PGC Worldwide Lab Call Details

DATE: Friday, June 21st, 2013
PRESENTER: Pamela Sklar, Mount Sinai School of Medicine
TITLE: "Psych Chip Overview"
START: We will begin promptly on the hour.

1000 EDT - US East Coast
0700 PDT - US West Coast
1500 BST - UK
1600 CET - Central Europe
0000 AEDT – Australia (Saturday, June 22nd, 2013)

DURATION: 1 hour

TELEPHONE:

- US Toll free: 1 866 515.2912
- International direct: +1 617 399.5126
- Toll-free number? See http://www.btconferencing.com/globalaccess/?bid=75_public
- Operators will be on standby to assist with technical issues. "*0" will get you assistance.
- This conference line can handle up to 300 participants.

PASSCODE: 275 694 38

Lines are Muted **NOW**

Lines have been automatically muted by operators as it is possible for just one person to ruin the call for everyone due to background noise, electronic feedback, crying children, wind, typing, etc.

Operators announce callers one at a time during question and answer sessions.

Dial *1 if you would like to ask a question of the presenter. Presenter will respond to calls as time allows.

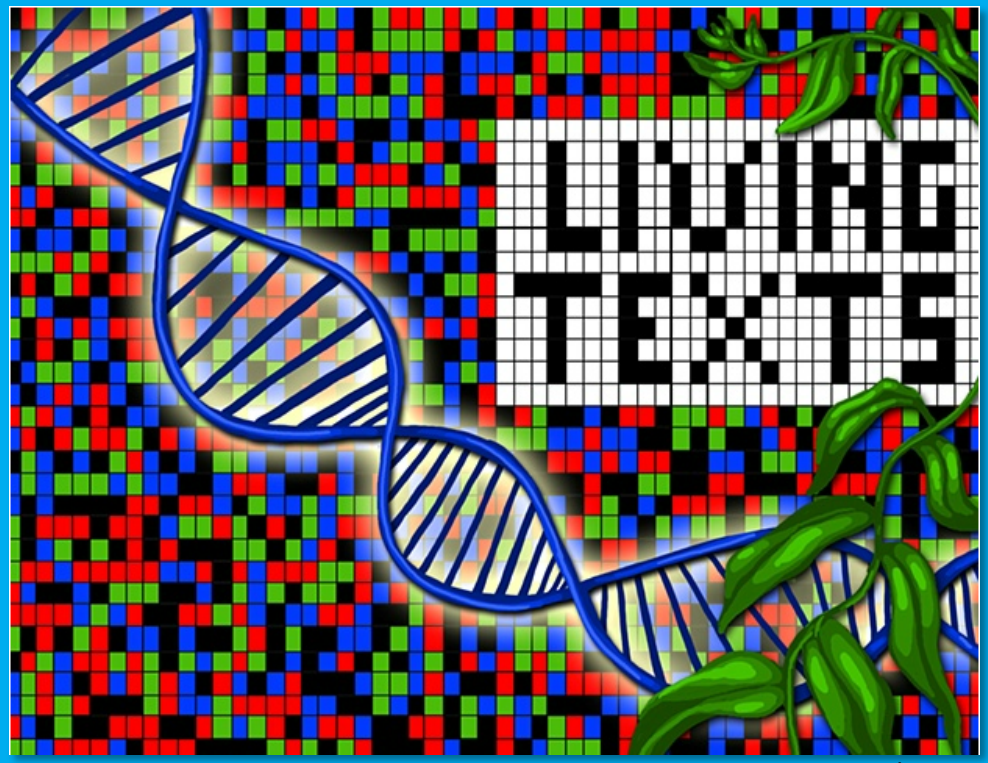
Dial *0 if you need operator assistance at any time during the duration of the call.

UPCOMING PGC Worldwide Lab

DATE: Friday, August 9th, 2013
PRESENTER: To Be Announced
TITLE: To Be Announced
START: We will begin promptly on the hour.
1000 EDT - US East Coast
0700 PDT - US West Coast
1500 BST - UK
1600 CEST - Central Europe
0000 AEST – Australia (Saturday, August 10th, 2013)
DURATION: 1 hour

TELEPHONE:
- US Toll free: 1 866 515.2912
- International direct: +1 617 399.5126
- Toll-free number? See http://www.btconferencing.com/globalaccess/?bid=75_public
- Operators will be on standby to assist with technical issues. "*0" will get you assistance.
- This conference line can handle up to 300 participants.

PASSCODE: 275 694 38



Jessica Johnson

PsychCHIP Overview

PGC Worldwide Lab Meeting
June 2013

Pamela Sklar
Chief, Division of Psychiatric Genomics
Friedman Brain Institute
Institute for Genomics and Multiscale Biology

Mount Sinai

PsychChip Design Committee

Ben Neale
Patrick Sullivan
Stephan Ripke
Jackie Goldstein
Kaitlin Samocha
Doug Ruderfer
Mark Daly
Shaun Purcell
Pamela Sklar

PGC successes

- Publications in 4 disease areas and the CDG.
  - ADHD; J Am Acad Child Adolesc Psychiatry. 2010 Sep;49(9):884-97
  - SCZ; Nat Genet. 2011 Sep 18;43(10):969-76
  - BD; Nat Genet. 2011 Sep 18;43(10):977-83
  - MDD; Mol Psychiatry. 2013 Apr;18(4):497-511
  - CDG; Lancet. 2013 Apr 20;381(9875):1371-9

- Wave 2 analyses with much larger samples

Evolution of PGC1 to PGC2

- 2007 Established
- 2009-2010 Initial meta-analytic efforts
- PGC1 funded
- 2010-2011 Major publications of genome-wide significant loci
- PGC2 funded
- 2013
  - Added OCD/TD, PTSD, anorexia
  - Wave 2 analyses completed and PsychChip genotyping to begin

Current state of affairs

- **SCZ GWAS with 108 GWS loci**
  - Sample size matters
  - Common variants matter
- **Other disease areas** ramping up for similar analyses
  - Sample sizes are often limiting
- **Many exome studies** ongoing with results in flight, early results suggest
  - No goldilocks alleles
  - Sample size matters
- **NEEDED:**
  - Cost efficient method for extensive GWAS of rare and common variants with additional focus on Psych relevant markers not well-covered by common chip types

Goal of the PsychChip

- Integrate and prioritize results
  - CNV group
  - Each individual disease area
  - Next-generation sequencing
  - Cross-disorder analyses
- Genotype in >100,000 samples
  - NIMH funding for 45,000 samples
  - Philanthropic funding for ~30,000
  - Additional supplemental funding pending

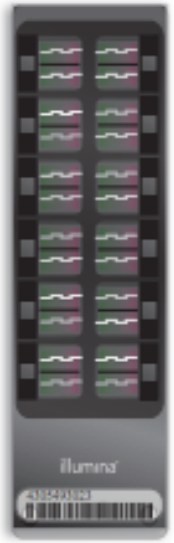
Custom Chips in other disorders

- Immunochip – 200K chip
  - New loci identified across multiple related diseases including ankylosing spondylitis, atopic dermatitis, rheumatoid arthritis, juvenile idiopathic arthritis, narcolepsy, psoriasis, biliary cirrhosis

- Metabochip – 200K chip

PsychChip Platform

Figure 1: HumanCoreExome BeadChip



The HumanCoreExome BeadChip enables genotyping of markers informative across diverse world populations, including exome-focused markers, delivering high-quality data that can be used for a variety of downstream applications.

Feature	Description
Number of tagSNP markers	> 240,000
Number of exome-focused markers	> 240,000
Headroom for additional custom markers	50,000
Number of samples	12
DNA requirement	200 ng
Assay	Infinium HD
Instrument support	HiScanSQ, HiScan or iScan
Sample throughput**	1,400 samples / week
Scan time / Sample	5 minutes

GWAS selected as random grid
Exome in loss of function variants from 16K individuals in additional to basic exome content

Chip and reagent cost for all PGC members = $45 US

Performance statistics

% Variation Captured[†] ($r^2 > 0.8$)	1kGP[†] MAF > 5%	1kGP[†] MAF > 1%
CEU	0.59	0.45
CHB + JPT	0.62	0.51
YRI	0.27	0.17

Data Performance	Value[‡] / Product Specification
Call frequency	99.9% / > 99.9% avg.
Reproducibility	99.9% / > 99.9%
Log R deviation	0.17 / < 0.30[§]

Spacing	Mean
Spacing (kb)	1 marker / 5.5 kb

What's missing?

Augmenting coverage of common SNPs in psychiatrically relevant regions given limited GWAS backbone

Augmenting coverage across rare CNVs, and detection of smaller CNVs

Ability to focus on psychiatrically relevant very rare variants not on exome chip

Diseases Contributing to Design ~85,000 cases

- Disease Groups: SCZ*, BD*, AUT*, ADHD, MDD*, ANOREXIA*, OCD/TS, CDG*, CNV

  * enhanced samples over PGC1 analyses

	Cases	Controls
SCZ	35,476	46,839
BD	15000	28,088
AUT	6,000	-
ADHD	6,000	-
MDD	15,000	25,000
Anorexia	3,000	

PsychChip Content – 50K

- **Common Variant (GWAS) content** – 25,000
  - SCZ, AUT, BIP, ADHD, MDD, CDG,
  - Pairwise tagging P value < 10^{-4}
- **CNV content** – 10,000 (387 CNVs)
  - P value <0.001 + 25 kb flanking region
  - Average probe spacing of 3kb
- **Special Requests** – 5,000
  - Smaller groups, anorexia, OCD/TD
  - MHC, sex, ancestry if needed
  - Family sequencing studies
- **Rare Variant Content** – 10,000
  - Exome sequencing studies – ~5000
  - Genotyping by sequencing – ~5,000

High resolution of topmost signals

- Strategy for GWAS selection
  - a) Identify SNPs above a given P-value threshold
  - b) Identifying all SNPs with a r^2 of > 0.6 with the SNPs in group a with P < 0.01
  - c) Clump [which means to select SNPs preferentially on P-value] at an r^2threshold of 0.9
- Potential refinement for strongly associated regions
  - LD filtering at r^2 with upper bound at 0.95 and lower bound at 0.6
- Improve the imputation in associated regions
- May aid in the identification of causal SNPs

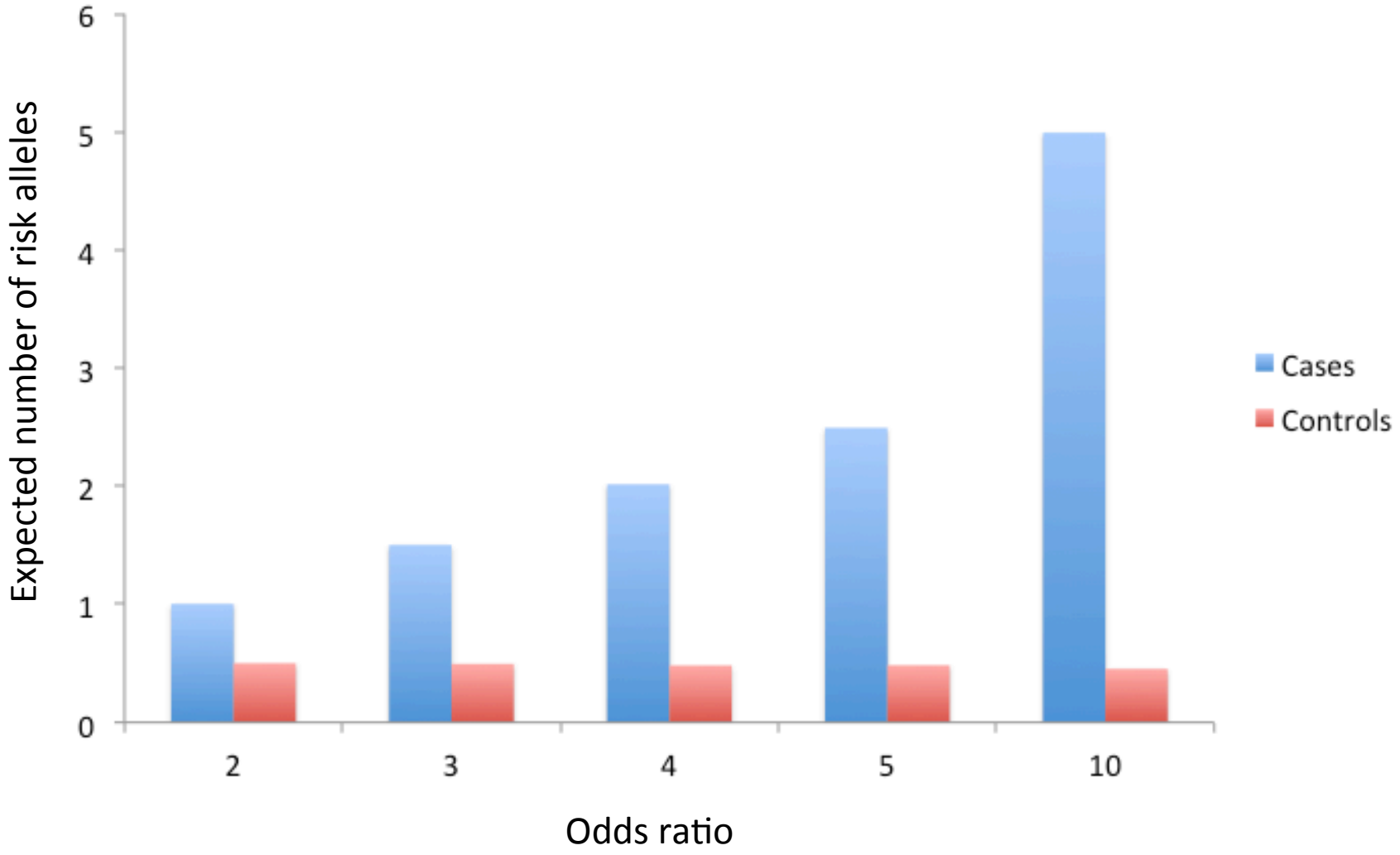
Rare variant inclusion strategy
available sequence data

- Schizophrenia – ~2500 cases; ~700 trios
- Bipolar disorder – ~2500 cases combined from three studies through the Bipolar Sequencing Consortium
- Autism - ~1,000 trios

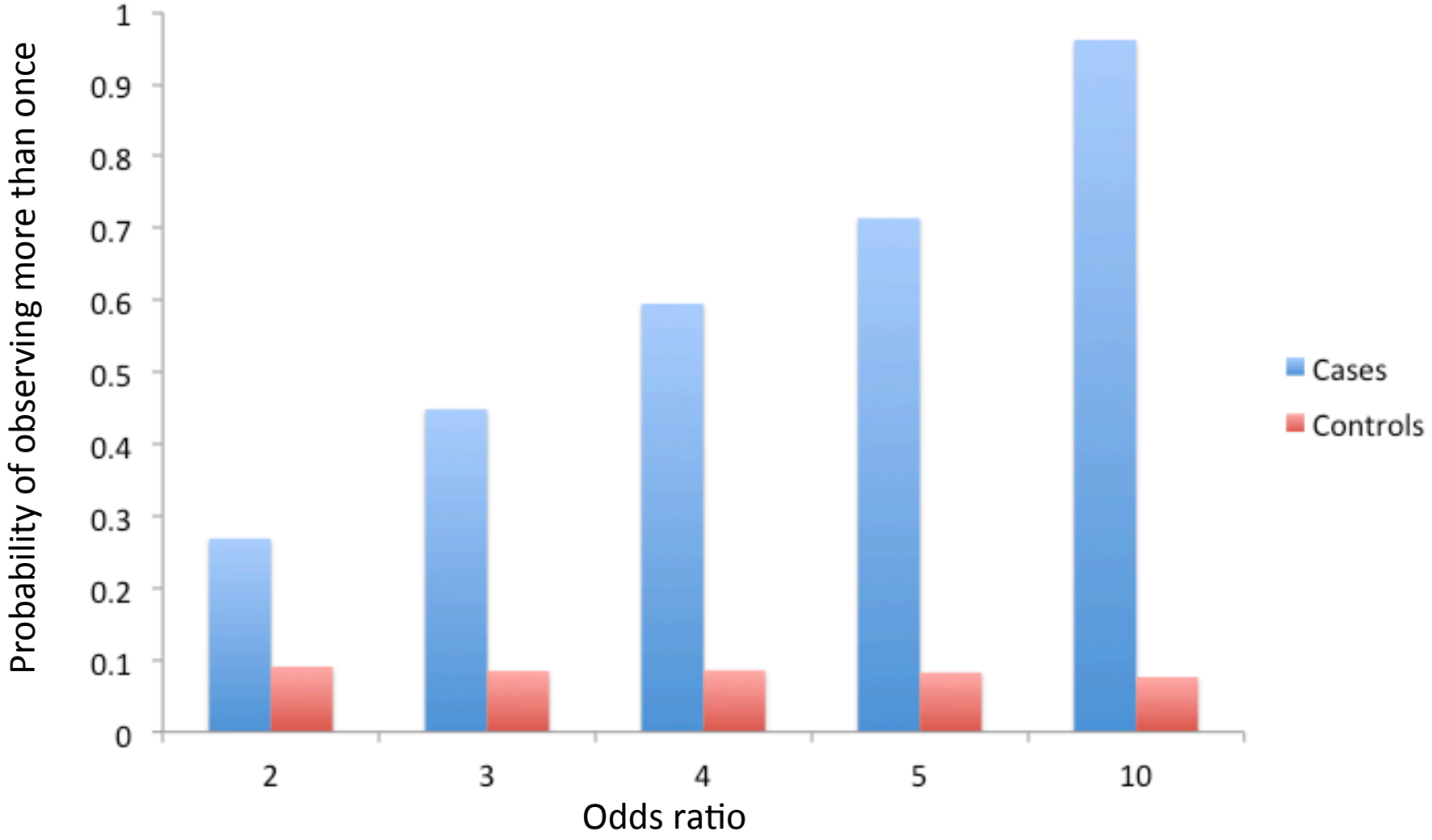
Rare variants & PsychChip

- Most low frequency coding variation
  - ExomeChip
- For a few specific genes of high interest
  - sequencing by genotyping, focus on disruptive/nonsense mutation
- **Leverage existing exome sequence data**
  - **sufficiently rare to likely not be on ExomeChip**
  - **but sufficiently penetrant to be recurrently observed in 1000s of SCZ and BP exomes**

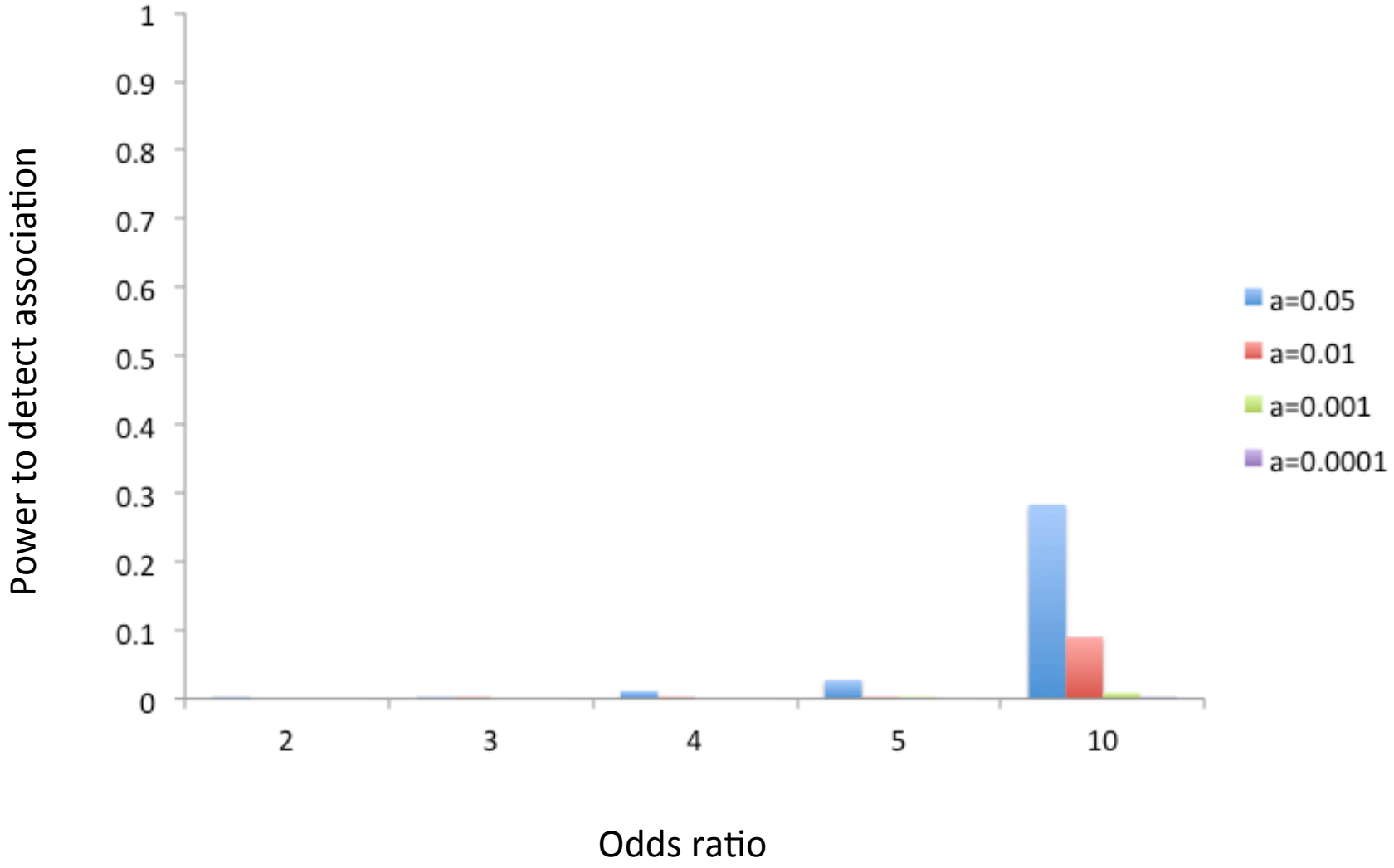
Swedish exome sequencing (2,500 SCZ, 2,500 controls)



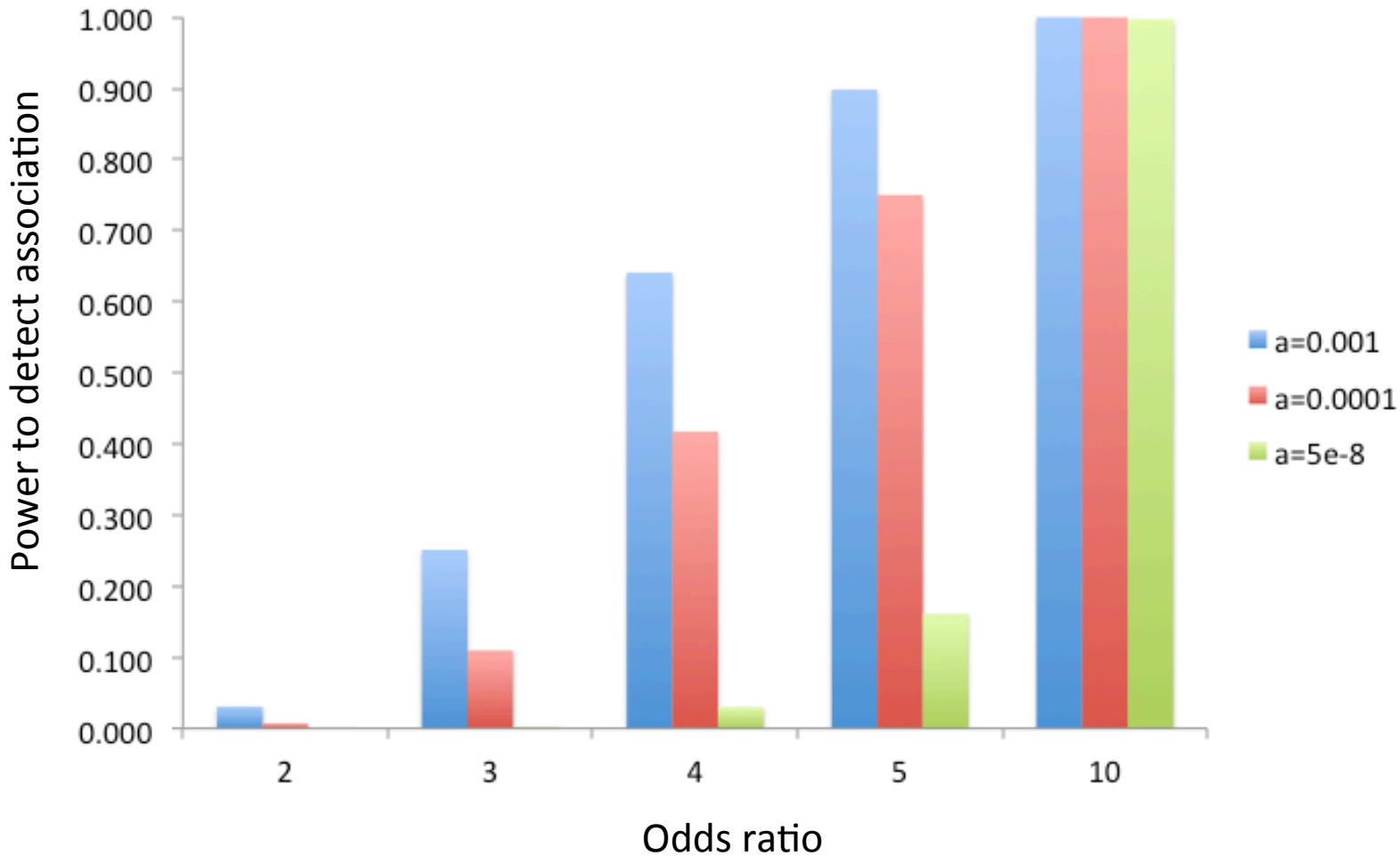
Swedish exome sequencing (2,500 SCZ, 2,500 controls)



Swedish exome sequencing (2,500 SCZ, 2,500 controls)



Potential PsychChip sample sizes: 30K cases 40K controls (e.g. for BP/SCZ contrast)

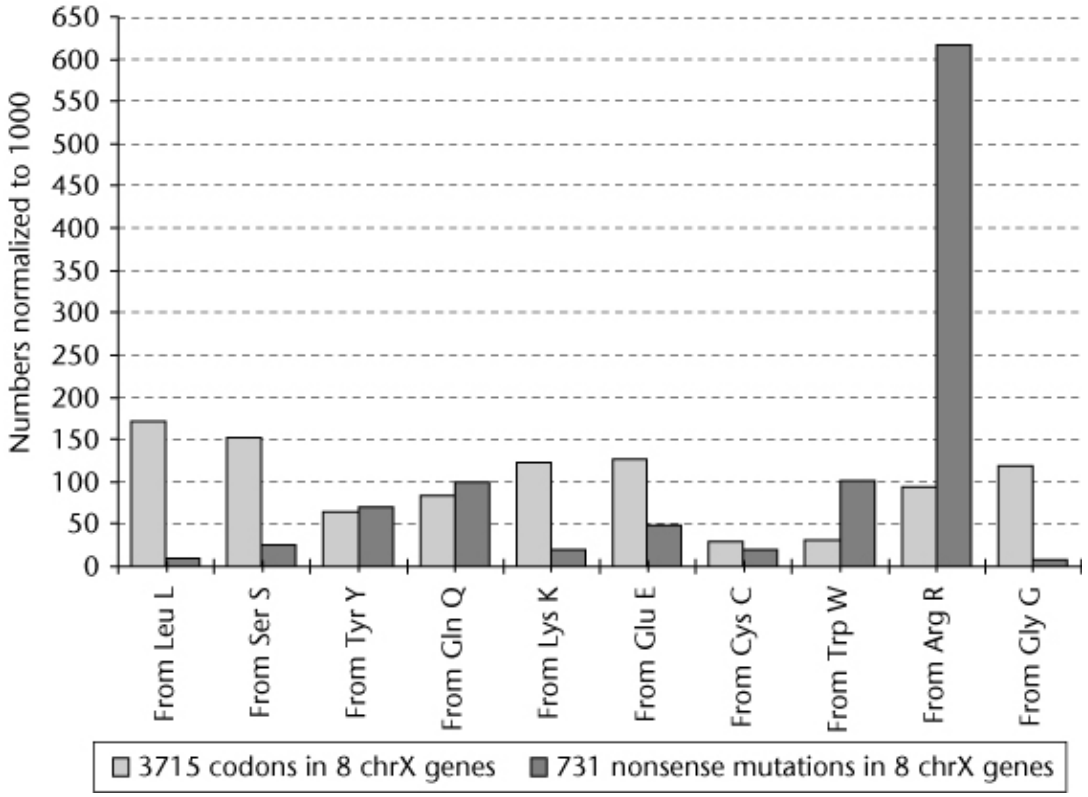


Rare variant inclusion strategy part II

- Autism ~1,000 trios; Schizophrenia ~700 trios
  - Identified significant genes with *de novo* loss of function mutations
  - Nominate a set of genes for "sequencing by genotyping"
  - Aim to estimate the penetrance of rare loss of function mutations
- Can also be used to identify penetrant functional mutations in genesets identified in other ways

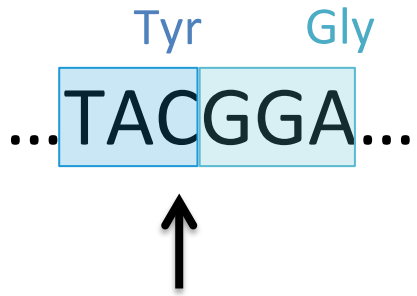
Sequencing by genotyping

- Different codons show different probabilities of nonsense mutation (Antonarakis, 2006)



Sequencing by genotyping

1. Use tri-nucleotide sequence context to determine probability of mutation
2. Annotate each base pair change
3. Select the highest probability nonsense/splice mutations for genotyping



ACG	AAG	7.3e-09	nonsense mutation
	AGG	6.7e-09	nonsense mutation
	ATG	1.2e-07	silent mutation

Example gene coverage

Gene	% mutation	Count
ARID1B	0.860	618
CACNA1C	0.811	627
CACNA1D	0.811	692
CACNB2	0.814	190
CHD8	0.865	799
DYRK1A	0.841	220
GRIN2A	0.805	341
POGZ	0.878	379
SCN2A	0.831	564
SYNGAP1	0.861	358
TCF4	0.794	213

Top 5,000 variants

Excludes same base pair mutations

Chip Design Timeline

- July 15th – final GWAS group submissions for AUT, ADHD, BIP, MDD, SCZ through Stephan Ripke, NOTE: special requests due also
- August 1st – final requests from other groups, OCD/ TD, exome sequencing groups, anorexia, CNV, CDG
- August 15th – drop dead date for any changes or critical additions
- September 1st – Draft submission to Illumina
  - Initial design/return of design scores ~ 2 weeks, expect 5-10% drop out
- September 21st – Submit final PsychChip design
- Oct-Nov 15th – chip manufacture
- End of November – genotyping starts

Criteria for Sample Inclusion

Inclusion criteria for PGC1 and 2 across diseases:

- Each sample had some type of full diagnostic workup, but did not have to be face-to-face interview, broadly accepted many schemas.
- Hospital discharge records, with validation study (e.g. Swedish and Danish samples).

Inclusion criteria for PsychChip:

- For samples meeting 1 or 2 above, for new samples, nominating collaborator simply needs to provide the information for double-checking by appropriate disease group.
- Non-standard diagnostic mechanisms, (e.g. web-based, patient rating only etc), should be reviewed by appropriate individual disease group for quality, and then rationale submitted to the central management group (with the assumption the workgroup recommendation is likely to prevail).
- Nominating group agrees that all data to be deposited on Dutch server for use of PGC, consistent with all prior MOUs and agreements.
- The nominating group confirms that all participants have provided written informed consent and that current Ethical Committee approvals are consistent with the Declaration of Helsinki and with inclusion in the PGC2 single disorder and cross-disorder aims.
- A high molecular weight DNA sample of 1-2 ug (at concentration > 50 ng/ul) is now available. This means that DNA has been extracted using a standard method. Preferred DNA source is blood or cell line. DNA from blood spots, saliva and whole genome amplified will be considered on a case by case basis.
- Nominating group will provide data that samples have successfully been genotyped in some format. Alternatively, a pilot of 3-6 samples will be accessed prior to committing to genotyping an entire cohort.

When completed and meta-analyzed with pre-existing PGC data there will be:

	GWAS		EXOME		CUSTOM PSYCH	
	cases	controls	cases	controls	cases	controls
SCZ	56000	60000	30000	18000	18000	11000
BD	25000	5000	25000	6500	12500	5000
ADHD	30000	11000	30000	11000	30000	11000
AUT	12000	5000	12000	5000	5000	5000
TOTALS	**123000**	**81000**	**97000**	**40500**	**65500**	**32000**

Note: sample sizes are close approximations, samples are frequently both added and removed as they are identified or genotyped by other methods

How to get your samples in the queue?

- Complete form that will be available on PGC website next week, and return to me
- Basic information required;
  - Sample size, diagnostic scheme, confirmation of proper IRB for DNA transfer, confirmation of MOU for PGC use, extraction method, DNA quantification method, estimated date DNA can be transferred, any limitations
- Information/registration of samples due by August 1st
- Samples due by Dec 1st

Genotyping locations and contact individuals

- Broad Institute
  - Kim Chambert; chambert@broadinstitute.org
  - Rich Belliveau (ADHD); rbell@broadinstitute.org
- Icahn School of Medicine at Mount Sinai
  - Jessica Johnson; jessicas.johnson@mssm.edu

Acknowledgements

All members of the PGC!!!!!

Questions?