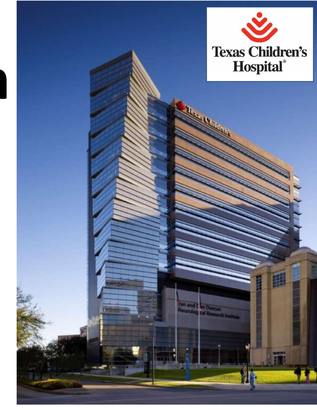
Functional assessment of *de novo* missense variants associated with Autism Spectrum Disorders through an overexpression-based screen in *Drosophila*

J. Michael Harnish[1,2], Samantha Deal[2,3], Paul Marcogliese[1,2], Jonathan Andrews[1,2], Hillary K. Graves[1,2], Sharayu V. Jangam[1,2], Xi Luo[1,2], Ning Liu[1,2], Hsiao-Tuan Chao[2,3,4,5], V. Hemanjani Bhavana[1], Yiu-Huen Tsang[1], Pei-Tseng Lee[1,2], Wen-Wen Lin[1,2], Hongling Pan[2,6], Kenneth L. Scott[1,7], Michael F. Wangler[1,2,3,4,5] Shinya Yamamoto[1,2,3,5]

[1]Department of Molecular and Human Genetics, BCM, [2]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, [3]Program in Developmental Biology, Baylor College of Medicine (BCM) , [4]Department of Pediatrics, BCM, [5]Department of Neuroscience, BCM, [6]Howard Hughes Medical Institute, [7]Program in Integrative Molecular and Biomedical Sciences, BCM, Houston, TX, 77030



Rationale

Whole-exome sequencing (WES) is becoming less expensive and more widespread as research and diagnostic tools for human diseases. WES identifies hundreds of rare variants in an individual's genome that needs to be interpreted by human geneticists and clinicians. However, current *in silico* tools are insufficient to predict the pathogenicity of many missense variants. This becomes especially problematic when undertaking large sequencing efforts to uncover rare variants that contribute to the pathology of relatively common conditions. For example, numerous recent efforts have uncovered rare gene variants associated with intellectual/developmental disabilities and Autism Spectrum Disorder (ASD), but thousands of *de novo* variants are missense variants that are of unknown significance. Here, we utilize an overexpression-based screen in *Drosophila* to investigate several such *de novo* missense variants identified in the Simons Simplex Collection (SSC), a large cohort of over 2,500 ASD simplex families. Using this method, we prioritized 78 genes and generated UAS-human cDNA transgenic lines that allow exogenous expression of reference or variant alleles in flies. From these experiments we discovered 22 gene variants (28%) which had a scorable phenotypic difference between overexpression of the variant and reference allele. These data strongly suggest these variants have a consequence on gene function and may contribute to the pathogenicity of ASD.

Candidate Gene Selection

Exome Data from SSC (*Iossifov et al 2014*)

SFARI SIMONS FOUNDATION AUTISM RESEARCH INITIATIVE

2,717 genes: *de novo* events → Independent genomic database searches

1,513 genes: *de novo* missense variants

396 genes: no *de novo* siblings

979 genes: conserved in *Drosophila*

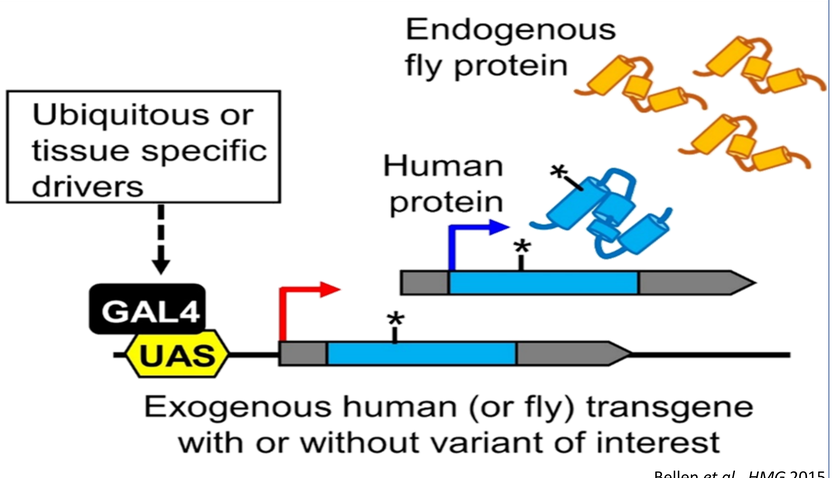
Prioritize genes with available genetic tools

Independent genomic database searches: Genes with variants at same nucleotide; ASD-related phenotypes

Prioritization: 122 fly genes, 130 human genes and >140 variants

Project Overview and Strategy

We utilized the GAL4 UAS system to overexpress human cDNA's in *Drosophila* and compare the wild type and variant for phenotypes. Our Gal4 drivers used were GMR-Gal4 (eye driver), nubbin-Gal4 (wing driver) and tub-Gal4 (ubiquitous expression). We could score gross morphological phenotypes as well as survivability using these assays. If the reference phenotype was more severe than the variant phenotype, that variant was considered a loss of function allele. Conversely, if the variant allele had a stronger phenotype, the variant was considered gain of function. In a few assays we noticed a loss of function with one driver, but a gain of function in another, suggesting some level of tissue-specificity. Examples of these phenotypes are presented in the following figure. Further behavioral assays were performed, data from which are presented in poster X by colleague Jonathan Andrews.



Bellen *et al.*, *HMG* 2015

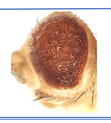
Reference phenotype > Variant phenotype = Loss of function

Reference phenotype < Variant phenotype = Gain of function

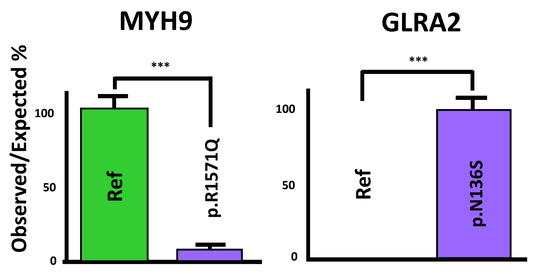
Drivers:
Ubiquitous (tub-GAL4)
Wing (nubbin-GAL4)
Eye (GMR-GAL4)

Phenotype:
Survival
Morphology
Morphology

Results: Morphologically Discordant Phenotypes

Gal4	Reference	Variant	Human cDNA
nub>Gal4  **UAS LacZ control**		p.T99N 	*HTR1D*
		p.V916M 	*EPHB1*
GMR>Gal4  **UAS LacZ control**		p.V916M 	
		p.R873P 	*SOGA3*

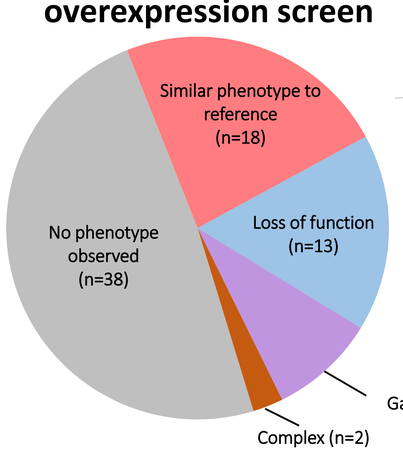
tub>Gal4



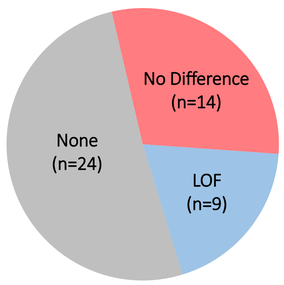
Example phenotypes for nubbin, GMR and tubulin driven human cDNA's. The HTR1D variant allele p.T99N acts as a gain of function allele. The SOGA3 variant allele p.R873P acts as a loss of function allele. EPH1B is a complex allele, as it appears to be a gain of function in the wing under the control of nubbin-Gal4 but a loss of function in the eye under the control of GMR-Gal4. Reference and variant alleles can also affect the viability of flies when driven with tub Gal4. Both loss of function and gain of function alleles were identified in all 3 Gal4 lines use (nub, tub and GMR).

Results: Discordant Variants Identified

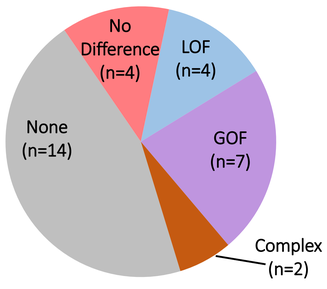
Phenotypes observed from overexpression screen



Human orthologs of essential fly genes



Human orthologs of non-essential fly genes



Our screen identified 22 variants for which there was a different phenotype when the variant allele was expressed as compared to the reference allele. In 49% of cases (n=38) no phenotype was observed when either allele was expressed. Interestingly, human orthologs of essential fly genes were more likely to yield loss of function alleles, while homologs of non-essential fly genes had gain of function phenotypes or complex phenotype as the majority of phenotypic discordance.

Conclusions

- **We identified 22 missense variants from ASD patients with scorable phenotypic differences, suggesting the variant has a functional consequence on the gene of interest**
- **There appears to be a difference in the rate of loss of function and gain of function variants seen in the groups of human orthologs correlating to genes that are essential vs non essential in *Drosophila***
- **13 of the missense variants were identified are in genes with no known disease association and may be novel disease-causing genes**
- **Variants in genes previously associated with human diseases are possible candidates for phenotypic expansion.**
- **We identified variants that were not predicted as deleterious *in silico* by missense scoring algorithms including CADD and PolyPhen**
- **This screening method is a rapid, cost effective way to predict the functional consequence of missense variants *in vivo* and capture both damaging loss of function and gain of function variants that may be incorrectly predicted by current *in silico* methods**