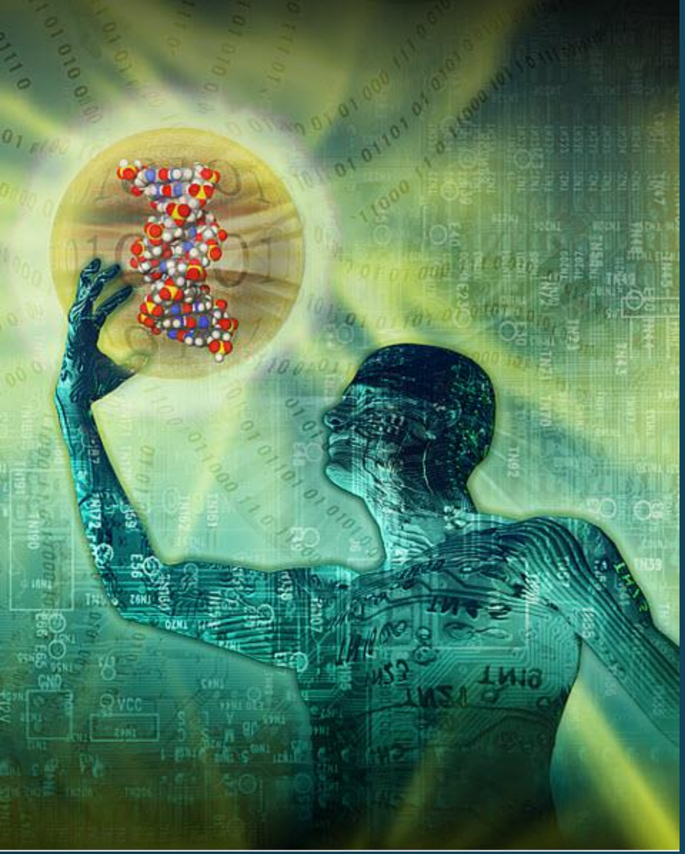


Genetic risk prediction for complex traits and its relationship to sub-phenotypes in vitiligo

Stephanie A. Santorico, Ph.D.

Dept of Mathematical and Statistical Sciences, Human Medical Genetics and Genomics Program, and Dept of Biostatistics & Informatics

University of Colorado Denver

Stephanie.Santorico@ucdenver.edu

Prediction of individual risk for the purpose of preventative intervention and/or selection of optimal treatments will be critical to precision medicine related initiatives and to the future of health care.

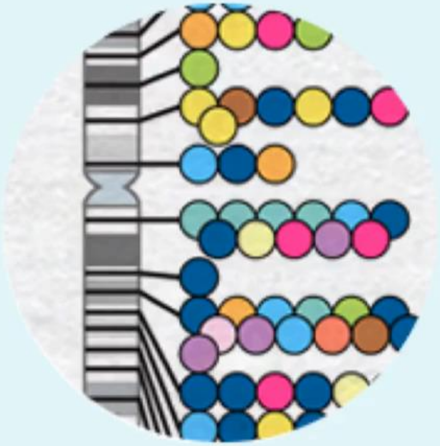
PMC PERSONALIZED MEDICINE COALITION

All of Us[SM]
THE FUTURE OF HEALTH BEGINS WITH YOU
The Precision Medicine Initiative®

European Alliance for Personalised Medicine

Beyond traditional risk factors: Genome-wide association study (GWAS)

GWAS Catalog



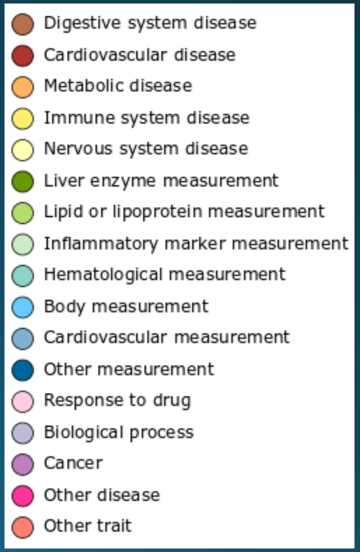
The NHGRI-EBI Catalog of published genome-wide association studies provides a publicly available curated resource of all published GWAS and association results

www.ebi.ac.uk/gwas

MacArthur J, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Research, 2017, Vol.45 (Database issue): D896-D901. PMID: 27899670

As of May 2018

- >5000 studies
- >69,000 unique SNP-trait associations
- From 3378 publications



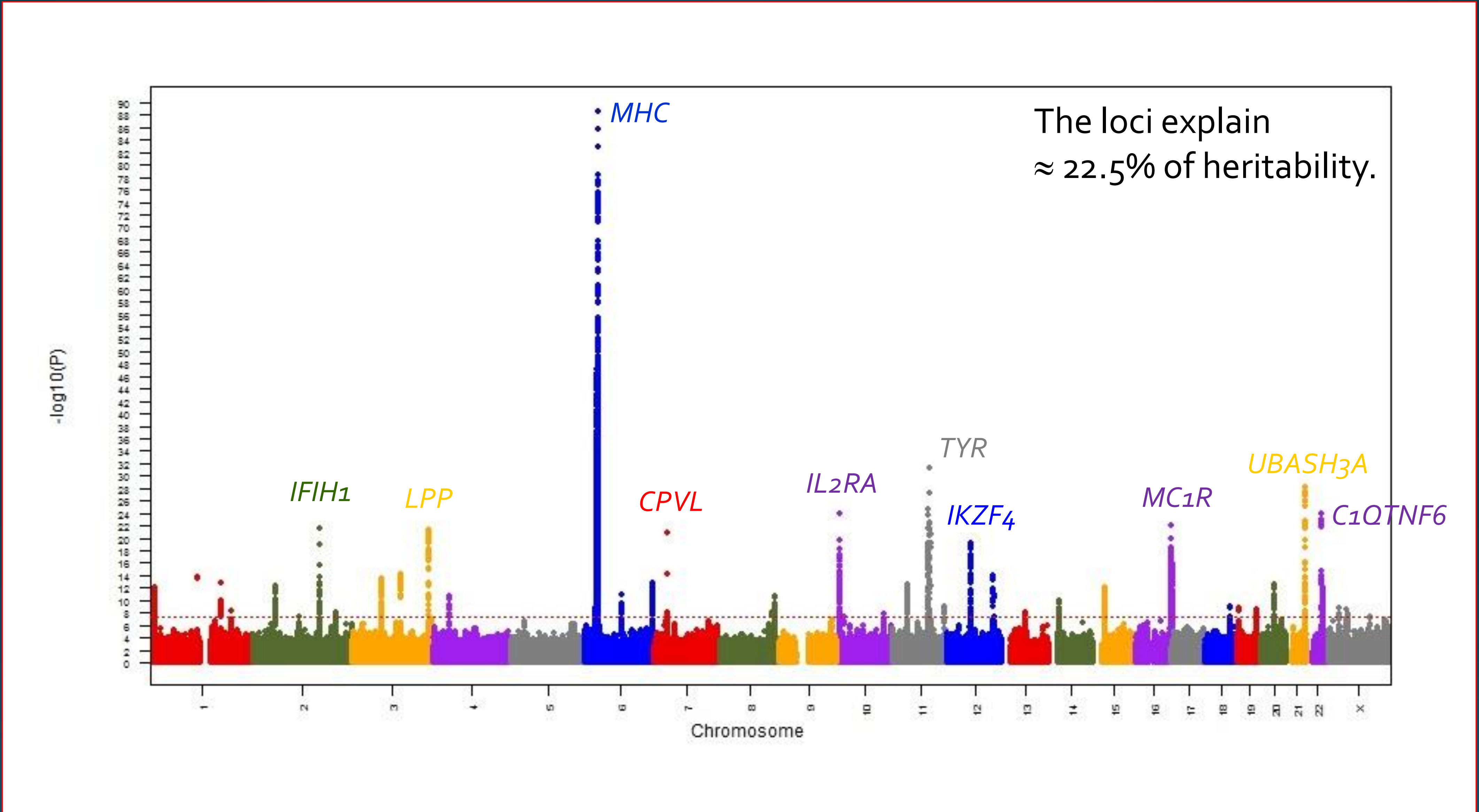
Generalized Vitiligo (GV)

- Complex autoimmune disease
- White patches of skin and hair resulting from destruction of melanocytes
- Prevalence varies from 0.001 to 0.02 with populations of European descent having a prevalence of ≈0.004[2]
- Heritability of ≈75%
- Sibling recurrence risk of ≈15 and AUC for single-sibling family history of 0.53

Our lab organized the International VitGene Consortium and completing three genome-wide association studies (GWAS) in European-derived whites (EUR).

2. *Alkhateeb A, et al. Pigment Cell Res (2003)*

48 loci genome-wide significant (p-value $< 5 \times 10^{-8}$) with replication[3]



3. Jing et al. Nature Genetics (2016)

Study design and methods

- *Subjects*:
  - 2,841 cases and 37,255 controls of European (EUR) descent
  - Collected over 3 GWAS stages
  - An independent replication set of 1,827 EUR vitiligo cases and 2,181 controls

- *Genetic data:*
  - Genotypes from Illumina Human OmniExpress Array
  - Genome-wide imputation to 1000 Genomes Project Phase I-August, 2012.
  - A total of 8,801,562 autosomal variants passing QC and present in all GWAS were used for analysis.
    - Genotyped SNP data and genotype posterior probabilities for variants with imputation INFO > 0.5 were used.
    - For each GWAS, variants were excluded based on MAF < 0.01, deviation from Hardy-Weinberg equilibrium ($P < 10^{-4}$) and SNP call rates >98%.
  - Genetic sub-structure for each GWAS was determined by Spectral-GEM.
  - Controls were matched to cases based on Spectral-GEM
  - Spectral-GEM principal components were selected for inclusion as covariates in subsequent analyses based on a family-wise error rate of 0.1 in logistic regression for association with case-control status.

- *Primary GWAS Method:*
  - Cochran-Mantel-Haenszel test using matched cases and controls.

GWAS gave us:

- 48 loci associated with vitiligo that replicated in an independent set of data
- Loci largely broke into categories involved in regulation of immune cells and apoptosis or encode melanocyte components or transcription factor with roles in both immune cells and melanocytes
- Loci were well connected in terms of biological networks
- Full results of study available in Jin et al. Nature Genetics. 2016

……how well do these loci predict disease risk? Can additional genetic variation improve upon risk prediction? Can risk be subdivided to lead to sub-types of disease?

Genetic Risk Scores

- For each variant, a logistic regression for disease is fit as a function of the number of the number of minor alleles giving a corresponding $\hat{\beta}_i$ for each variant

- Given a threshold, α_T, a risk score is built from all variants with a p-value $\leq \alpha_T$:

$$\hat{S} = \sum_{i=1}^{m} \hat{\beta}_i G_i$$

- A logistic regression is fit to model association of the risk score with disease:

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + X_{anc}\beta_{anc} + \beta_S \hat{S}$$

- Ten-fold cross-validation was used to fit and assess models based on area under the curve (AUC) for the receiver operator characteristic (ROC) curve.

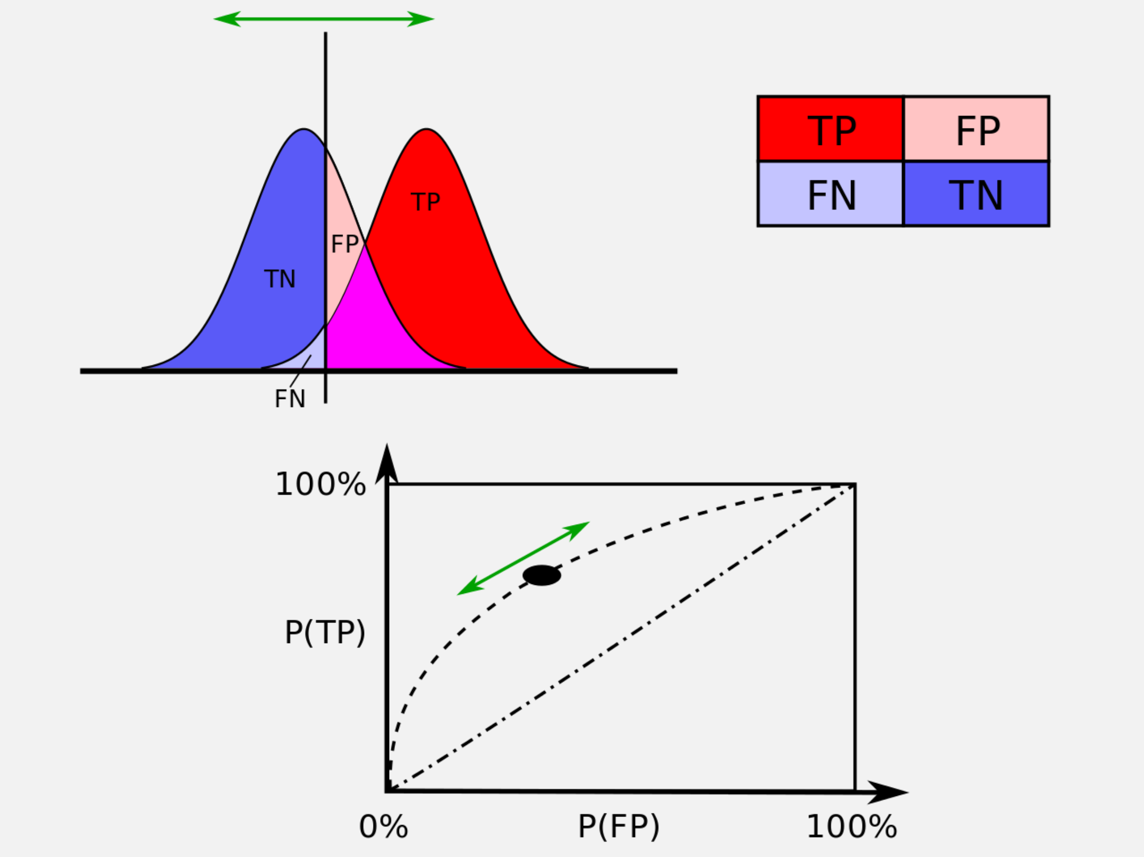
Risk prediction models considered

- ALL LOCI: Using all genotyped and imputed variants

- CLUMPED: An initial filtering of variants using clumping
  - Index SNPs with p-value < 0.1
  - Clumps formed by SNPs within 250kb with an $r^2 > 0.2$

- CONFIRMED: Using variants from confirmed major loci only

Comparison of Models by AUC

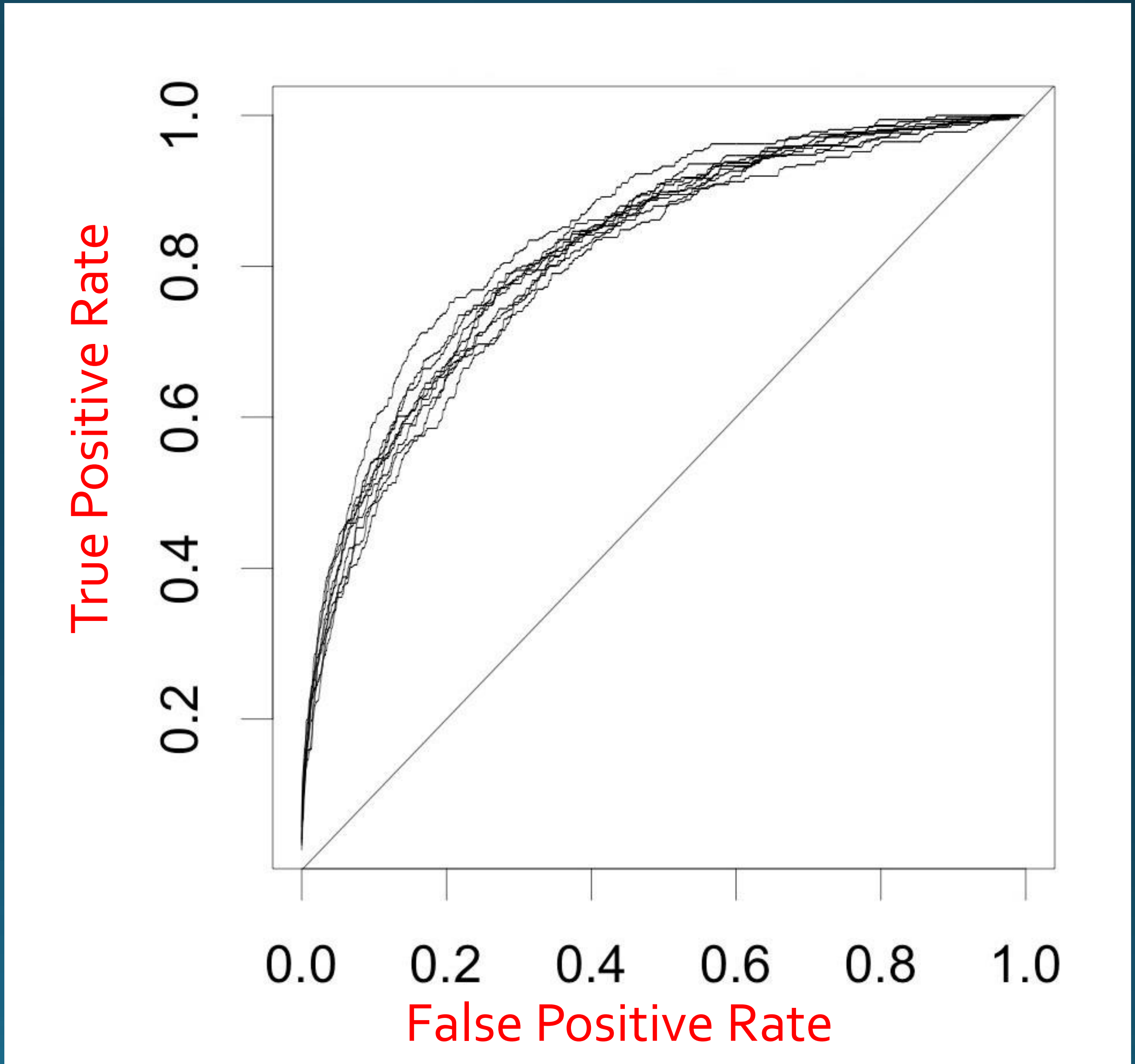


Receiver Oprator Characteristic (ROC) curves and Area Under the Curve (AUC)

ROC curves from 10 fold cross validation for the major locus risk score

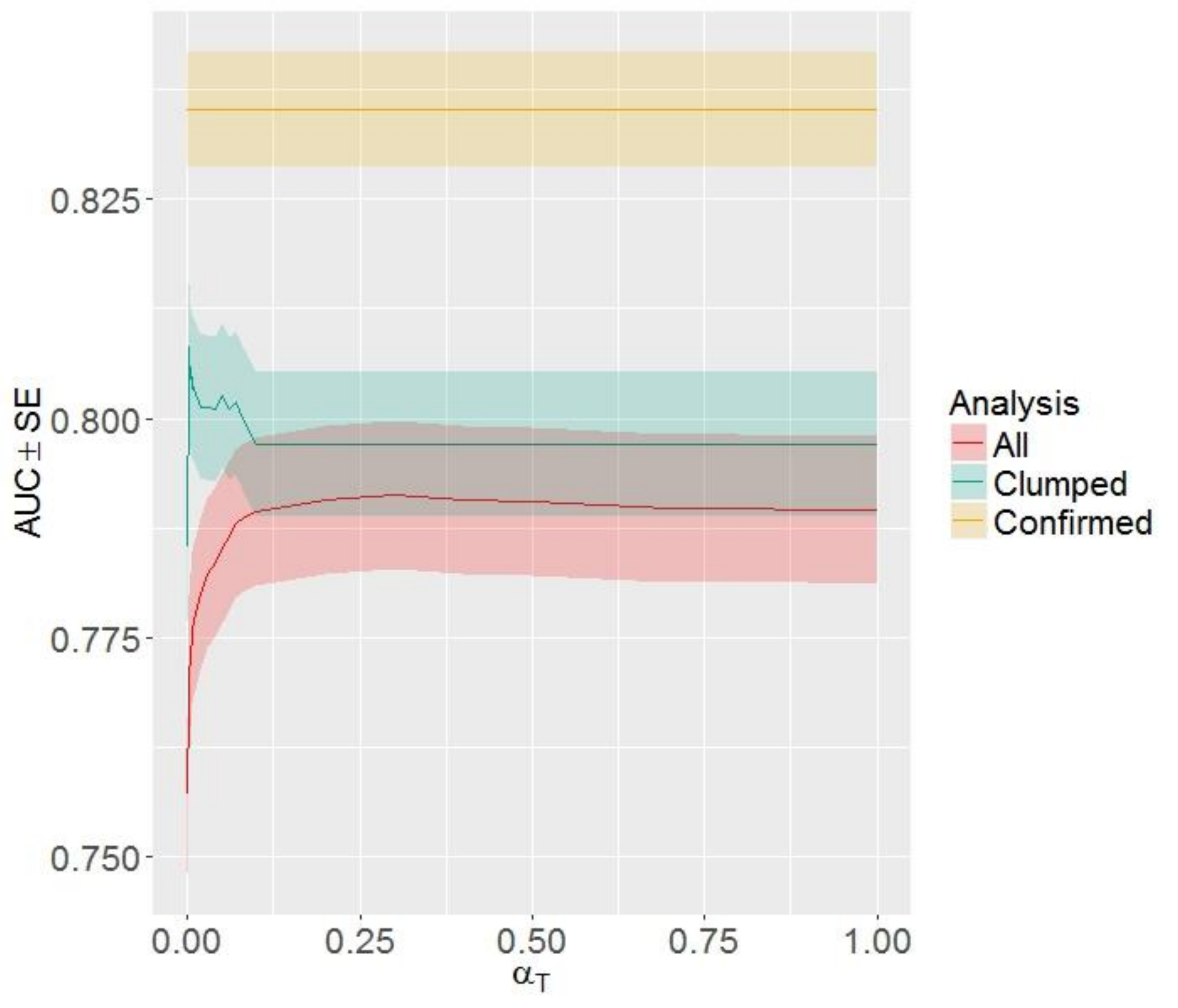
AUC ≈ 0.84

Recall: AUC for single-sibling family history = 0.53



Comparison of Models by AUC

CONFIRMED LOCI



AUC for genetic risk score as P-value threshold varies.

Three approaches are demonstrated based on use of all genotyped and imputed variants, a set of filtered variants using LD clumping, and restriction to a set of confirmed loci.

Clumping is based on an index P-value of 0.1 within 250kb from the index SNP with R2>0.2.

Secondary phenotypes

Phenotype	Proportion (out of N)
Halo Nevi	0.40 (336)
Koebner Phenomenon	0.68 (199)
Body Surface Involvement	
$\leq 25\%$	0.687 (2501)
$26 - 50\%$	0.175 (2501)
$51 - 75\%$	0.078 (2501)
$> 75\%$	0.060 (2501)
	Mean (SD, N)
Age of Onset in years	25.76 (16.6, 2749)

HN

KP

Secondary phenotypes

Phenotype	Proportion (out of N)	P-value for Association with CRLS
Halo Nevi	0.40 (336)	0.112
Koebner Phenomenon	0.68 (199)	0.004
Body Surface Involvement $\leq 25\%$ $26 - 50\%$ $51 - 75\%$ $> 75\%$	 0.687 (2501) 0.175 (2501) 0.078 (2501) 0.060 (2501)	 0.739
	Mean (SD, N)	
Age of Onset in years	25.76 (16.6, 2749)	1.9×10^{-8}

HN

KP

Loci largely break into categories

Role of Locus	Loci
Regulation of Immune Cells	PTPN22 (rs2476601), PTPRC (rs16843742), IFIH1 (rs2111485), CTLA4 (rs231725), FARP2–STK25 (rs41342147), UBE2E2 (rs35161626), FOXP1 (rs34346645), CD80–ADPRH (rs148136154), LPP (rs13076312), FBXO45–NRROS (rs6583331), HLA-A (rs60131261), HLA-DRB1–DQA1 (rs9271597), BACH2 (rs72928038), RNASET2–FGFR1OP–CCR6 (rs2247314), CPVL (rs117744081), TG–SLA–WISP1 (rs2687812), IL2RA (rs706779), ARID5B (rs71508903), CD44–SLC1A2 (rs1043101), IKZF4 (rs2017445), SH2B3–ATXN2 (rs10774624), TNFSF11 (rs35860234), TNFRSF11A (rs8083511), TICAM1 (rs4807000), SCAF1–IRF3–BCL2L12 (rs2304206), PTPN1 (rs6012953), UBASH3A (rs12482904), IL1RAPL1 (rs73456411), CCDC22–FOXP3–GAGE (rs5952553)
Regulators of apoptosis (particularly involving immune cells)	RERE(rs301807), FASLG (rs78037977), BCL2L11–MIR4435-2HG (rs4308124), SERPINB9 (rs78521699), NEK6 (rs10986311), CASP7 (rs12771452), PPP1R14B–PLCB3–BAD–GPR137–KCNK4–TEX40–ESRRA–TRMT112–PRDX5 (rs12421615), GZMB (rs8192917), C1QTNF6 (rs229527)
Encode or regulate melanocyte components	IRF4 (rs12203592), TYR(rs1126809), OCA2–HERC2 (rs1635168), MC1R (rs4268748), RALY–EIF252–ASIP–AHCY–ITCH (rs6059655)
Functions are either unknown or not obviously relevant to vitiligo or autoimmunity	PPP4R3B (rs10200159), PPP3CA (rs1031034), Gene desert (rs11021232), KAT2A–HSPB9–RAB5C (rs11079035), ZC3H7B–TEF (rs9611565)

Association of risk score with age of onset

Risk score	Estimate	SE	P-value
Regulation of Immune Cells	-1.46	0.31	3.0×10^{-6}
Regulation of Apoptosis	-0.61	0.31	0.053
Encode Melanocyte	-0.77	0.31	0.014
Unknown	-0.28	0.31	0.408

The risk score is associated with age of onset of vitiligo

	$\hat{\beta}\ (SE)$	P-value
HLA-A	-0.38 (0.66)	0.39
HLA-DRB1	-2.89 (0.44)	8.5×10^{-17}
Remaining risk score	-1.02 (0.31)	0.0011

As risk score increases, age of onset decreases.

Additional secondary phenotypes:

		P-value for Score				
Phenotype	Proportion (out of N)	**Full**	Immune	Apoptosis	Melanocyte	Unknown Function
Halo Nevi	0.40 (336)	**0.112**	0.158	0.955	**0.013**	0.514
Koebner Phenomenon	0.68 (199)	**0.004**	**0.009**	0.883	0.642	0.077
Body Surface Involvement $\leq 25\%$ $26-50\%$ $51-75\%$ $> 75\%$	0.687 (2501) 0.175 (2501) 0.078 (2501) 0.060 (2501)	**0.739**	0.829	0.122	0.581	0.947
	Mean (SD, N)	P-value for Score				
Age of Onset in years	25.76 (16.6, 2749)	$\mathbf{1.9 \times 10^{-8}}$	$\mathbf{2.9 \times 10^{-6}}$	0.053	**0.014**	0.408

That is,

- Addition of genetic factors improves prediction beyond that of sibling family history
- Additional genetic variation beyond the validated loci did not improve our ability to predict vitiligo risk
- Higher disease risk scores correlates with earlier age of onset
  - With this correlation coming from regulation of immune cells (primarily) and encoding of melanocytes (secondary)
- Genetic risk score was not associated with body surface area, implying other genetic factors influence this secondary phenotype

What next?

- While genetic risk prediction using confirmed loci improved upon family history, there is much more room for improvement.
  - We are currently exploring incorporation of heterogeneity.
  - Heterogeneity is being explored in the context of genetic association as well as through the use of sub-phenotype data.

- Results of the above work may improve prediction by taking into account heterogeneity and may lead to better grouping of patients for development of treatments

- Given the lack of association of the risk score with body surface involvement (and with any of the individual confirmed loci), we are pursuing a GWAS focused on body surface involvement.

Acknowledgements

Genotyping: Center for Inherited Disease Research (CIDR)

Janus supercomputer, which is supported by the National Science Foundation (award CNS-0821794), the University of Colorado Boulder, the University of Colorado Denver, and the National Center for Atmospheric Research.

This work was supported by grants R01AR045584, R01AR056292, X01HG007484, and P30AR057212 from the US National Institutes of Health and by institutional research funding IUT20-46 from the Estonian Ministry of Education and Research.

CU Denver

- Richard Spritz, M.D.

- Ying Jin, M.D.

- Subrata Paul, M.S.

- Tracey Ferrara, Ph.D.

- Paulene Holland

- Janet Siebert, M.S.

- Gen Andersen, B.S.

Purdue Fort Wayne

- Daniel Yorgov, Ph.D.