Genotyping Polymorphic MicroRNA Target Sites in Individuals with Schizophrenia

Michael Moreau
Marco Azaro
Anthony Marcketta

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Overview of Presentation

- Schizophrenia background
- Experimental design:
  - Samples, selecting candidate genes
  - Identifying polymorphic microRNA targets sites
  - An improved bead-based SNP genotyping method
- Results of family-based genetic association analysis

Schizophrenia

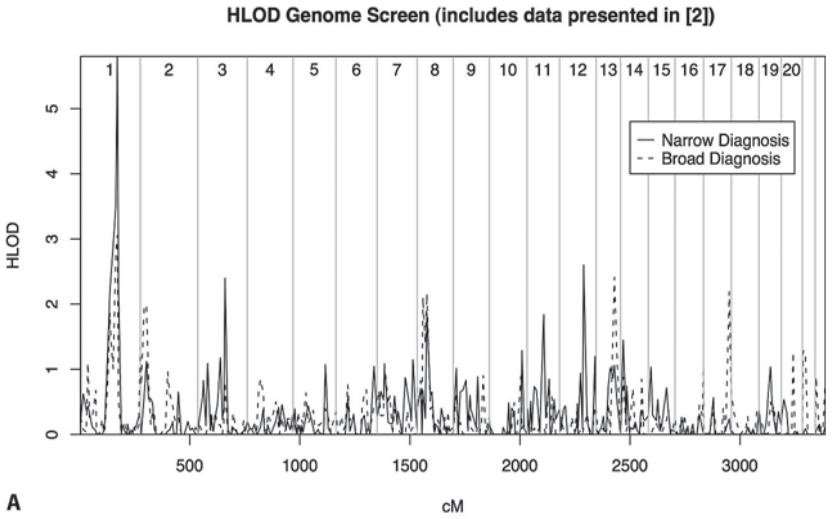
- Schizophrenia is a type of psychosis characterized by disturbed thought, perception, emotion, and behavior
  - Prevalence is about 0.5-1% in the general population
- The Psychotic Symptoms: Active manifestation of abnormal behavior, distortions of normal behavior
  - Delusions, hallucinations
- The Negative Symptoms: Absence or lack of normal behavior
  - Apathy, alogia, anhedonia, affective flattening
- The Cognitive Symptoms: Severe disruptions in speech, attention and concentration, psychomotor function, learning and memory

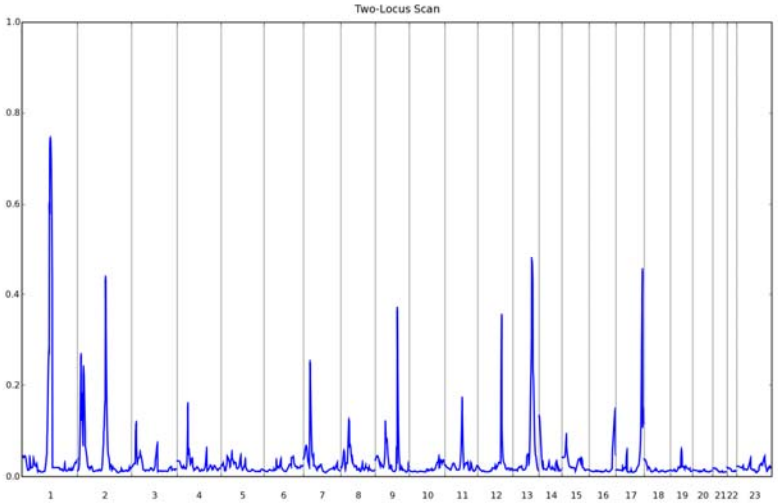
Study Design - Samples

- Family-based genetic association study
  - 392 subjects from 25 extended pedigrees of Canadian Celtic ancestry
  - On average, 13.8 individuals per family and 3.6 affected with schizophrenia or schizoaffective disorder
  - Illness appeared to segregate in a unilineal, autosomal dominant manner

Study Design – Candidate Gene Selection

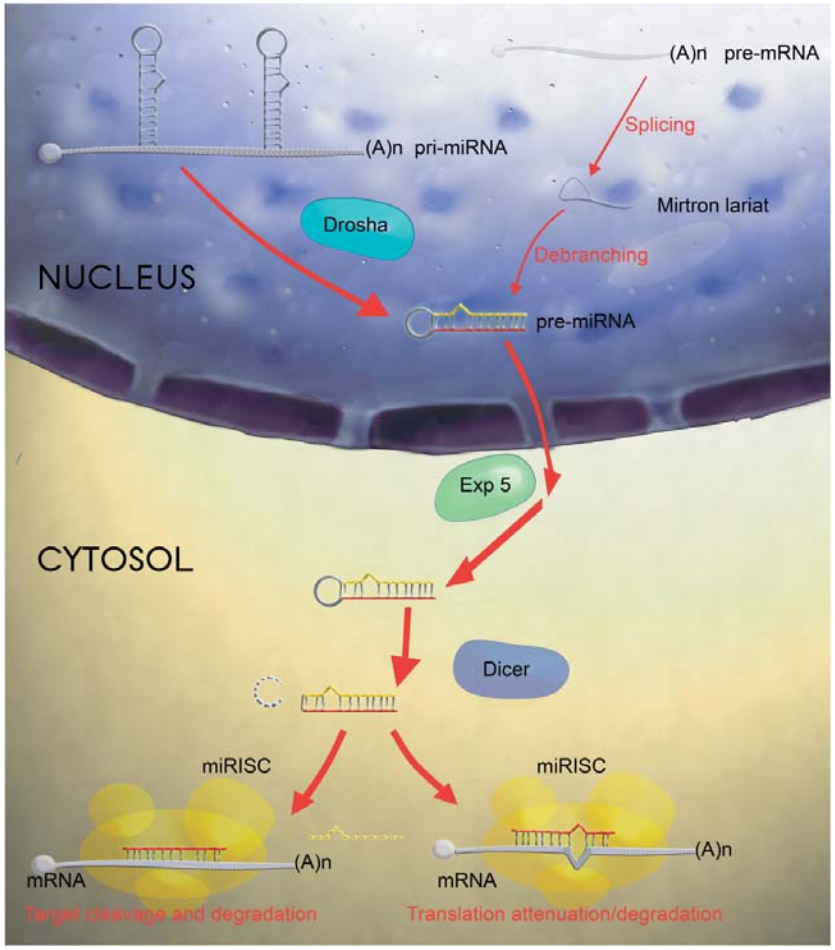
- In 2000, a genome-wide linkage scan revealed strong linkage to 1q21-q22
- Reanalyzed using the PPL in 2006
- Fine mapping and mutation analysis identified putative functional variant in NOS1AP
- NOS1AP risk allele used to define liability classes, conditioned linkage scan performed to find factors that interact epistatically
- Peaks from conditional scan used to define genomic boundaries, mined for polymiRTs





microRNAs

- microRNAs are ~21 nt non-coding RNA molecules that regulate gene expression post-transcriptionally
- miRNAs biosynthesis consists of two enzymatic processing steps
- Bind to 3’ end of mRNA target
- Sequence complementarity affects efficacy of gene repression and mechanism of action

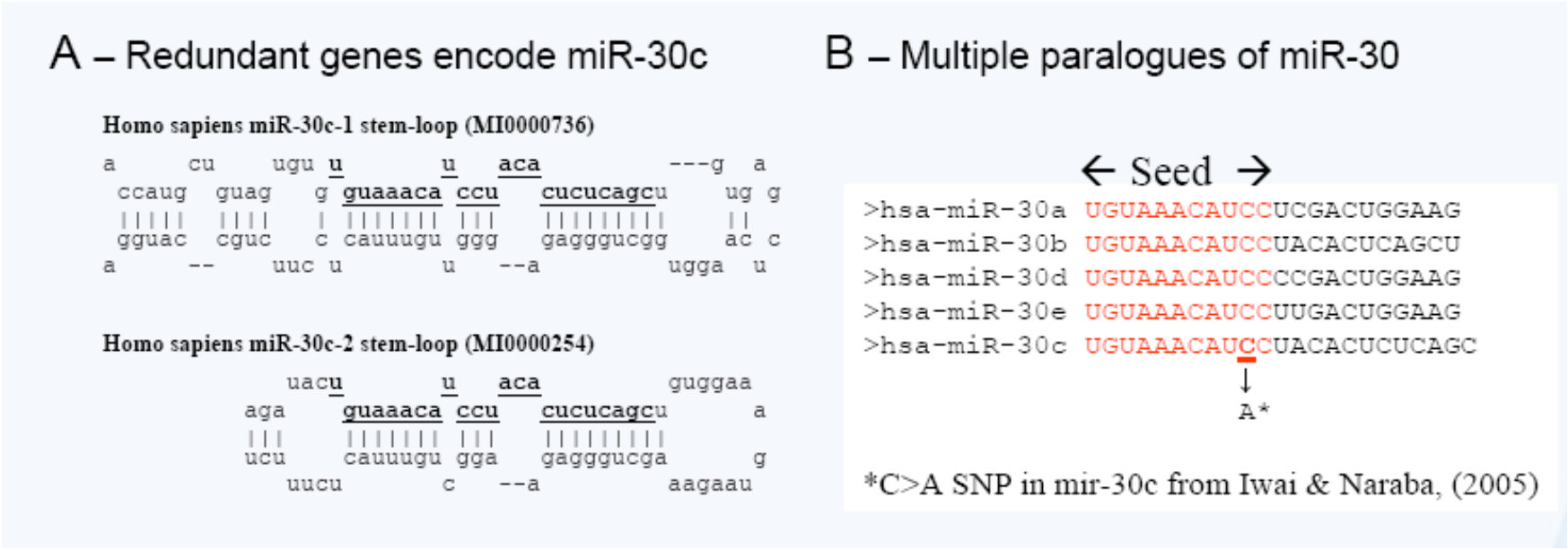


from Mendes et al., Nucl. Acids Res. March 18, 2009 (epub)

Variation in miRNA genes

- Mutations in miRNA genes and targets have very different implications
  - Mutations in miRNA genes have broad implications
  - One miRNA can target hundreds of transcripts
  - Evolutionary selection against disrupting entire regulatory networks

Redundancy may ameliorate impact:



Barnes, M. R., S. Deharo, et al. (2007). "The micro RNA target paradigm: a fundamental and polymorphic control layer of cellular expression." Expert Opin Biol Ther **7**(9): 1387-99.

Variation in miRNA target sites

- Variants in miRNA target sites are more likely
  - Biological effects can be subtle or pronounced
  - SNPs may create, strengthen, weaken, or destroy miRNA target sites

Clop, A., F. Marcq, et al. (2006). "A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep." Nat Genet **38**(7): 813-8.

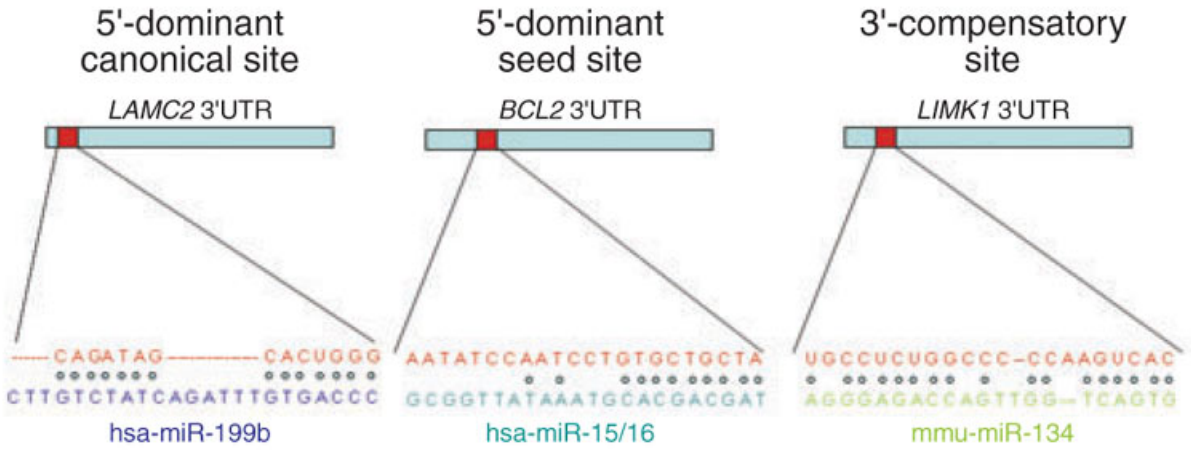
Prediction of miRNA targets that overlap validated SNPs

- Given a list of genes, what do we need to identify polymorphic target sites?
  - in this particular case, we are interested in a set of 6467 genes that fall beneath linkage peaks

- We need:
  - 1) a source of accurately annotated mRNAs (AceView)
  - 2) a database of SNPs with accurate positional and validation status
  - 3) a high-throughput method to predict miRNA targets that does not rely on interspecies conservation
    - PROBLEM: this requirement invalidates the majority of existing target prediction programs!

Why is miRNA target prediction so challenging?

miRNA-target heteroduplexes fall into 3 general classes:



- although perfect complementarity in the seed region (positions 2-7) is characteristic of a large class of alignments, the number and position of remaining matches, mismatches and bulges is highly variable
- standard signature-based alignment algorithms prioritize perfect seed matching
  - Many chance alignments, improved signal/noise by factoring in context determinants, evolutionary conservation
- may be better to look for conserved patterns that are not necessarily contiguous

from Sethupathy et al., 2006, Nature Methods, Vol.3 No.11

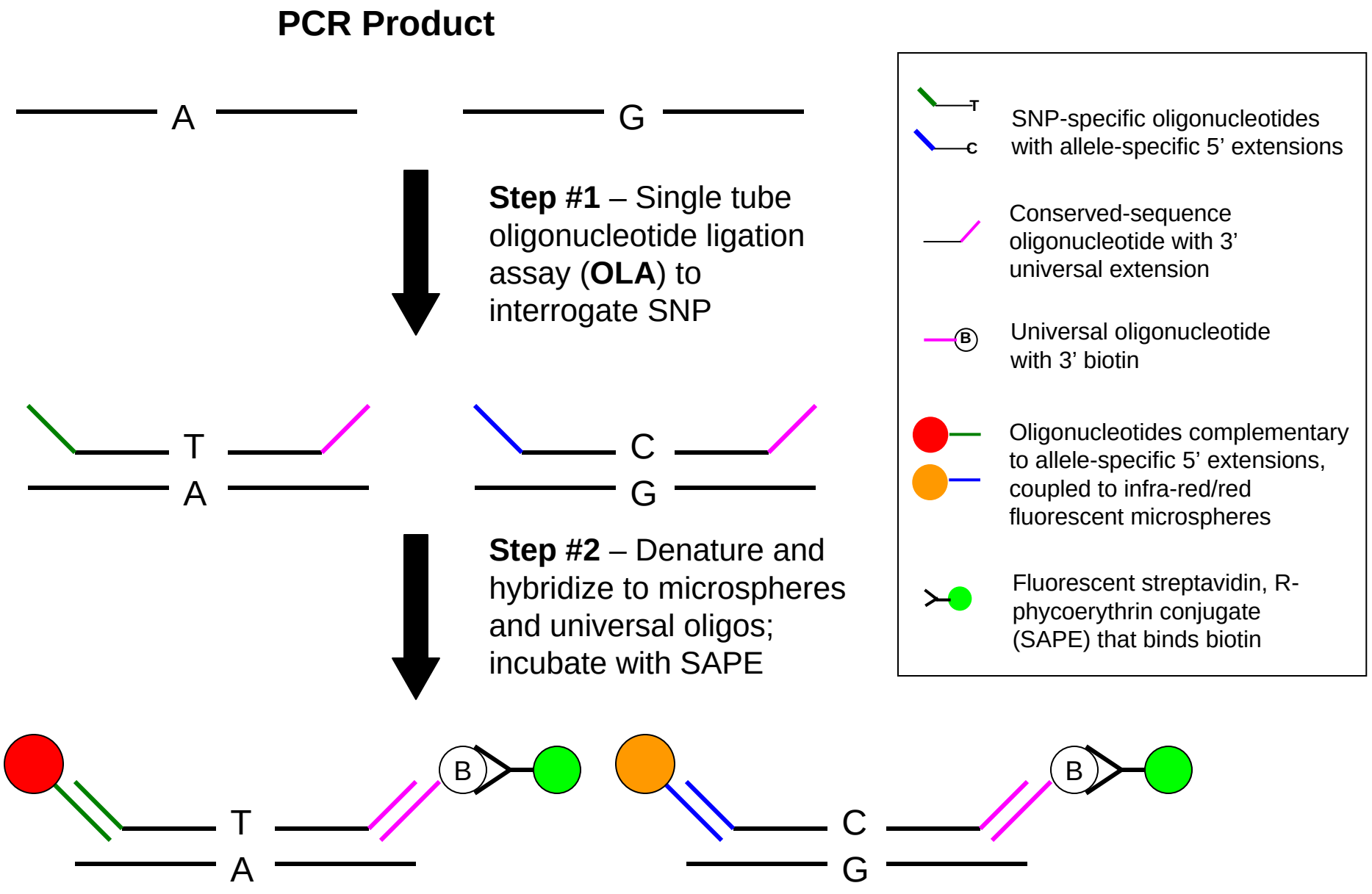
Identifying Polymorphic miRNA Target Sites

- *rna22* is a pattern-based approach
  - Training set from (Rfam) rel 7.1
  - Teiresias algorithm used to discover variable-length motifs
    - e.g. [AT][CG].TTTTT[CG]G..[AT]
  - Map reverse complements and define “target islands”
  - Pair each target with a candidate miRNA
- Experimentally demonstrated repressions of 30% or more for 168 of 226 tested targets

Our procedure is a mutant offspring of Rna22

- with a few differences:
  - 1) We have newer and greatly expanded sets of mature miRNAs
    - Release 7.1 (10/06) had 319 human miRNAs in a total of 3102
    - Release 12.0 (9/08) has 692 human miRNAs in a total of 9169
      - we are considering patterns derived from human miRNAs only!
  - 2) We consider all alternative splice variants curated by AceView
    - NOTE: most target prediction programs ignore tissue-specific splice variants and alt. poly-A termini
  - 3) Our background probability computation uses a random sample of human genomic fragments that reflect size distribution of miRNA set
  - **4) We substitute 3'UTRs with all relevant valid SNPs**

The Use of A Novel Genotyping Method

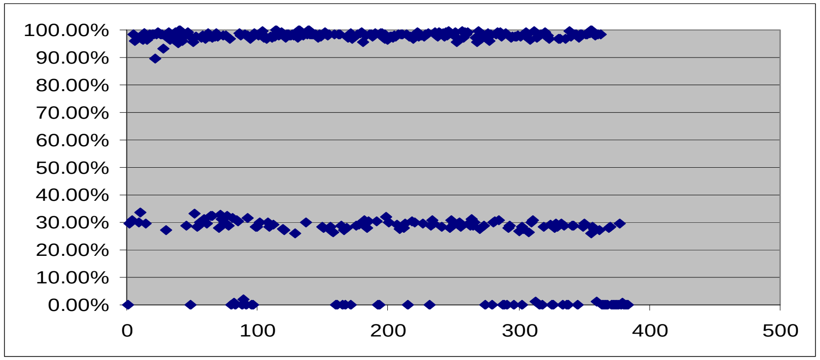


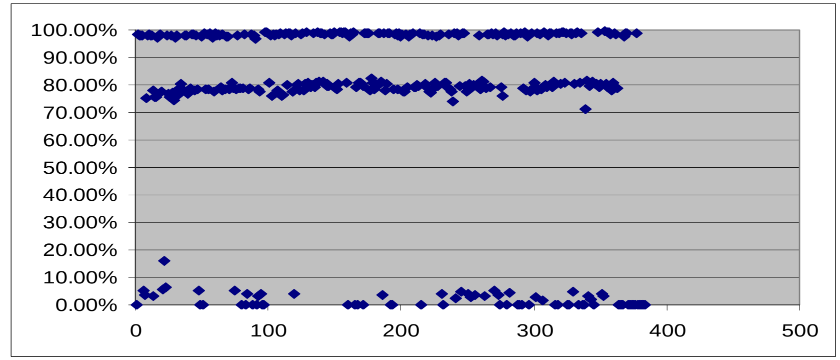
The Use of A Novel Genotyping Method

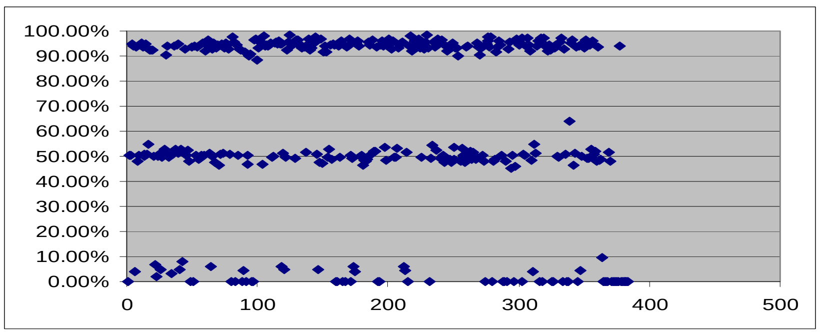
- **Step #3** Detect using two color flow cytometric analysis (Luminex 100)

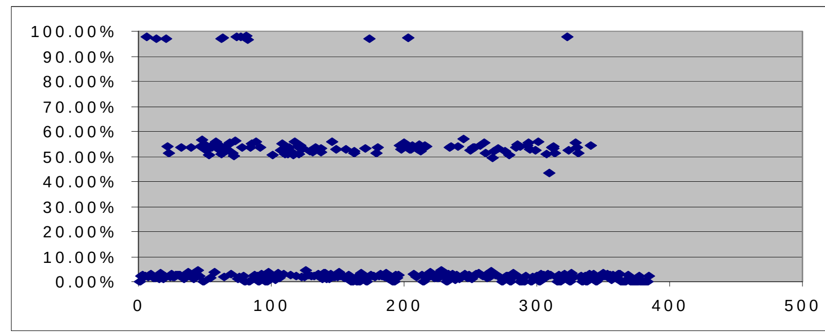
	rs75797_C	rs75797_G	rs79974_C	rs79974_T	rs80104_C	rs80104_G	rs98566_C	rs98566_T	rs98932_C	rs98932_G
5	417	6202	188	91.5	69	44	52	34	53	82.5
9	119	4758	1707	1670	4668	6168.5	8399.5	10675.5	4883	504
5	207	4903	1878	1845	4548.5	6147	8849	11570	4965	333
5	233.5	4871.5	3370	192.5	4271	6266	9006	10236	4487.5	530
5	81	5132	3552	226	6777.5	240	599.5	15745	4371.5	347.5
5	304	4495	143	3463	7870.5	325.5	8815	10833	4218.5	405
5	230.5	5223	3675	257	5035	6047	13317	1217.5	5319	352.5
5	187	4734.5	1911	2084	4381	5751	12484.5	1352	4733	376
5	248	5213.5	1803	1782	4778	6661	9092.5	11030	5367	297
5	218.5	5115	3075	190	4486	6518	9744.5	11392	5074	454.5
0	170	4486	3367	172	7660	342	13771	1351	4980	476
5	248	5346	3525	247.5	8015	304.5	12279.5	1337	5556.5	532
5	365	4641	1841.5	1790	7793	151	8356.5	11476	5217.5	407
3	300	5306	1873	1871	4166.5	5967	8544	11016	4435	325.5
9	264	4867	3560	195.5	329.5	9530	12861	1248	4291	344
7	118	5231	2204	2126.5	4627	6374	9197	11821	4806	430
5	167	4936.5	2398	1985.5	8078	271.5	14065	1428	4998	393.5

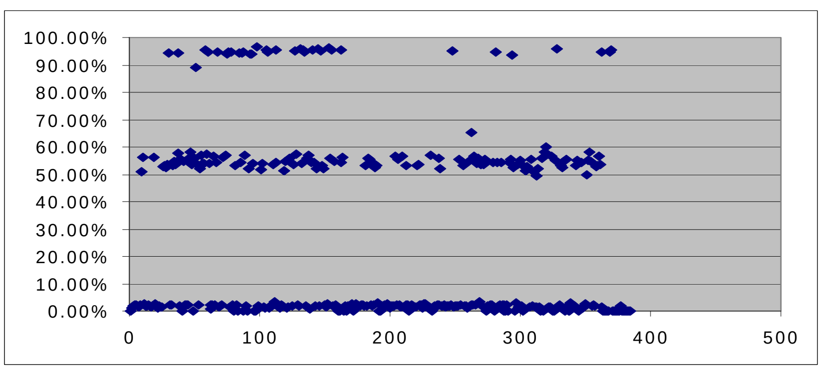
Genotype Calling

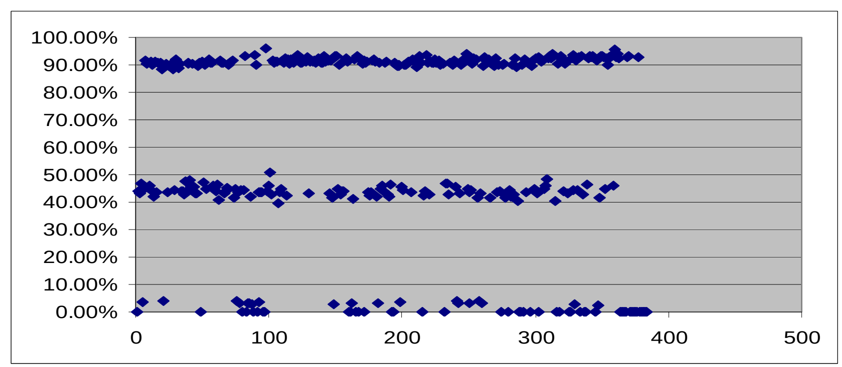












Improvements to bead-based oligonucleotide ligation SNP genotyping assays

- Refinement and validation of a robust method for multiplexed SNP genotyping on the Luminex platform:
  - Oligonucleotide ligation assay (OLA)
  - Use of fewer beads
  - Incorporation of biotin using a universal oligo
  - Multiplexed PCR
  - No wash steps

Improvements to bead-based oligonucleotide ligation SNP genotyping assays

Table 1. Inter-well CV and MFI Range Using Low and High Number of Beads in Control Experiment

200 input/20 count	Positive control	Negative control
Avg. inter-well CV	2%	26%
Avg. MFI	21,968	192
MFI range	21,003–22,842	62–299
5000 input/300 count	Positive control	Negative control
Avg. inter-well CV	2%	20%
Avg. MFI	22,619	162
MFI range	21,496–23,022	115–205

Median fluorescence intensity (MFI) was assessed using the high PMT setting of the Luminex 100 flow cytometer. Inter-well CVs were calculated manually and are based on the median.

Concordance between Luminex OLA and other genotyping methods

- Comparison of 8 SNPs in 14 human genomic DNA samples using direct sequencing; 100% concordance
- Comparison of 6 SNPs in 84 samples using a gel-based version of the OLA; 100% concordance
- Comparison using pyrosequencing:

SNP	concordant calls	% concordant
ID345549	231/232	99.60%
rs1123005	287/289	99.30%
rs11806859	168/170	98.80%
rs12122048	58/58	100.00%
rs4657179	290/290	100.00%
rs4657187	167/168	99.40%
rs905720	263/264	99.62%
Total	**1464/1471**	**99.52%**

Cost breakdown

	Cost/rxn	Cost/ genotype
PCR reagents	0.156	
OLA reagents	0.085	
Luminex reagents	0.053	
Fixed Cost	**0.294**	
Primers	0.30	0.02
Beads (200/allele)	0.6	0.04
Total	**1.194**	**0.080**

Results Summary

- 48 SNPs genotyped in 392 samples
  - 3 multiplexed assays: 15-plex, 16-plex, 17-plex, total of 12 luminex reaction plates
  - All genotyping was completed within 2 days
  - 45/48 (95.6%) converted to working assays, standard conditions with no optimization required
  - 2 markers were non-polymorphic in the Canadian pedigrees
  - 43 markers analyzed for genetic association to schizophrenia or schizoaffective disorder

Association Analysis

- Posterior probability of linkage disequilibrium (PPLD)
  - A Bayesian method which estimates the probability of linkage disequilibrium between a marker and a causal mutation
  - Integrates out nuisance parameters related to mode of inheritance (MOI) of disease, allowing for a “model-free” analysis
  - The prior probability is set to 2%; values above 2% represent evidence for LD, below 2% represent evidence against LD
  - Simulations indicate that 95% of PPLD values are below 4%; >99% are below 10%

Marker	PPLD|L narrow	PPLD|L broad	Marker	PPLD|L narrow	PPLD|L broad
rs01271	0.0190	0.0178	rs64149	0.0132	0.0131
rs07539	0.0181	0.0190	rs75427	0.0300	0.0185
rs11768	0.0153	0.0166	rs77494	0.0175	0.0181
rs20360	0.0234	0.0226	rs82136	0.0248	0.0151
rs24834	0.0148	0.0152	rs88303	0.0175	0.0165
rs35575	0.0144	0.0165	rs89885	0.0147	0.0131
rs36147	0.0400	0.0221	rs11895	0.0168	0.0181
rs53550	0.0159	0.0180	rs12311	0.0158	0.0150
rs69391	0.0167	0.0177	rs15793	0.0154	0.0197
rs75797	0.0177	0.0177	rs17507	0.0161	0.0149
rs79974	0.0185	0.0165	rs18910	0.0102	0.0135
rs80104	0.0145	0.0121	rs23444	0.0600	0.0197
rs98566	0.0154	0.0197	rs24663	0.0188	0.0179
rs98932	0.0151	0.0114	rs28687	0.0115	0.0041
rs02294	0.0152	0.0155	rs32305	0.0181	0.0154
rs11218	0.0179	0.0171	rs37530	0.0180	0.0187
rs11988	0.0149	0.0132	rs50892	0.0165	0.0158
rs25003	0.0196	0.0182	rs60120	0.0500	0.1700
rs34228	0.0219	0.0213	rs72916	0.0195	0.0201
rs43293	0.0126	0.0144	rs84430	0.0157	0.0151
rs57007	0.0159	0.0160	rs98931	0.0125	0.0097
rs63196	0.0172	0.0165			

Results Cont.

Marker	PPLD Narrow	PPLD Broad	Gene
rs3736147	0.0400	0.0221	BIN3
rs10923444	0.0600	0.0197	GDAP2
rs1060120	0.0500	0.1700	H3F3B

Conclusions

- We have developed an efficient, robust, and economical medium throughput genotyping assay for the Luminex platform

- Ideally suited for:
  - Candidate gene studies - tagSNP approach
  - Follow-up, high density SNP genotyping after WGAS
  - Custom panels of putative functional variants
    - Functional elements such as PolymiRTS
    - Polymorphisms in drug metabolizing enzymes

Acknowledgements

- Dr. Linda Brzustowicz
- Dr. Marco Azaro
- Anthony Marcketta
- Dan Lee
- Academic Collaborators
  - Dr. Anne Bassett
  - Dr. Veronica Vieland
- Luminex Corp.
  - Tim Betit
- Perkin Elmer
  - Brian Polizzi
  - Nicole Flaherty