Concentration of Double Strand Break Specification Signatures in Autism Associated Genes

A. Omaiye[1], N. Doan[2], T. A. Deisher[2], S. D. Solares[3]

(1) Seattle University, Seattle, WA; (2) Sound Choice Pharmaceutical Institute, Seattle, WA; (3) University of Maryland, College Park, MD

Introduction

Study of autism spectrum disorder (ASD) on an individual basis has identified hundreds of diverse de novo mutations, deletions and duplications present in about 10% of cases of simplex ASD, with an estimate of 15% when taking into account the inability to measure smaller copy number variations (CNVs). Multiplex ASD and unaffected patients carry much smaller percentages, indicating environmental influences on the genetics of ASD. These factors point at mutations as a primary cause of autism (Sabat et. al. 2007). Although ASD is highly heterogeneous in terms of genes[5] and mutations[1], the affected genes associate within biological networks, explaining the common phenotype from hundreds of diverse mutations (Gilman et. al. 2011). A more urgent question to be posed is "What causes hundreds of diverse de novo mutations to arise?"

Altered double strand break (DSB) formation and repair pathways may be a commonality among the diverse genetic mutations that have been documented in ASD. Much of what is known about DSB comes from the study of meiotic recombination (MR). A degenerate 13 mer sequence, CCNCCNTNNCCNC, has been associated with 41% of MR hotspots as well as with disease causing nonallelic homologous recombination hotspots and common mitochondrial deletion hotspots (Myers et. al. 2008). The zinc finger protein, PRDM9, has been shown to be a global regulator of MR hotspots through its binding to the 13 mer consensus sequence (Berg et. al. 2010). Despite this, PRDM9 allelic expression patterns do not predict actual MR rates on a population basis (Kong et. al. 2010). However, the relationship between MR activity and PRDM9 allelic binding affinity to constrained 13 mer sequences has never been examined, and this relationship may reveal population based associations.

Sites of HR are susceptible to subsequent DSB and mutations (Hicks et.al 2010), and therefore are a potential player in the hundreds of diverse de novo mutations documented for simplex ASD. Therefore, we have determined the genomic locations of constrained 13 mers, modeled PRDM9 binding affinity to the constrained 13 mers, and looked for associations between these DSB regulating factors and genes associated with ASD.

Figure 1: Distance of 13 mer from closest SNP vs SNP recombination rate (all X chromosome genes)

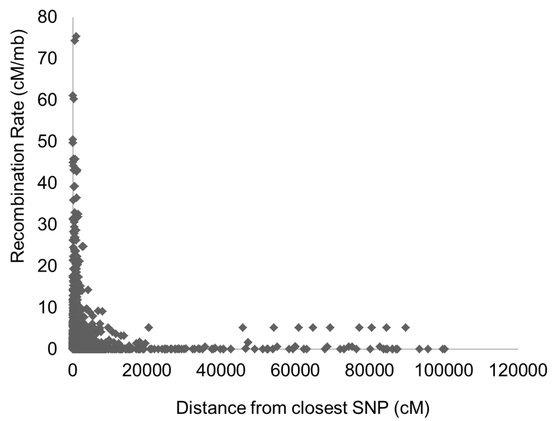

Fig 1 Legend: Calculated binding energies for constrained 13 mer motifs did not correlate to distance from the center of hotspots, nor did they correlate to recombination rate at that hotspot. However, as has been published, the distance of the 13 mer from the hotspot did correlate to recombination rate at that hotspot, as shown above in Fig 1, validating our coding programs.

Methods

Data Collection
All currently identified autism association genes were compiled into a single file by combining the AutDB[7] and ACGMap[8] databases. Human genomic nucleotide sequences were collected in their entirety by chromosome from UCSC's FTP site[9], and gene positions were obtained from UCSC's Table Browser.[10] Meiotic recombination hotspot locations were retrieved from the International HapMap Project[11] and their coordinates were changed from build 35 to build 37 with the UCSC LiftOver tool. An algorithm was written to generate all 1024 specific 13 mer sequences and their reverse complements.

Determination of Motif Position
NCBI's stand-alone BLAST, version 2.2.24[12] was utilized to locate all the 13 mer sequences and reverse complements on each chromosome. Multiple overlaying softwares were written to match the 13 mer locations from BLAST with MR hotspot locations on aags and all genes under chromosomal and exon levels. These were further utilized to match simply the hotspot locations within aags and all genes.

Table 2: Comparison of 13 mer presence within all genes and autism associated genes

Chr	% Genes Containing HS		% Genes Containing 13 mer in HS		% Genes Containing 13 mer	
	All genes	aags	All genes	aags	All genes	aags
1	23.42%	50.00%	13.05%	27.78%	72.97%	72.22%
2	16.97%	62.50%	16.97%	45.83%	78.14%	95.83%
3	27.56%	60.87%	14.99%	52.17%	79.94%	86.96%
4	35.24%	50.00%	17.62%	32.14%	72.17%	78.57%
5	28.68%	36.67%	16.32%	20.00%	75.27%	83.33%
6	26.15%	33.33%	12.70%	23.33%	71.85%	83.33%
7	25.52%	63.00%	18.78%	54.00%	77.00%	97.00%
8	28.95%	53.85%	16.42%	30.77%	74.70%	92.31%
9	23.20%	75.00%	14.42%	62.50%	72.10%	75.00%
10	31.63%	60.00%	17.28%	40.00%	79.78%	90.00%
11	17.85%	36.00%	9.38%	24.00%	65.83%	92.00%
12	20.12%	37.50%	12.59%	37.50%	77.40%	100.00%
13	37.65%	77.78%	20.05%	66.67%	74.57%	88.89%
14	26.49%	66.67%	16.41%	66.67%	64.34%	100.00%
15	23.57%	46.67%	11.79%	46.67%	68.04%	86.67%
16	17.81%	46.15%	10.75%	38.46%	76.61%	92.31%
17	17.10%	26.67%	12.51%	20.00%	74.24%	100.00%
18	32.53%	0.00%	19.88%	0.00%	79.52%	100.00%
19	10.11%	0.00%	8.06%	0.00%	76.38%	100.00%
20	23.88%	44.44%	15.25%	33.33%	74.58%	77.78%
21	25.00%	100.00%	20.42%	100.00%	69.00%	100.00%
22	17.65%	11.11%	13.24%	11.11%	79.78%	100.00%
X	14.00%	45.00%	13.45%	39.00%	65.00%	95.00%
Average	24.43%	47.10%	14.95%	37.91%	74.28%	90.75%
T-Test significance	$6.795*10^{-5}$		$1.300 * 10^{-9}$		$3.577* 10^{-5}$	

Prediction of PRDM9's Relationship to the 13 Mer
PRDM9 is predicted to bind positions 1-11 or 2-12 of the 13 mer motif, and binds via 3 of its 12 zinc finger domains.[13] We tested each position for every combination of its zinc finger domains for 22 different potential binding scores for each constrained 13 mer motif. The Persikov algorithm was applied to generate all possible binding scores for PRDM9, based on the fingers it uses and the locations on the 13 mer it can bind to. Programs were written to match a given PRDM9 binding score to its corresponding 13 mer consensus sequence, and several combinations of data were applied to define PRDM9's relationship to the 13 mer. Binding score and a 13 mer's distance from the a SNP with a significant recombination rate or the center, start, and end of a hotspot inside of a gene were all studied, using all possible scores and then the highest scores for fingers 8-12.

Figure 2: Distance vs. Linear Binding Score for SNPs closest to a 13 mer in X chromosome genes

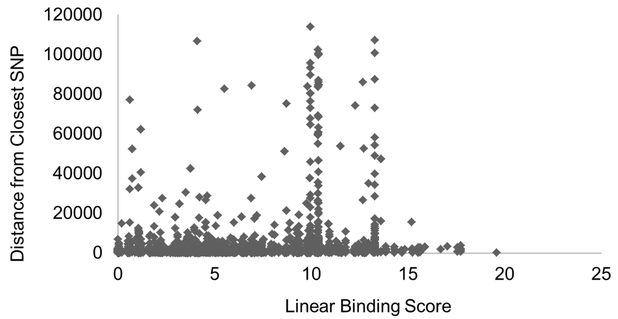

Figure 3: Distance vs. Polynomial Binding Score for SNPs closest to a 13 mer in X chromosome genes

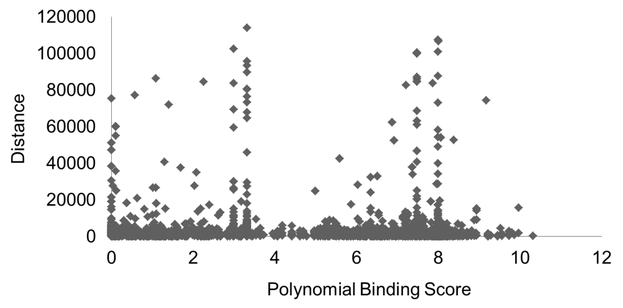

Table 2: Calculated binding energies of all possible zinc finger PRDM9 configurations to the two 13 mer motif potential counter-moieties.

13 mer binding positions	PRDM9 Zinc Fingers									
	10-12	9-11	8-10	7-9	6-8	5-7	4-6	3-5	2-4	1-3
1-11	1.305	4.974	4.16	6.47	3.061	3.795	3.776	--	6.752	3.604
2-12	--	7.726	--	3.591	3.891	3.822	5.188	5.355	--	4.604

Results

1) The vast majority of 13 mers are in non-genic regions.
2) The number of 13 mers per gene is the same in AAGS and all genes.
3) A higher % of AAGS have hotspots, and AAGS have more hotspots per gene.
4) A higher % of AAGS have 13 mers in hotspots, and AAGS have more 13 mers iin hotspots per hotspot overall.
5) Studying only the constrained 13 mer sequences with PRDM9 shows no relation between binding score and the motif's distance from the center, end, or start of a hotspot or the recombination rate of a SNP.

Table 1: Comparison of chromosomal average hot spots and 13 mer concentrations

	Hotspots Per Gene	13 Mers Per Gene	13 Mers in Hotspot Per Hotspot
All Genes	2.38	16.8	1.63
aags	4.21	17.7	2.22

Discussion

Figure 1 demonstrates that we were able to validate published data demonstrating the association of the degenerate 13 mer with meiotic recombination. The closer a 13 mer was to the middle of a recombination hotspot, the higher the recombination rate at that hotspot. However, binding scores for the constrained 13 mers did not demonstrate any relationship with meiotic recombination.

Our demonstration that aags have more HS as well as more 13 mers within HS calls for further study of the role of DSB & PRDM9 in de novo autism-associated gene mutations. However, our tests of PRDM9's binding scores were not informative. The algorithm that calculated binding scores however, did function as expected since our calculated binding score yielded most favorable results for PRDM9 zinc fingers 8-12, as published by others (Table 2). More likely, it is the presence of flanking sequences which inhibit or promote PRDM9's binding affinity to the motif. Another note is that most meiotic recombination hotspots occur outside of genes. Taking into account the binding score of PRDM9 to the 13 mer with respect to these hotspots in addition to intergenic hotspots may ultimately reveal a correlation.

References

1.Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, et. al. 2007. Strong Association of De Novo Copy Number Mutations with Autism. *Science*. **316**: 445-449.

2.Gilman, S.R.; Iossifov, I.; Levy, D.; Ronemus, M.; Wigler, M.; Vitkup, D. 2011. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*. **70**: 898-907.

3.Myers, S.; Freeman, C.; Auton, A.; Donnelly, P.; McVean, G.2008. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature*. **40**: 1124-1129.

4.Berg, I. L.; Neumann, R.; Lam, K. G.; Sarbajna, S.; Odenthal-Hesse, L.; May, C.A.; Jeffreys, A.J.; 2010. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genetics*. **42**: 859-863.

5.Kong, A.; Thorleifsson, G.; Gudbjartsson, D. F.; Masson, G.; Sigurdsson, A.; Jonasdottir, A.; Walters, G. B., et. al. 2010. Fine-scale recomination rate differences between sexes, populations and individuals. *Nature*. **467**: 1099-1103.

6. Hicks, W.M. 2010. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science*. **329**: 82-85.

7.Basu SN, Kollu R, Banerjee-Basu S. Nucleic Acids Res. 2009 Jan;37(Database issue):D832-6.

8.Churbanov A, Vorechovsky I, Hicks C. 2010. A method of predicting changes in human gene splicing induced by genetic variants in context of cis-acting elements. *BMC Bioinformatics*. 11-22.

9.The Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860-921.

10.Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004 Jan 1;32(Database issue):D493-6

11.International HapMap Consortium. The International HapMap Project. Nature. 2003;426:789–796.

12.Kent WJ. BLAT - the BLAST-like alignment tool. *Genome Res*. 2002 Apr;12(4):656-64.

13.Ségurel L, Leffler EM, Przeworski M. 2011. The Case of the Fickle Fingers: How the PRDM9 Zinc Finger Protein Specifies Meiotic Recombination Hotspots in Humans. PLoS Biol 9(12): e1001211. doi:10.1371/journal.pbio.1001211

Conflict of Interest: None (?)

Acknowledgments: This work was funded by the M.J. Murdock Charitable Trust and private donations.