OPEN ACCESS Freely available online

Nov, 2010

PLoS one

Genome-Wide Analysis of Copy Number Variation in Type 1 Diabetes

Britney L. Grayson[1], Mary Ellen Smith[2], James W. Thomas[1,2], Lily Wang[3], Phil Dexheimer[4¤a], Joy Jeffrey[5], Pamela R. Fain[5], Priyaanka Nanduri[5¤b], George S. Eisenbarth[5], Thomas M. Aune[1,2]*

1 Department of Microbiology and Immunology, School of Medicine, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Department of Medicine, Division of Rheumatology, School of Medicine, Vanderbilt University, Nashville, Tennessee, United States of America, **3** Department of Biostatistics, School of Medicine, Vanderbilt University, Nashville, Tennessee, United States of America, **4** Functional Genomics Shared Resource, School of Medicine, Vanderbilt University, Nashville, Tennessee, United States of America, **5** Barbara Davis Center for Childhood Diabetes, University of Colorado, Aurora, Colorado, United States of America

Gerstein Lab Journal Club

March 2nd, 2012

Gili Zilberman

Outline

- Biological Background
  - Type 1 Diabetes
  - Variations
  - Monozygotic twins
- Hypothesis & Goal
- Method & Analysis
- Experiment Overview
- Results
- Discussion

Background: Type 1 Diabetes (T1D, formerly juvenile diabetes)

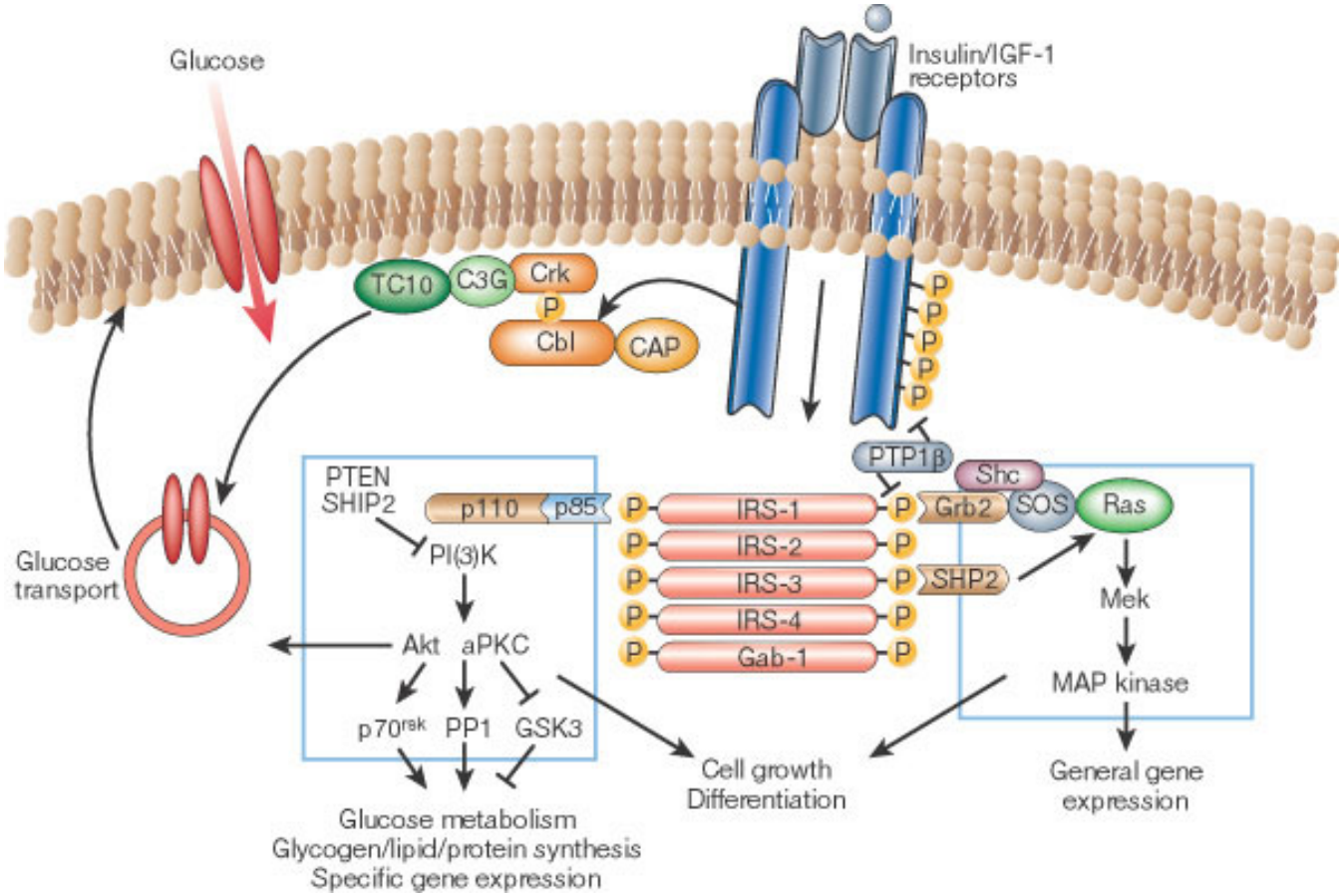
- Can occur at any age, but most often diagnosed in children, adolescents, and young adults.
- 5–10% of all diabetes cases, 11–22 million worldwide.
- Results from immune-mediated selective destruction of pancreatic islet (or beta) cells.
- Causes insulin deficiency and hyperglycemia.
- Symptoms include polydipsia, polyuria, polyphagia and weight loss, when significant numbers of islet cells have been destroyed.

Background: Insulin & Glucose

- **Insulin** is produced by beta cells in the pancreas.
- Facilitates **glucose** transport into cells, where it is stored and later used for energy.
- **Glucose** builds up in the bloodstream and is unavailable for use.
- Exact **causes of T1D** are unknown.
  - Most likely an autoimmune disorder.
  - A trigger (infection) causes the body to create autoantibodies.
  - Genetic susceptibility

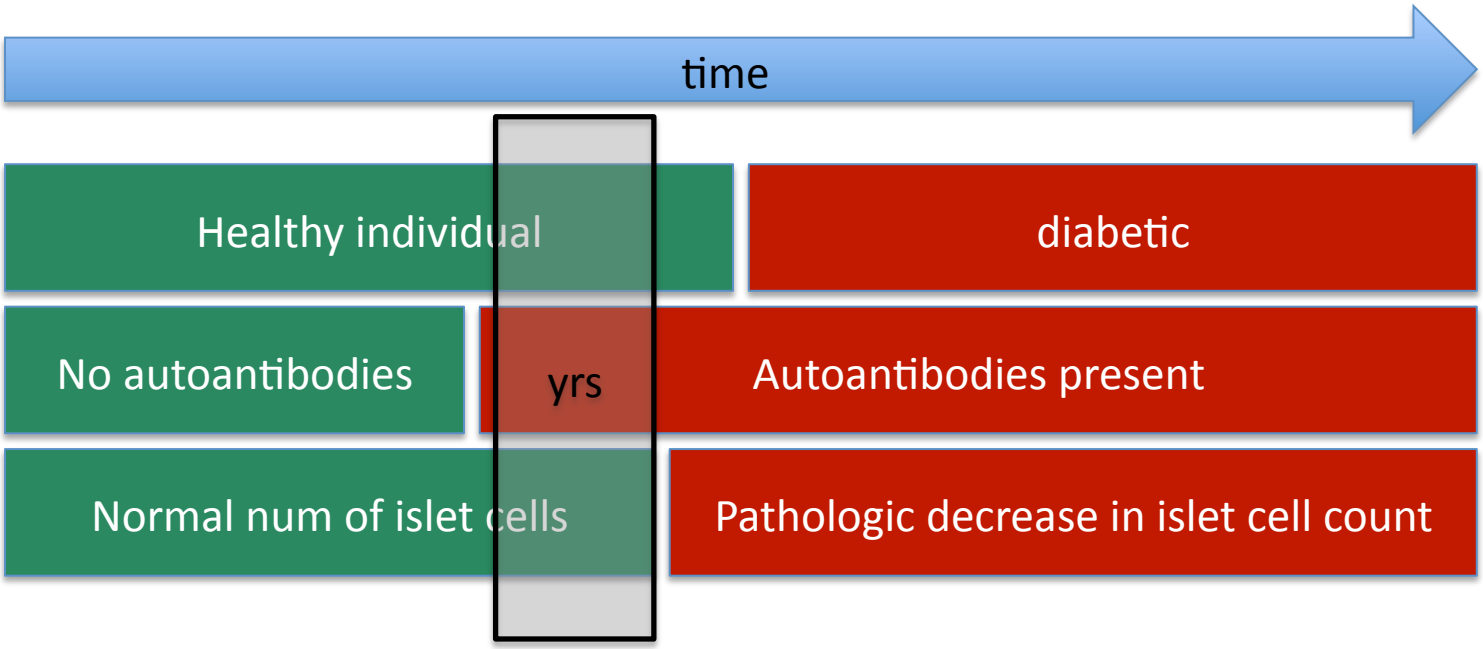
Signal Transduction in Insulin Action

1. **Receptor auto-phosphorylation**
2. **Signaling molecules phosphorylation**
3. **Translocation to the plasma membrane**



Saltiel *et al.,* Nature, 2001

T1D Disease Progression



Starting Point & Previous Studies (1)

- Pancreas transplantation
  - Invasive, immunosuppressive drugs
- Islet cell transplantation
  - Into the patient's liver, produce insulin
  - Immunosuppressive drugs (unless self engineered)
  - Pig islets within a protective capsule (Living Cells Technologies Limited)
- Genetically engineered insulin producing cells
  - No reaction to external signal

Starting Point & Previous Studies (2)

- More than 90% of type 1 diabetics carry HLA alleles DR3- DQ2 or DR4-DQ8 compared to no more than 40% of the general population.

- Recent genome-wide association study of ~2,000 patients: association of 10 SNPs (5 novel) with T1D.

- However, these regions (except HLA on chromosome 6) confer only modest effects on T1D, and “account for only a small proportion of overall familiality”.

Background: Variations

Copy-number variant (CNV):

A segment of DNA that is 1 kb or larger and is present at a variable copy number in comparison with a reference genome.

Includes: insertions, deletions and duplications.

Copy-number polymorphism (CNP):

A CNV that occurs in more than 1% of the population.

Feuk *et al.*, Nature Genetics, 2006

Background: Monozygotic (MZ) Twins

- “Identical” twins, developed from one fertilized oocyte.
- Monozygotic twins do not have identical genetic sequences and are known to vary in CNV.
- 27.3% concordance for T1D in MZ twins, while only a 3.8% in dizygotic twins.

Hypothesis & Goal

Hypothesis: CNVs contribute to susceptibility to and/or protection from T1D.

Goal: Identify CNVs either enriched or depleted in T1D and in the disease onset.

Assumption: In disease discordant monozygotic twins, if a CNV were associated with a disease, the twin affected would have the variant, while the unaffected twin would not.

Method & Analysis

Methods: Genome-wide analysis on 2 cohorts:

- 20 unrelated adults diagnosed with T1D and 20 unrelated control subjects.
  - Identify CNVs that distinguish T1D patients from controls.
- 10 MZ twin pairs disease discordant for T1D.
  - Validate the frequencies found.

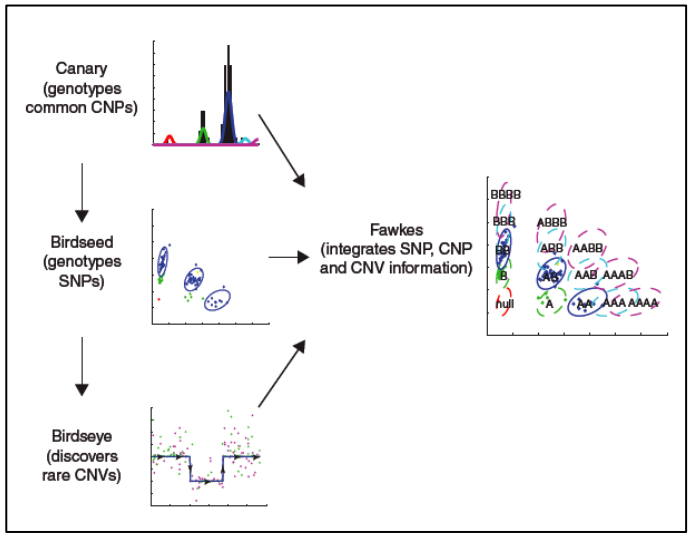
Analysis:

- Affimetrix SNP array 6.0 and BirdSuite
- qPCA and CellCaller 1.0 (validation)

Overview of BirdSuite (1)

- A software for deriving integrated copy number and SNP genotypes

- **Output**: copy number values across the chromosome with a confidence score for each individual call.



Korn *et al.*, Nature genetics, 2008

Overview of BirdSuite (2)

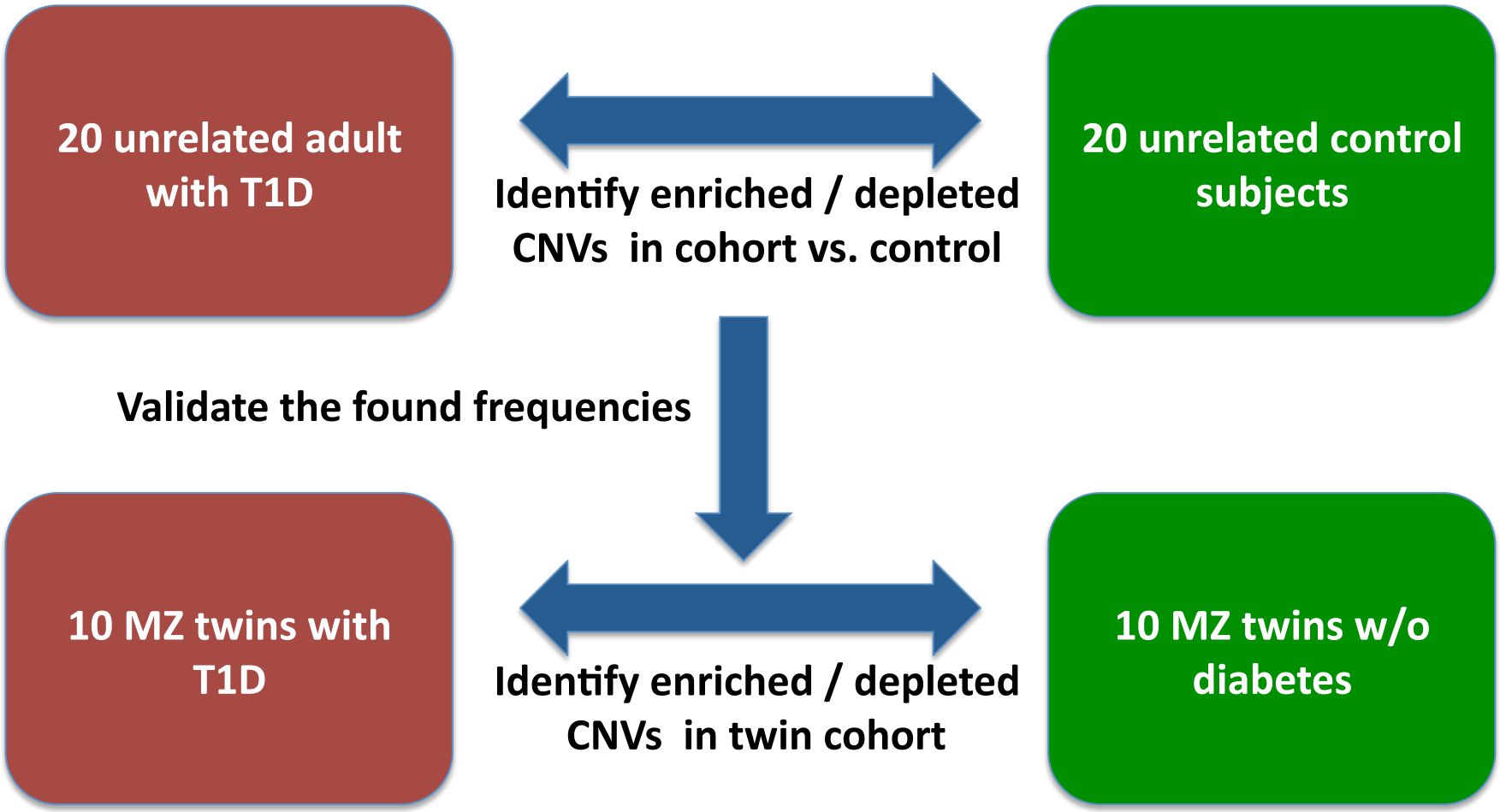
1. Assigns copy number across regions of known common CNPs (based on HapMap).

2. At each SNP locus, samples expected to have two copies of the locus and are assigned genotypes: AA, AB or BB (based on expected allele intensity).

3. Informed by probe-specific mean and variance (step 2), a HMM is used to discover rare or de novo CNVs.

4. Copy number and SNP allele information are combined to provide an integrated genotype at each locus.

Experiment Overview



Thresholds

- Rare diseases: CNV found 70% of patients.
- More common diseases (T1D): expect ~40%.
- **Significance conditions**: CNVs that were present in **more than 40%** of the T1D (control) group at a **1.5 fold change** greater frequency as compared to the Ctrl (T1D) group were determined to be enriched (depleted).

Results: T1D vs. Control

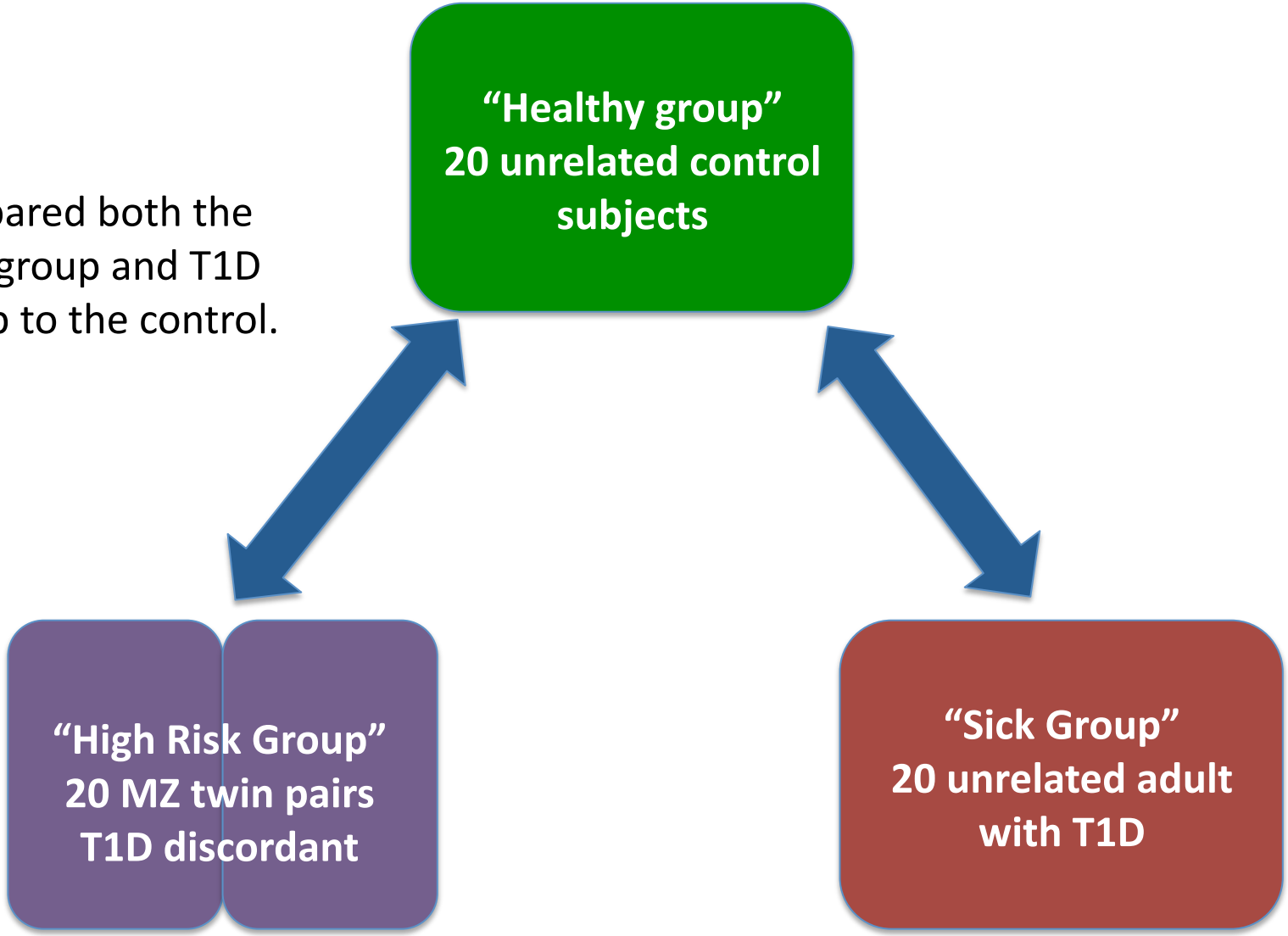
- 18 CNPs **enriched** in T1D vs. control.
  - > 40% & 1.5 fold greater frequency in T1D.
- 20 CNPs and 1 novel CNV **depleted** in T1D vs. control
  - > 40% & 1.5 fold greater frequency in control.
- **Total**: 39 CNVs

Results: MZ twins

- CNVs present in only one of the twins in a pair were grouped based on disease status.
- **No overlaps** with the 39 CNVs in the T1D vs. control found.
- No CNVs were present in more than 2 twin pairs
  - No trend.
- However, CNVs may be enriched in this group as a whole:
  - Unaffected twins have 75% incidence of developing autoantibodies.
  - 65% will progress to T1D.

Experiment Overview (plan B)

compared both the twin group and T1D group to the control.



Results: MZ *vs.* Control

- Same criteria: >40%, 1.5 fold change.
- 49 CNPs were enriched in twin cohort.
- 23 CNPs were depleted in twin cohort.
  - 1 novel CNV
- Total of 72 CNVs.

Results: MZ and T1D *vs.* Control

- The 72 CNVs from the twin cohort were compared with the 39 CVNs in the T1D cohort.
- 10 were enriched in both.
- 11 were depleted in both.
- Later: **select** CNVs > 1,000 bp, identified by >=3 consecutive probes.
- Total of **9 CNVs** selected.

Statistical Significance

Permutation test (repeated 1000 times)**:**

- keeping the number of patients fixed in each of the three groups, randomly permute group status for the samples.
- re-calculated the number of enriched CNVs in both disease groups relative to Ctrl.

$$p.val = \frac{\sum all\_permutations\_with\_overlap \geq 10}{\sum all\_permutations} = 0.005$$

Enriched CNVs, > 1,000 bp

Table 1. CNVs enriched in T1D and Twin cohorts, relative to Ctrl.

CNP ID[a]	Chr	Start	End	Amplification or Deletion	Ctrl%	T1D%	Ctrl: T1D p[b]	Twin%	Ctrl: Twin p	Sequence
253	2p11	87,600,933	87,609,093	deletion	42	72	0.13	70	0.15	NCRNA00152
934	6p21	32,700,999	32,710,085	deletion	42	89	0.01	65	0.26	CNS[c]
1162	7q33	133,435,735	133,449,694	deletion	37	78	0.03	80	0.02	CNS
1303	8q11	51,194,577	51,195,974	deletion	21	61	0.03	50	0.12	SNTG1
1956	13q21	71,375,556	71,378,557	deletion	58	89	0.08	95	0.02	-

[a]CNP ID as defined in McCarroll, et al. Nature Gen 40(10):1166-74.
[b]p-value derived from chi-square analysis.
[c]CNS= Conserved Noncoding Sequence, defined as a region >100bp with at least 70% similarity to sequence in mus musculus (as determined by ECR browser, ecrbrowser.dcode.org).
doi:10.1371/journal.pone.0015393.t001

- Length 1,400-14,000 bp.
- Frequency in twin and T1D cohorts 50%-95% (21%-58% in control)
- Each sequence encodes a potential TF binding site. Regulator function?

Depleted CNVs, > 1,000 bp

Table 2. CNVs depleted in T1D and Twin cohorts, relative to Ctrl.

CNP ID[a]	Chr	Start	End	Amplification or Deletion	Ctrl%	T1D%	Ctrl:T1D *p*[b]	Twin%	Ctrl:Twin *p*	Sequence
1102	7q11	66,266,764	66,282,667	deletion	68	39	0.14	10	0.001	TYW1
1879	12q23	98,319,424	98,322,865	deletion	47	22	0.20	10	0.02	ANKS1B
A588[c]	15q11	18,491,920	19,803,369	both	58	33	0.24	10	0.004	BCL8, POTEB, GOLGA6L6, GOLGA8C
2240	17p12	15,483,886	15,487,515	deletion	42	22	0.35	0	0.004	TRIM16

[a]CNP ID as defined in McCarroll, et al. Nature Gen 40(10):1166-74.
[b]*p*-value derived from chi-squared analysis.
[c]A588 is a novel variant identified in this study.
doi:10.1371/journal.pone.0015393.t002

- Length 3,400-15,900 bp.
- Absence frequency in twin and T1D cohorts 0%-39% (42%-68% in control)
- All copy number deletions are 0 or 1.

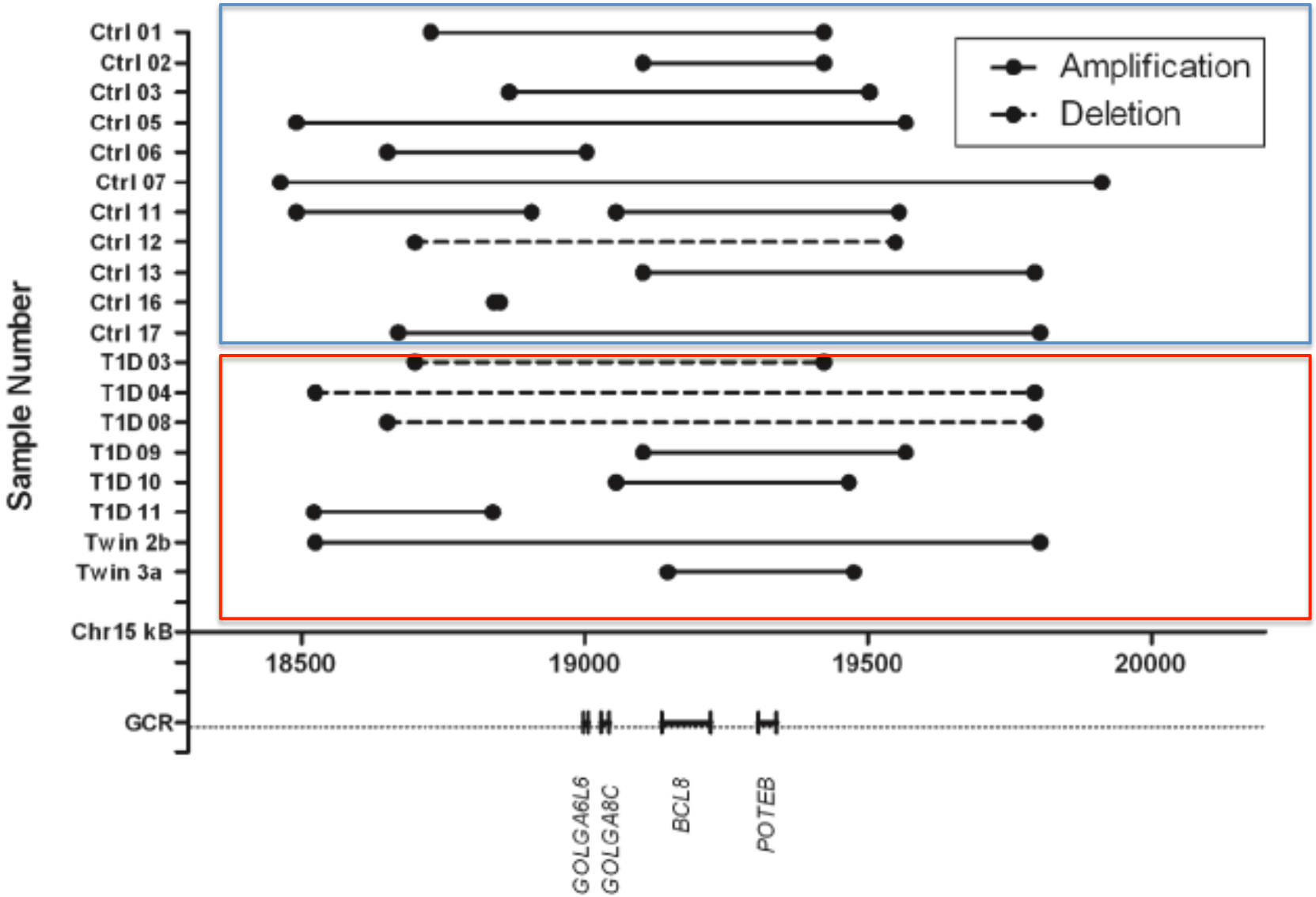
A588: A Novel CNV (1)

- Depleted in both T1D and Twin cohorts
- **Frequency**: 58% in control, 33% in T1D, 10% in Twin.
- Spans more than 1.3 million base pairs.
- Manifests as both an amplification and deletion

A588: A Novel CNV (2)

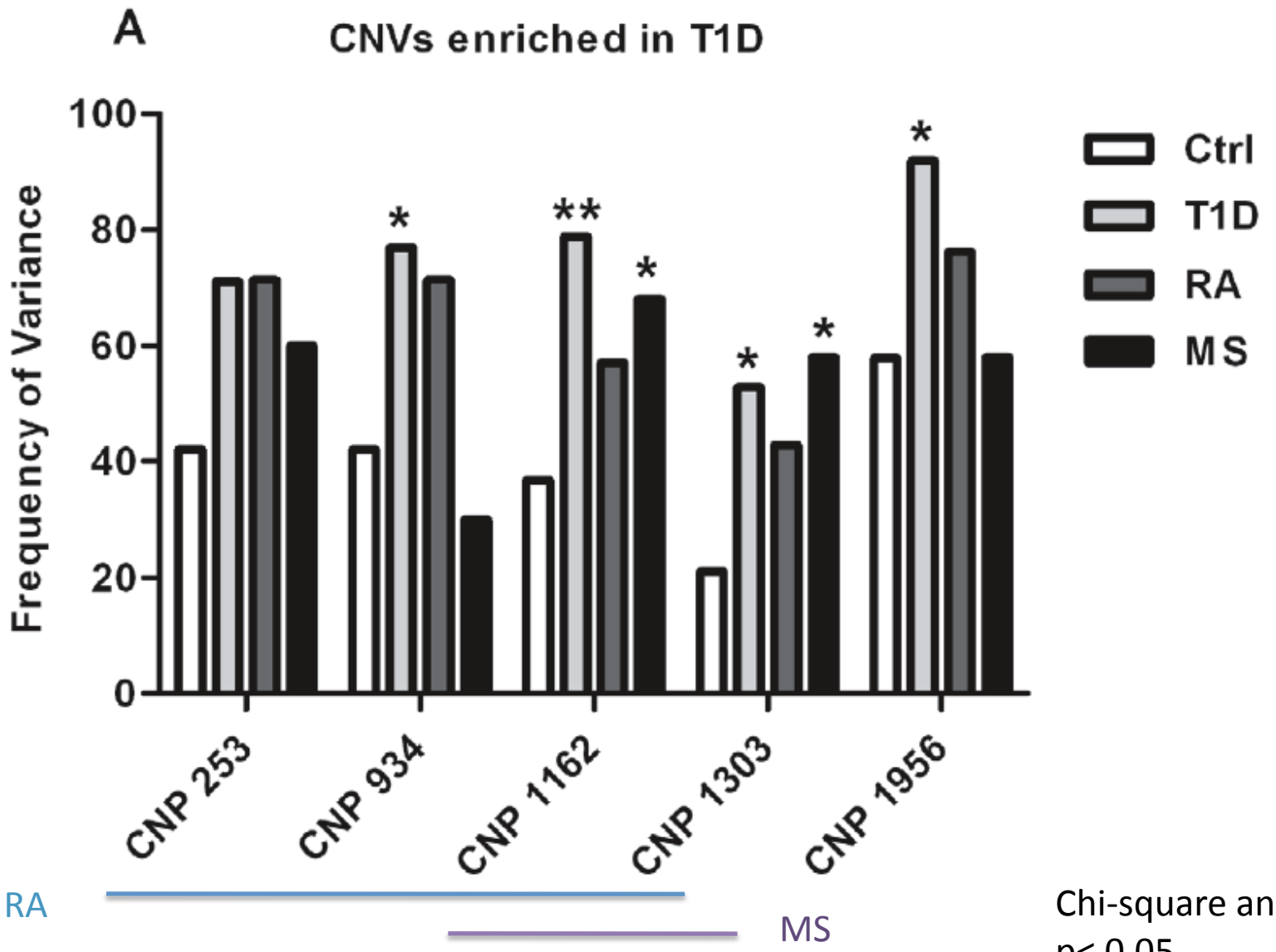
- Contains coding regions for genes like the golgin family (glycosylation and transport of proteins and lipids) B cell CLL/ Lymphoma gene BCL8.
- Many of the variants do not span the entire region.
  - Rather, overlapping and non-overlapping variants.
- Variants were grouped together as one singular CNV region (CNVR).
  - A single variant can impact regulation and expression of a gene more than 1 megabase away.

A588: A Novel CNV



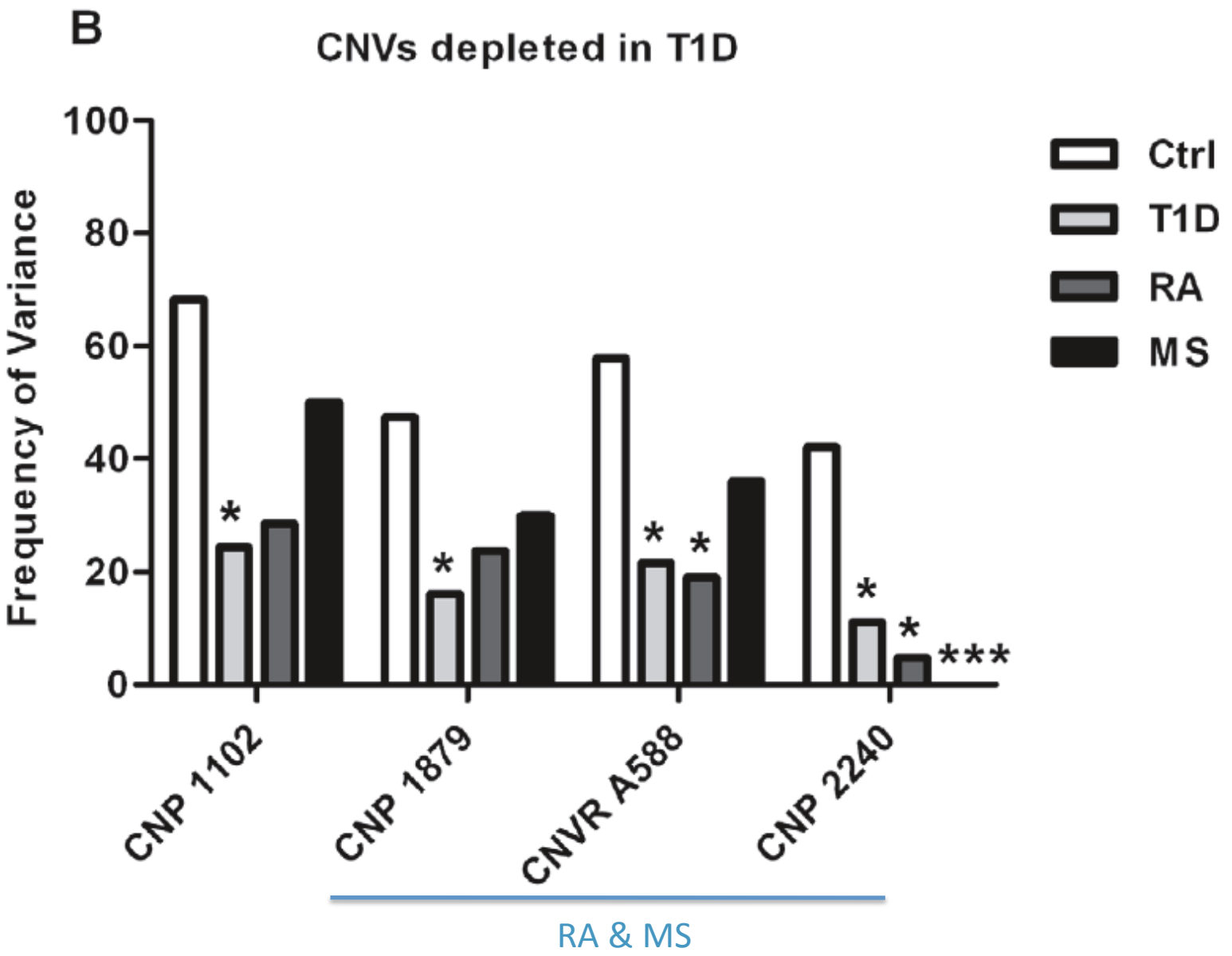
Strengthen The Findings With Other Autoimmune diseases

- The CNVs are potentially autoimmune related.
- Assess their frequencies in rheumatic arthritis (RA) and multiple sclerosis (MS) patients.
- RA- a chronic, systemic inflammatory disorder that principally attacks flexible joints.
- MS - the fatty myelin sheaths around the axons of the brain and spinal cord are attacked, leading to demyelination and scarring.



Meet the 40%, 1.5x criteria

Chi-square analysis
$p < 0.05$
$^{**}p<0.005$



2 significant in RA, 1 in MS

Discussion (1)

- 9 CNVs enriched or depleted in 2 independent cohorts T1D & high risk vs. control.
  - Genes & evolutionary conserved non-coding regions
- 5 CNVs were found to be enriched in patients with T1D and unaffected high risk twins.
  - Involved in the formation of autoantibodies /autoimmunity?
- 4 CNVs were less likely to be variant in the T1D and twin relative to the control group.

Discussion (2)

- CNVR A588 at a CN 2 is more common in patients with T1D.
  - Determine expression levels of the genes encoded in this region.
- 2 CNPs were found enriched in the diabetes cohorts and in RA. → Susceptibility to peripheral non-neurologic autoimmunity.
- 2 CNPs enriched in RA and MS → general autoimmune process.
- 2 CNPs and 1 novel CNPR depleted in all 3 autoimmune diseases.

Questions?