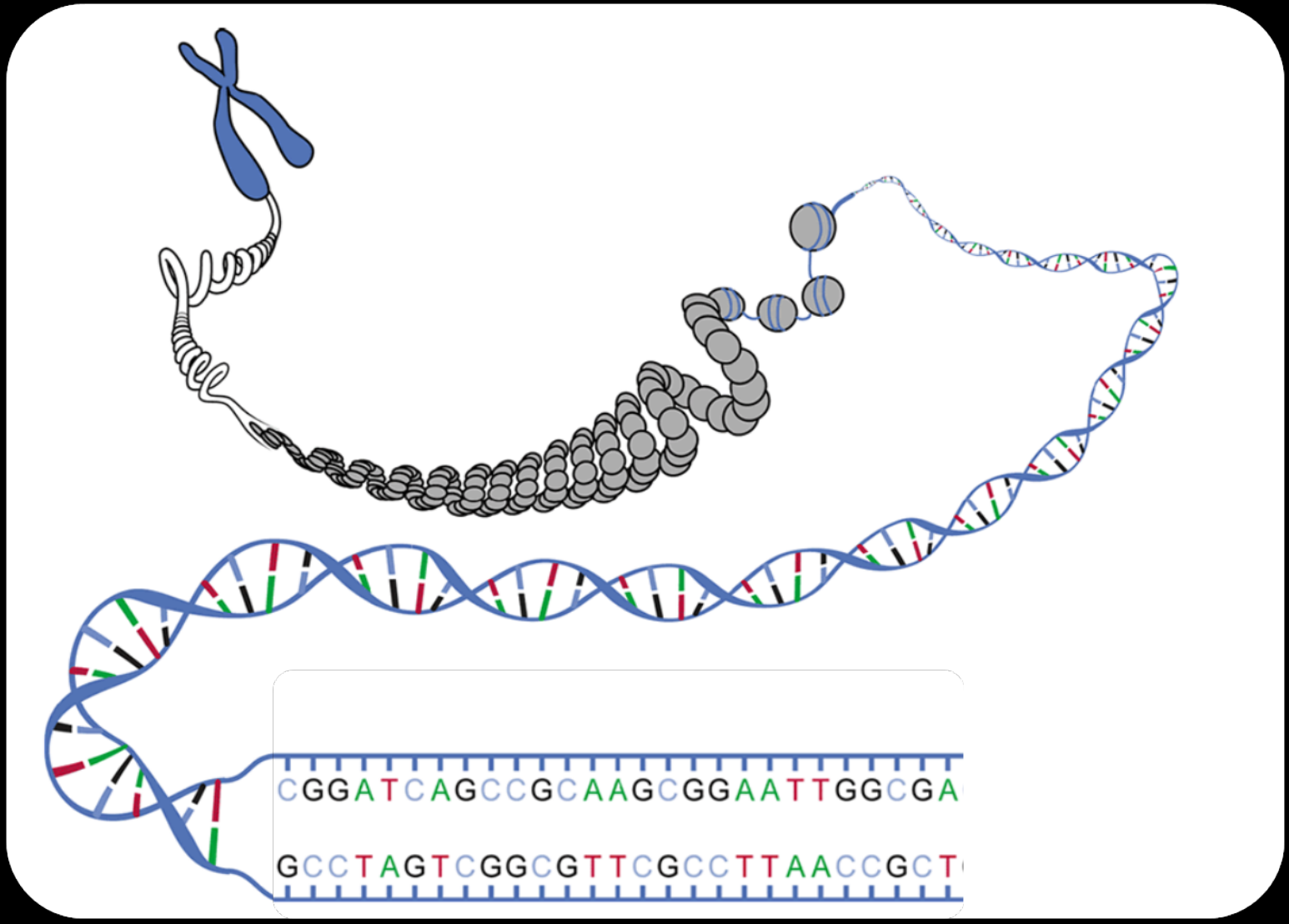
Autoimmune Informatics



Disease Mechanism

Therapeutics

DNA is a Sequence

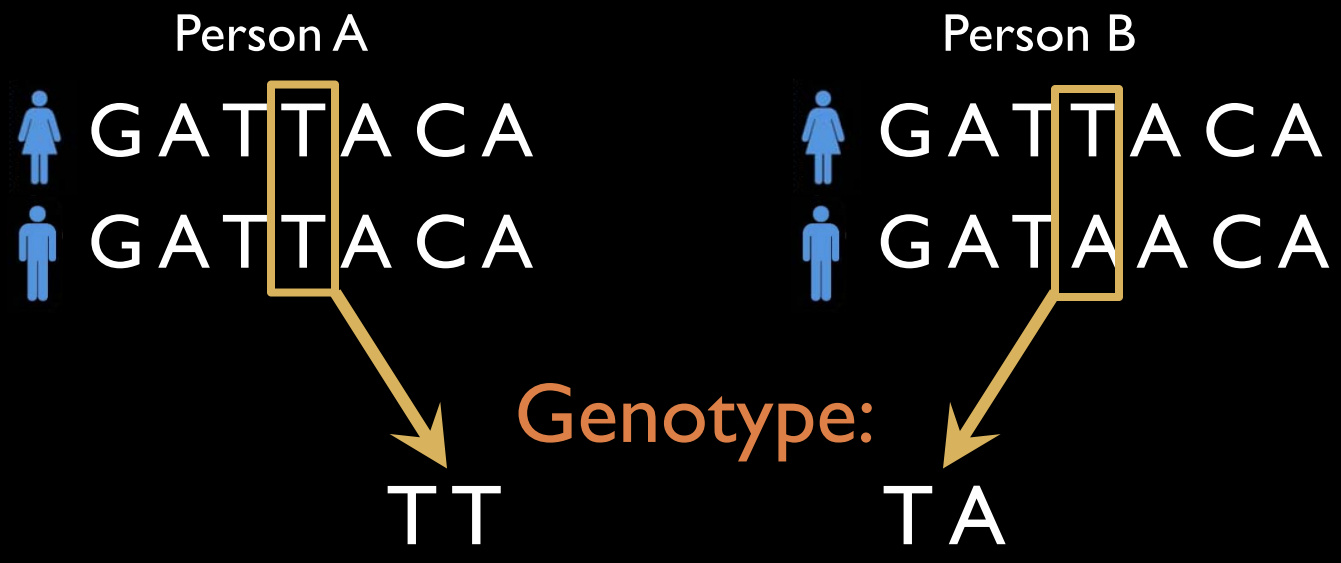


Genetic Variation is Responsible for Individual Differences

- Humans are 99.5% genetically identical
- 0.5% - genetic variation
- Explains phenotypic differences such as eye color and blood group
- Higher or lower disease risk

Single Nucleotide Polymorphisms (SNPs)

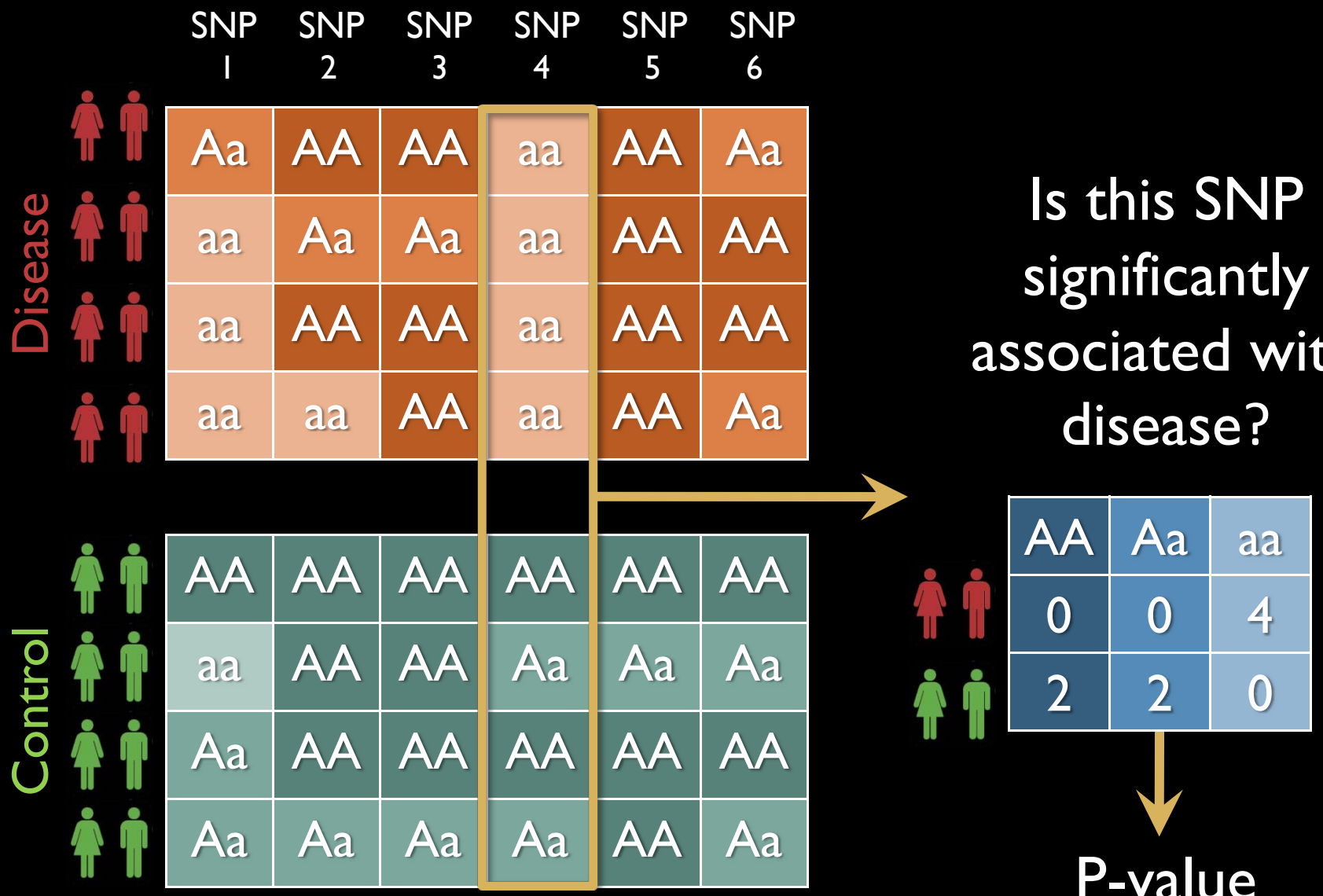
We inherit genetic material from both parents



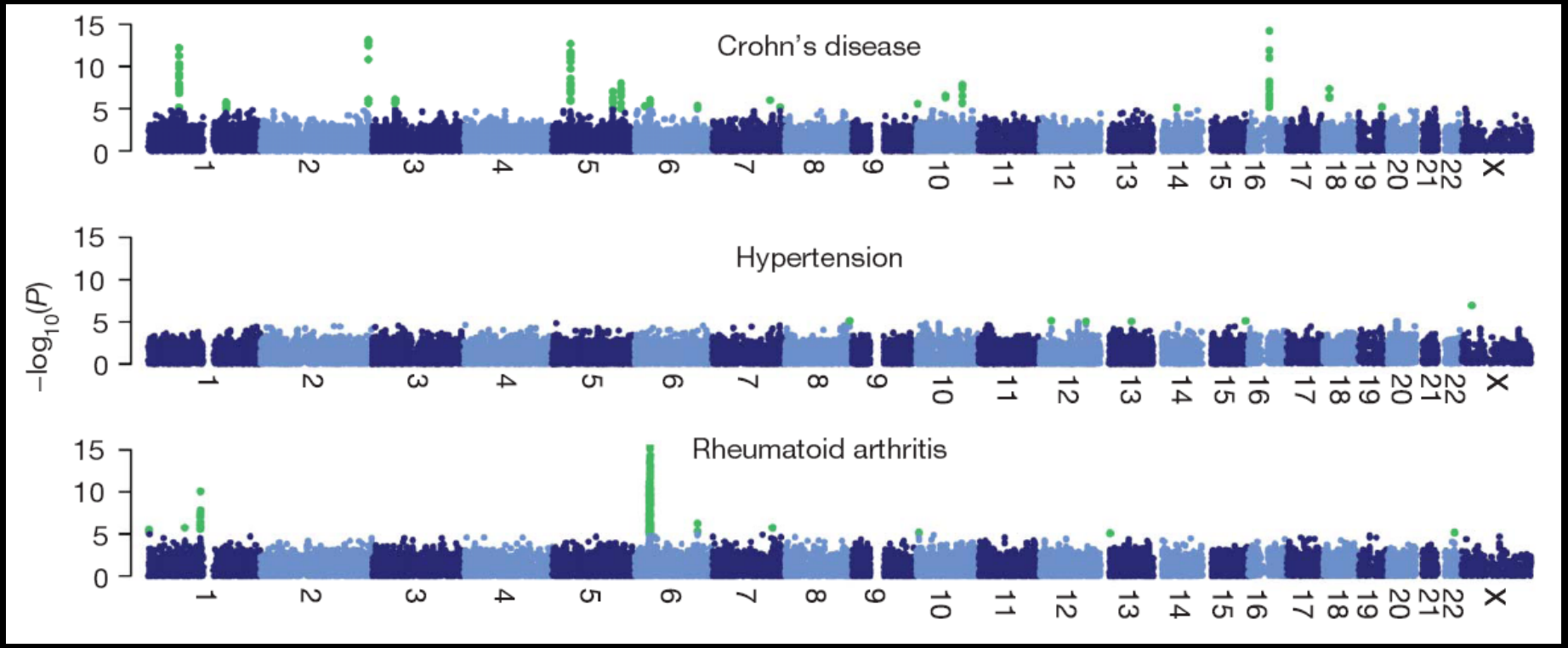
Alleles: T/A

Genetic variation at a single nucleotide position

Genome-wide Association Study



GWAS Results



Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007.

Motivation

- GWAS explore the relationship between genotype and a single disease
- Many such studies for autoimmune diseases present an opportunity to study relationships between diseases based on genetic variation
- Potential impact on both basic science and clinical care

Why is this Interesting?

- Studying similarities between diseases might lead to identifying novel treatment options

- Studying differences between diseases previously thought to be similar may point to key novel disease-specific genes and pathways

Question

Can we learn common disease mechanisms by comparing their genetic variation profiles?

Disease 1	
rs7743761	2.06E-141
rs4711269	7.83E-120
rs9266825	6.61E-119
rs9468877	6.09E-114
rs1051792	1.53E-110
rs2523586	6.34E-107
rs2248462	1.09E-99
rs3915971	1.76E-99
rs4713438	4.07E-90
rs2844529	1.95E-85
rs3998357	2.67E-81
rs2507987	1.11E-77
rs3128982	1.19E-75
rs1265155	8.58E-69
rs9380215	5.74E-68

Disease 2	
rs926070	7.49E-46
rs7192	8.73E-37
rs2227956	9.41E-31
rs2857201	6.99E-25
rs2596473	9.72E-22
rs2523618	8.08E-24
rs1033500	5.83E-21
rs9268384	5.83E-21
rs9268368	5.98E-21
rs2248462	5.94E-19
rs1077394	8.85E-19
rs3132468	9.56E-18
rs6910071	6.77E-17
rs2844529	6.46E-15
rs805262	7.16E-15

WTCCC Data

- Multiple Sclerosis (MS)
- Ankylosing Spondylitis (AS)
- Auto-immune Thyroid Disease (ATD)
- Breast Cancer (BC)

Illumina chip with 15K non synonymous SNPs

- Rheumatoid Arthritis (RA)
- Bipolar Disorder (BD)
- Crohn's Disease (CD)
- Coronary Artery Disease (CAD)
- Hypertension (HT)
- Type 1 Diabetes (T1D)
- Type 2 Diabetes (T2D)

500K Affymetrix chip

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007.

Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ et al: Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007.

WTCCC Data

- Multiple Sclerosis (MS)
- Ankylosing Spondylitis (AS)
- Auto-immune Thyroid Disease (ATD)
- Breast Cancer (BC)

Illumina chip with 15K non synonymous SNPs

- Rheumatoid Arthritis (RA)
- Bipolar Disorder (BD)
- Crohn's Disease (CD)
- Coronary Artery Disease (CAD)
- Hypertension (HT)
- Type 1 Diabetes (T1D)
- Type 2 Diabetes (T2D)

500K Affymetrix chip

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007.

Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ et al: Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007.

"Autoimmune Toggle" SNPs

SNP - Allele	Gene Symbol	P-value				
		RA	AS	T1D	ATD	MS
rs1063635 – A	LOC100129668	6.01E-08	1.83E-59	8.65E-10	8.30E-03	9.27E-05
rs1132200 – A	TMEM39A	2.24E-02	1.77E-02	8.28E-03	4.02E-03	4.56E-03
rs1634717 – A		1.80E-04	6.00E-13	4.94E-14	1.68E-06	3.34E-03
rs204991 – C	GPSM3	3.67E-08	9.34E-13	9.40E-24	5.04E-11	1.70E-03
rs2076530 – G	BTNL2	3.50E-57	8.76E-15	2.64E-14	3.93E-07	3.00E-19
rs2242655 – C	C6orf47	1.21E-03	5.75E-23	1.53E-05	1.13E-02	7.42E-05
rs2248462 – A		1.33E-03	1.09E-99	5.00E-25	1.45E-05	5.94E-19
rs2299851 – T	MSH5	4.91E-02	1.10E-22	1.04E-04	4.69E-02	5.52E-06
rs2517646 – G	TRIM10	1.98E-04	2.29E-03	1.86E-06	1.26E-02	7.59E-06
rs2844463 – T	BAT3	6.40E-07	1.56E-04	1.47E-05	1.15E-02	3.70E-02
rs3129953 – T	BTNL2	2.54E-11	2.13E-09	1.47E-40	2.66E-15	4.18E-05
rs3135363 – C		5.69E-22	7.21E-04	9.81E-12	4.46E-15	5.11E-07
rs4428528 – C		1.01E-18	2.19E-03	8.16E-23	7.11E-12	1.22E-03
rs887464 – A	PSORS1C3	3.20E-03	7.89E-09	7.43E-28	2.03E-05	7.28E-10
rs9267954 – T		2.89E-38	3.27E-13	4.40E-12	2.50E-02	2.17E-14

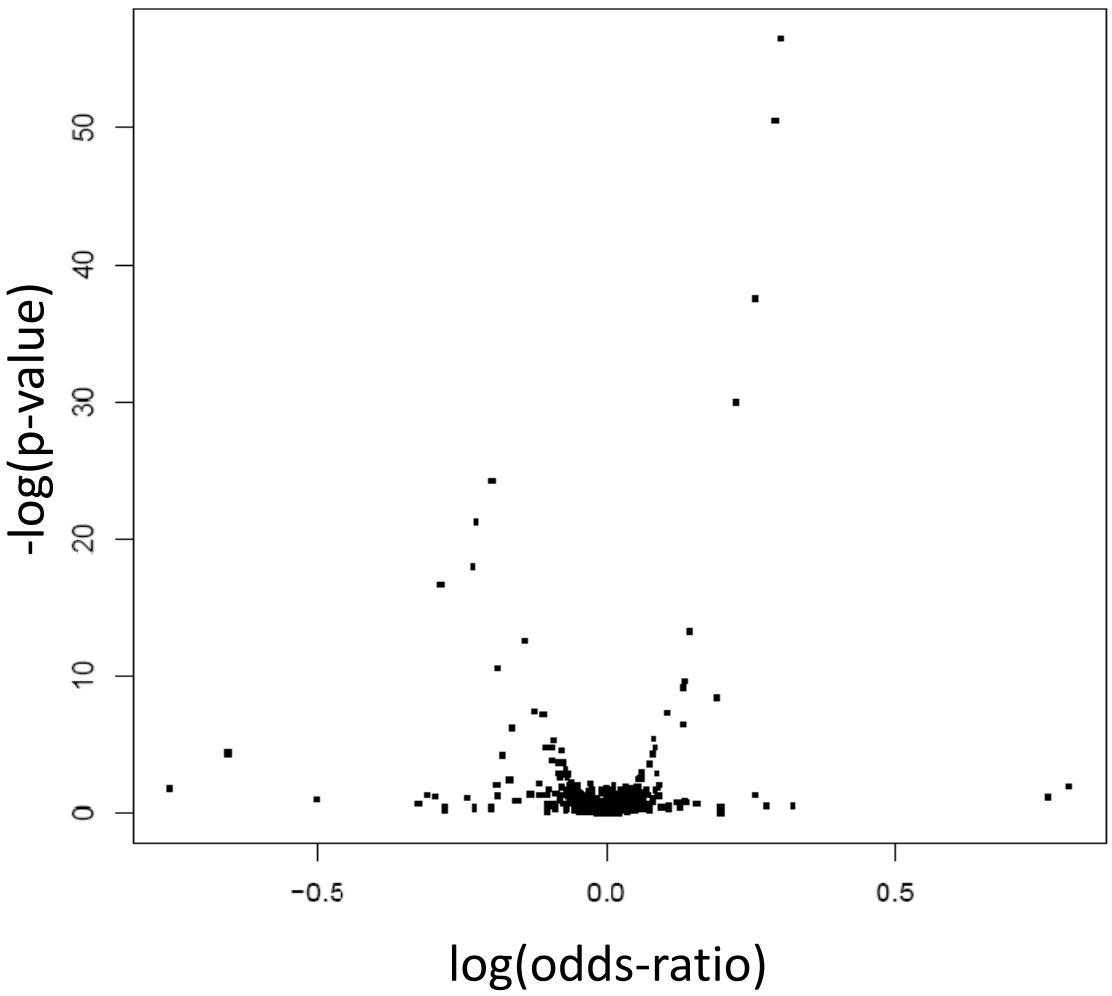
Susceptible to Disease

Protective of Disease

Odds-ratios and P-values

- P-values:
  - How much evidence there is against the hypothesis that the allele distribution in the control and disease populations is the same
  - Confidence of association but is not allele specific
- Odds-ratios:
  - The ratio of the probability that a disease individual has a certain allele to the probability of a healthy control having that allele
  - Capture whether an allele is susceptible or protective but might not be significant

Odds-ratios and P-values Are not Correlated



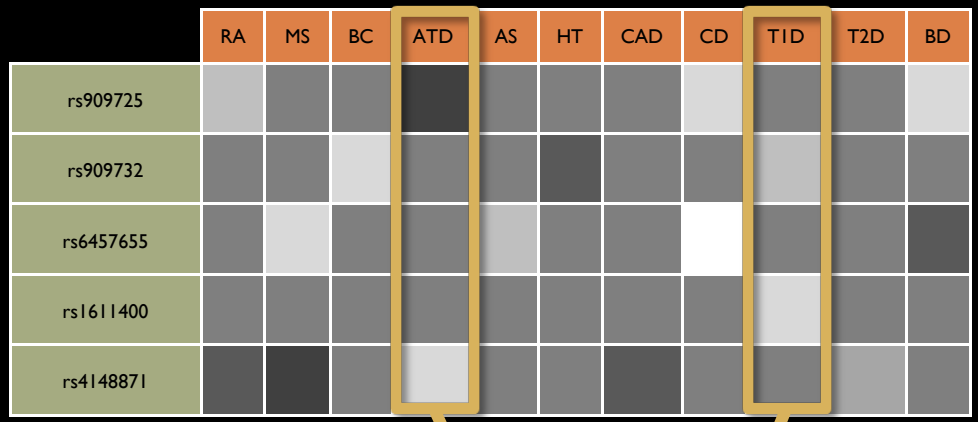
Genetic Variation Score

- For a disease d and a SNP s, define a genetic variation score gvs[d,s] :

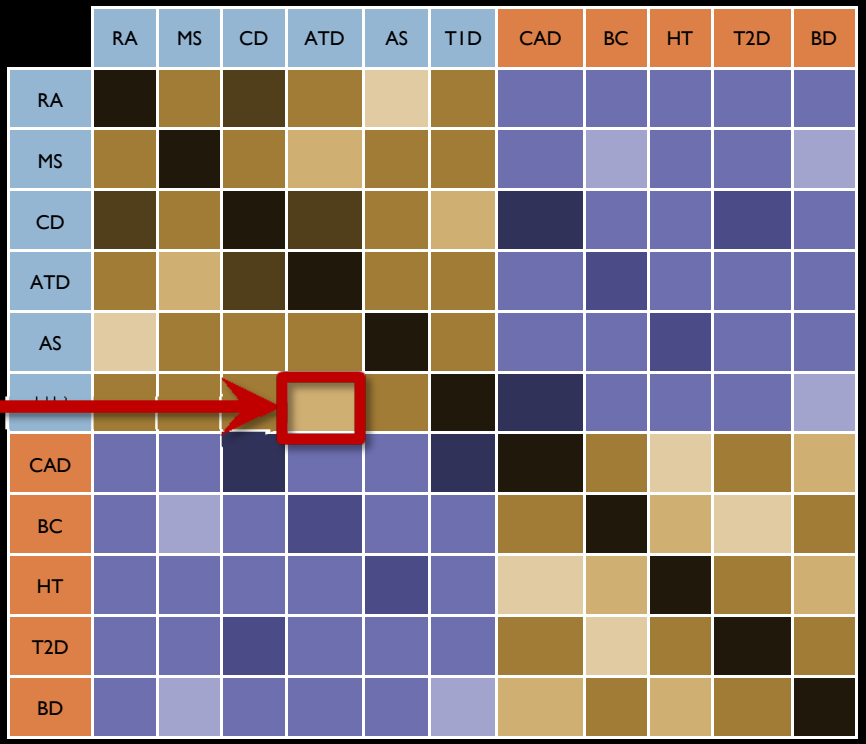
gvs[d,s] = sign(log-odds-ratio[d,s]) * (log(p-value[d,s]))

- Sign of log odds-ratio to capture whether the minor allele is protective or susceptible
- Log p-value to capture confidence of association

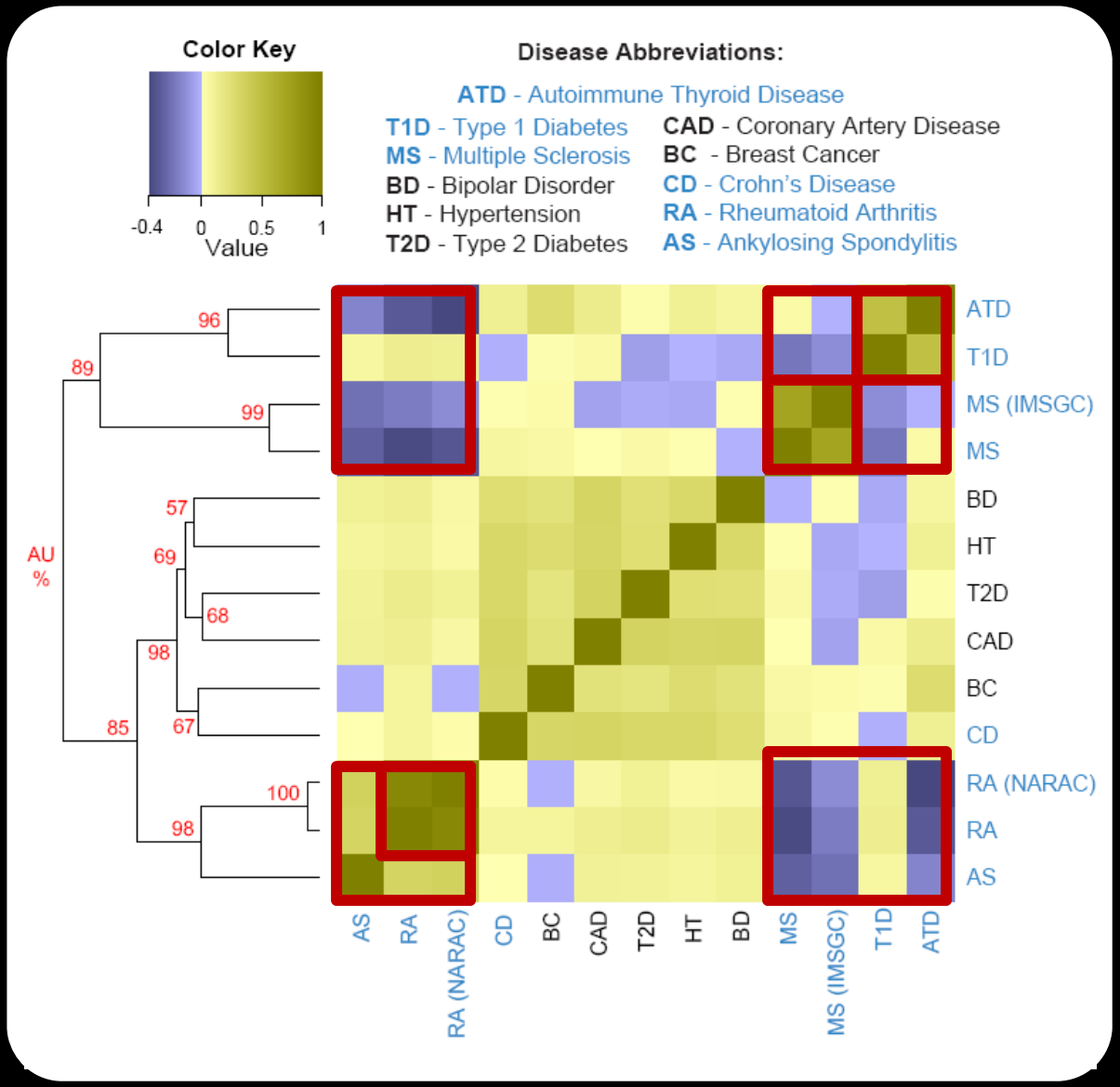
Pipeline



Cor (ATD, T1D)



Two Clusters of Autoimmune Disease



"Toggle" SNPs that Best Separate Autoimmune Clusters

SNP - Allele	Gene Symbol	Genetic Variation Score (GVS)						
		RA (NARAC)	RA	AS	T1D	ATD	MS (IMSGC)	MS
rs11752919 - C	ZSCAN23	-3.48	-3.21	-9.39	1.10	0.70	3.25	2.99
rs3130981 - A	CDSN	-0.46	-1.00	-9.47	-4.94	0.33	10.00	13.41
rs151719 - G	HLA-DMB	-6.71	-4.77	-1.08	-13.63	0.34	8.58	17.76
rs10484565 - T	TAP2	25.52	8.37	1.34	15.74	-1.36	-0.56	-0.30
rs1264303 - G	VARS2	11.51	7.36	18.76	0.89	-1.76	-1.85	-1.75
rs1265048 - C	CDSN	6.59	2.97	50.13	6.34	-0.85	-2.39	-4.16
rs2071286 - A	NOTCH4	5.30	0.78	6.42	4.04	-0.03	-1.89	-2.45
rs2076530 - G	BTNL2	67.49	56.46	14.06	13.58	-6.41	-9.50	-18.52
rs757262 - T	TRIM40	14.58	9.11	6.27	1.56	-0.79	-2.05	-7.34

Susceptible to Disease

Protective of Disease

Hypothesis

- Some SNPs predispose individuals to autoimmune disease in general

- Others specify which autoimmune class of disease one is more likely to have

Findings

- A novel approach to identify allele-specific relationships between disease pairs
- Compare genetic variation profiles of six autoimmune diseases and five non-autoimmune diseases
- Find two broad clusters of autoimmune disease where SNPs that make an individual susceptible to one class of autoimmune disease also protect from diseases in the other autoimmune class

Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ. Autoimmune disease classification by inverse association with SNP alleles. PLoS Genet. 2009.

Implications

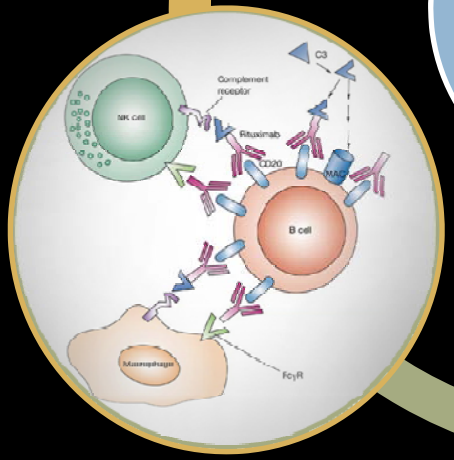
- May be used to illuminate potential new therapies
- Drugs useful for treating any particular disease may be useful for treating another disease in its class
- Learn more about disease mechanism by studying toggle SNPs?

Diagnostics

Autoimmune Disease Classification by Inverse Association with SNP Alleles

Discovery and Validation of Novel Drug Indications Using Gene Expression Data

Autoimmune Informatics



Disease Mechanism

Therapeutics

Acknowledgements

Atul Butte

Serafim Batzoglou

Bill Robinson

Pankaj Jay Pasricha

Biomedical Informatics

Joel Dudley

Annie Chiang

Alex A. Morgan

Reetesh Pai

Marc Schaub

Mohan Shenoy

Butte and Batzoglou Labs

National Library of Medicine

References

1. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007.
2. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ et al: Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007.
3. Schaub MA, Kaplow IM, Sirota M, Do CB, Butte AJ, Batzoglou S. A Classifier-based approach to identify genetic similarities between diseases. Bioinformatics. 2009.
4. Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ. Autoimmune disease classification by inverse association with SNP alleles. PLoS Genet. 2009.

Questions?

msirota@stanford.edu