SETBP1 as a novel candidate gene for neurodevelopmental disorders of speech and language

Sergey A. Kornilov, PhD

Yale University
School of Medicine

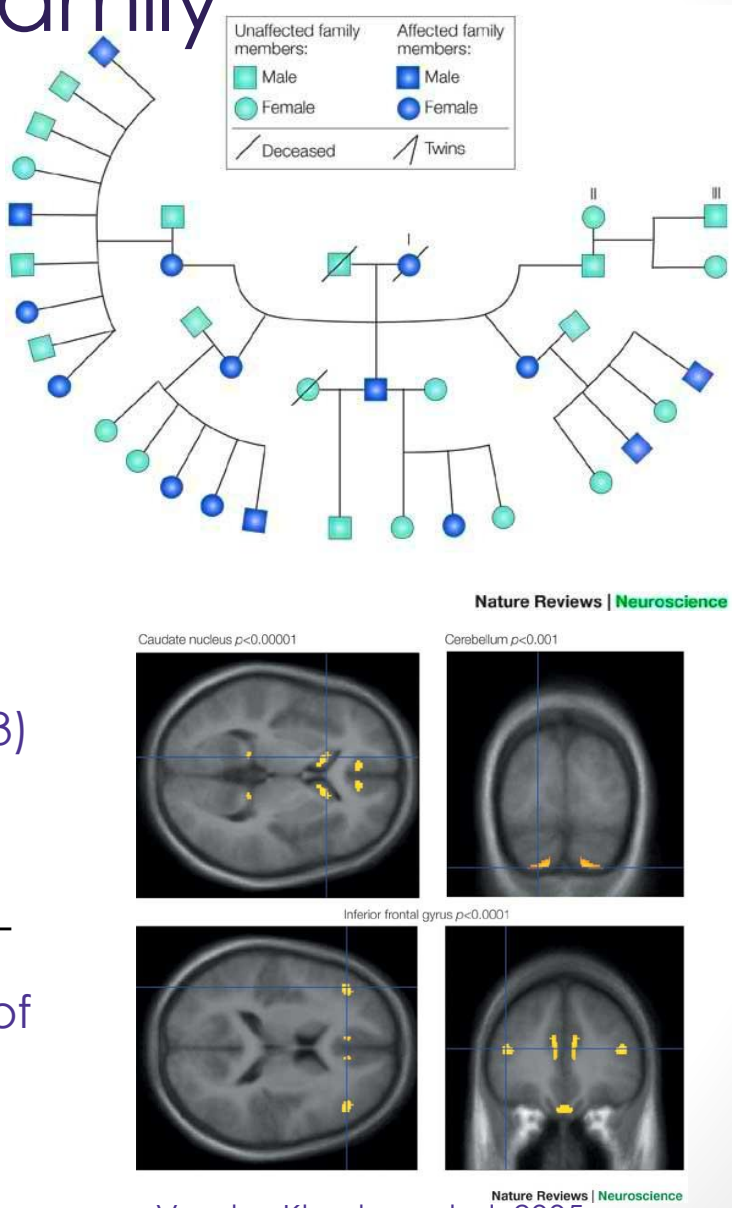
14 October 2015

Developmental Language Disorders (DLD)

- **Difficulties acquiring and using one's native language (APA, 2013)**
- In the absence of apparent sensory, cognitive, neurodevelopmental, psychiatric, genetic conditions
- Affect 7-10% of preschoolers (Tomblin et al., 1997; Law et al., 2001)
- Persist into adolescence and adulthood (Poll et al., 2010)
- Negative outcomes
  - Socio-emotional (Snowling et al., 2006)
  - Behavioral (Snow et al., 2011)
  - Academic (Dockrell et al., 2011)
  - Occupational (Conti-Ramsden & Durkin, 2012)
- Highly familial and heritable (h^2_g from .34 to 1.25; Stromswold, 1998)
- Yet, neurobiological and genetic etiology are largely unknown and heavily understudied

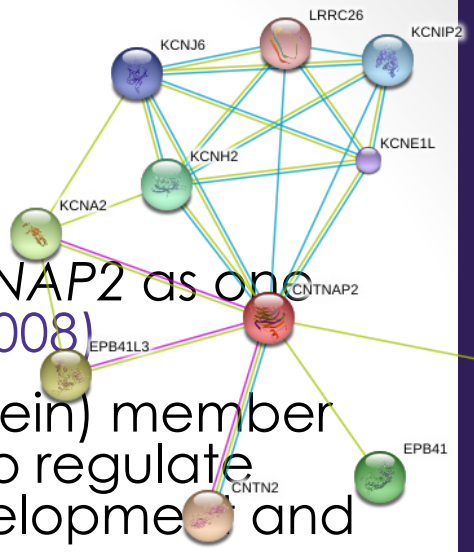
Genetics of DLD: KE family

- Rare case of autosomal dominant monogenic severe orofacial dyspraxia, intellectual disability
- Linkage to 7q31 (SPCH1), localized point mutation in *FOXP2* (Fisher et al., 1998)
  - Haploinsufficiency causes speech and language problems
  - Transcription factor (over 300 genes), highly expressed in the brain (Lai et al., 2003)
  - Associated with atypical brain activation patterns (Liegeois et al., 2003)
  - Involved in the development of vocalization systems in multiple species
  - Yet, only several individual cases have been reported to date; screening of at-risk populations failed to identify *FOXP2* as the causal factor in common forms of DLD (e.g., Claudio et al., 2013)



Vargha-Khadem et al, 2005

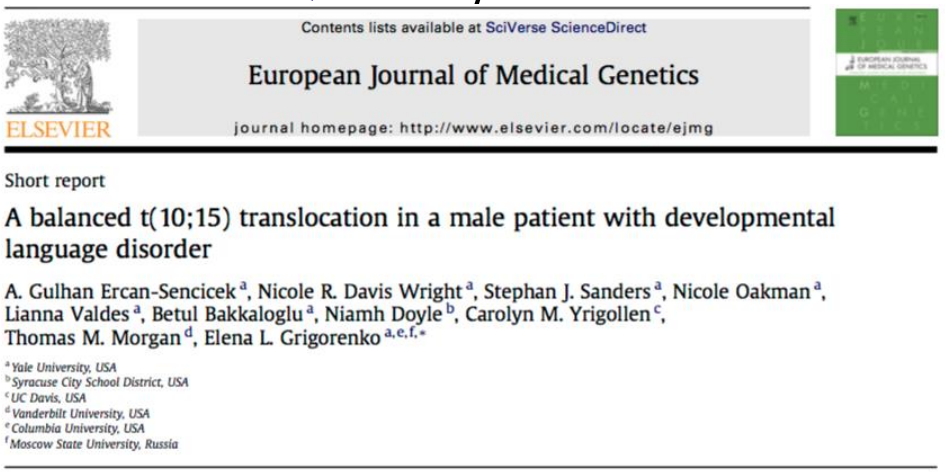
Genetics of DLD: *CNTNAP2*



- Downstream target screening identified *CNTNAP2* as one of the primary *FOXP2* targets (Vernes et al., 2008)
- *CNTNAP2* is a (neuronal transmembrane protein) member of neurexin family, interacts with neuroligins to regulate synapse formation, involved in neuronal development and axonal differentiation
- Associated with DLD endophenotype (pWM) and language measures in the SLIC cohort (Vernes et al., 2008)
- Associated with early language development in general population (Whitehouse et al., 2011) and neural indices of language processing (Kos et al., 2012)
- Like *FOXP2*, seems to be involved in avian vocal learning (Panaitof et al., 2010; Whalley et al., 2011)
- Associated with autism (Alarcon et al., 2008), Gilles de la Tourette (Verkerk et al, 2003), schizophrenia (Friedman et al., 2008), epilepsy (Strauss et al., 2008), intellectual disability (Zweier et al., 2009), ADHD (Elia et al., 2009) - ***erosion of phenotypic specificity*** (State, 2013)

Genetics of DLD: Other candidates

- Targeted associations - *CMIP* and *ATP2C2*, located in the SLI1 region (16q), are associated with the DLD endophenotype (pWM; Newbury et al., 2009)
  - *CMIP* involved in cytoskeletal remodeling (neuronal migration, synapse formation)
  - *ATP2C2* regulates translocation of cytosolic Ca and Ma ions to the Golgi; Ca homeostasis critical for neuronal function
- Structural events in *SEMA6D* (15q21) are associated with DLD (Ercan-Sencicek et al., 2012)
  - *SEMA6D* is part of the plexin/semaphorin/integrin signaling pathways, involved in axon guidance



Contents lists available at SciVerse ScienceDirect

European Journal of Medical Genetics

journal homepage: http://www.elsevier.com/locate/ejmg

ELSEVIER

Short report

A balanced t(10;15) translocation in a male patient with developmental language disorder

A. Gulhan Ercan-Sencicek [a], Nicole R. Davis Wright [a], Stephan J. Sanders [a], Nicole Oakman [a], Lianna Valdes [a], Betul Bakkaloglu [a], Niamh Doyle [b], Carolyn M. Yrigollen [c], Thomas M. Morgan [d], Elena L. Grigorenko [a,e,f,*]

[a] Yale University, USA
[b] Syracuse City School District, USA
[c] UC Davis, USA
[d] Vanderbilt University, USA
[e] Columbia University, USA
[f] Moscow State University, Russia

Genetics of DLD: recent GWASes

- Eicher et al. (2013) – GWAS of comorbid spoken/written DLD (AVON cohort) – tentative findings for *ZNF385D* (also associated with brain volume in PING cohort)
- Luciano et al. (2013) – GWAS of the BATS and ALSPAC cohorts,
- Gialluisi et al. (2014) – GWAS of UK-RD, SLIC, and CLDRC cohorts (PCA-based GWAS meta-analysis), no surviving signals
- Nudel et al. (2014) – GWAS of the SLIC cohort; only parent-of-origin (maternal) effects survived corrections (5p13, intergenic)
- Simpson et al. (2015) – CNV burden was associated with DLD status in the SLIC cohort; not driven by large de novo events; gene-based tests identified *CDC2L1, CDC2L2, LOC728661, and RCAN3* for spelling/reading measures
- Complex landscape of findings from studies with n ranging from 150 to 6,000, most findings do not survive corrections for multiple testing, and those that do are not language phenotypes per se

Why such limited success

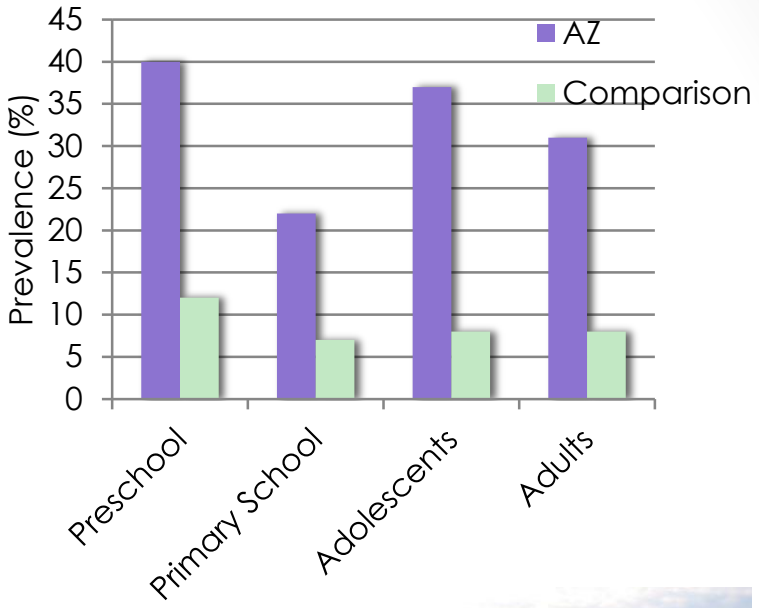
- Exclusionary nature of the diagnostic category
- Locus and allelic heterogeneity of common disorders
- Behavioral as well as etiological heterogeneity of DLD manifestations due to trait complexity
- Compared to other subfields and other neurodevelopmental disorders – understudied and underfunded

	2000–2001	2002–2003	2004–2005	2006–2007	2008–2009	Anuual rate of increase 2000–2010
Developmental dyslexia	18,770	18,199	19,971	22,975	27,283	633
Developmental dyscalculia	0	0	400	369	1,574	151
Developmental coordination disorder	0	623	602	1,379	1,166	55
Speech sound disorder	858	876	501	964	1,116	8
Specific language impairment	16,279	16,219	15,389	18,133	28,611	780
Attention deficit hyperactivity disorder	274,500	346,039	314,232	365,207	532,800	13,042
Autistic spectrum disorder	95,114	171,707	360,765	355,458	851,270	50,978
Tourette syndrome	35,604	42,119	37,031	41,626	59,587	1,150
Angelman syndrome	13,039	12,103	11,156	12,392	21,246	594
Cerebral palsy	18,197	31,412	49,467	66,507	94,578	4,409
Cornelia de Lange syndrome	194	655	677	3,541	5,789	352

Bishop, 2010

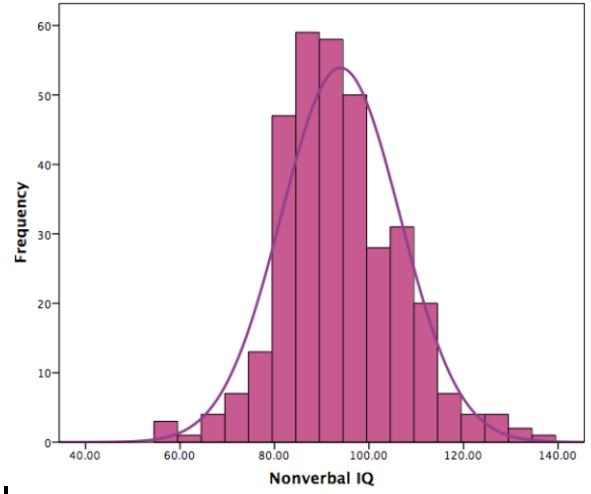
The AZ population

- Geographically isolated population residing in a cluster of villages in Russia's rural north
- 860 residents (118 between the ages of 3 and 18)
  - 82% are (distantly) related
  - Complex 11-generational pedigree (k=6,391)
- Atypically high prevalence of DLD (over 30%; Rakhlin et al., 2013)
- Reduced genetic/allelic variability due to potential founder effects typical of isolates (Wright et al., 1999)
- Uniform environment (same school(s)/similar SES)
- Potentially increased power to reveal genetic bases of DLD



Rakhlin, N., Kornilov, S.A., Palejev, D., Koposov, R.A, Chang, J.T., & Grigorenko, E.L. (2013). The Language Phenotype of a Small Geographically Isolated Russian-speaking Population: Implications for Genetic and Clinical Studies of Developmental Language Disorder. *Applied Psycholinguistics*, 34(5), 971-1003

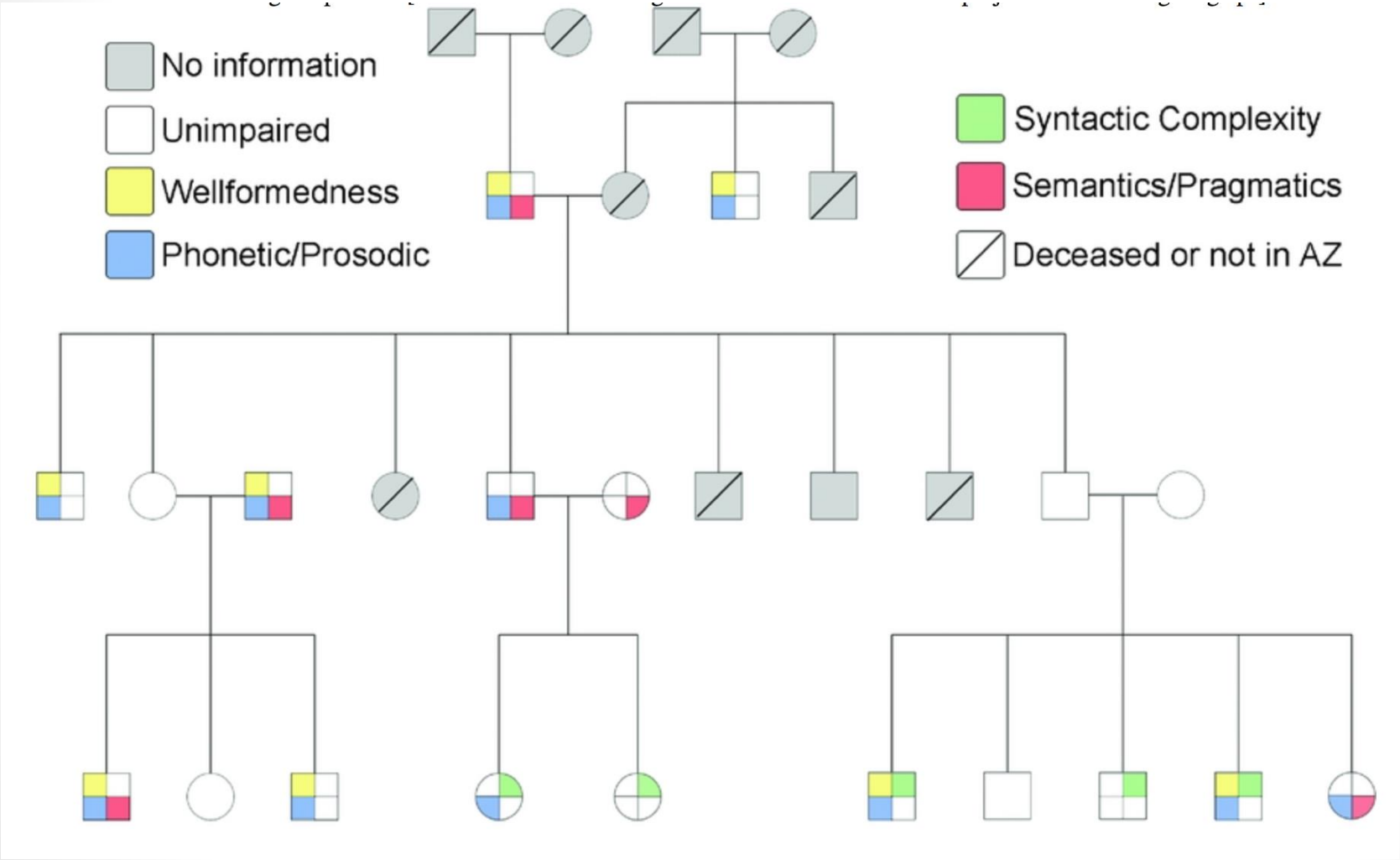
The AZ population



- Nonverbal intelligence normal
- No significant
  - Sensory, neurological, psychiatric problems
- Deficits in
  - Expressive language (Rakhlin et al., 2013)
  - Receptive language (Rakhlin, Kornilov, & Grigorenko, 2014)
  - Written language, reading, and spelling (Rakhlin et al., 2013)
  - Lexical-semantic development (Kornilov et al., 2015a)
  - Morphosyntactic development (Rakhlin, Kornilov, & Grigorenko, 2014)
  - Phonological development (Kavitskaya et al., 2011)
  - Social cognition and theory of mind (Rakhlin et al., 2011)
  - EEG/ERP indices of spoken word and attentional but not pre-attentive auditory processing (Kornilov et al., 2015a, 2015b)

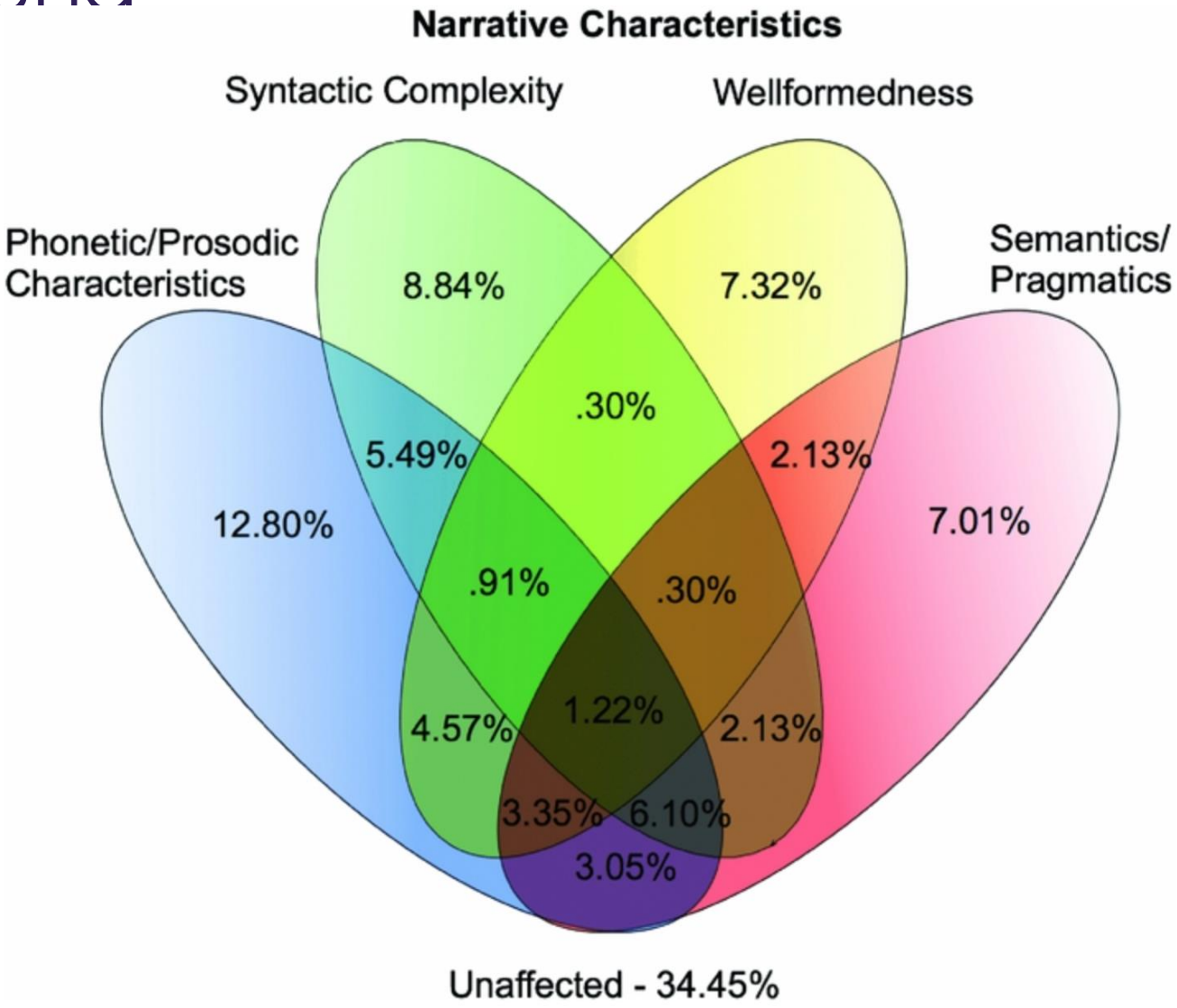
Rakhlin, N., Kornilov, S.A., Palejev, D., Koposov, R.A, Chang, J.T., & Grigorenko, E.L. (2013). The Language Phenotype of a Small Geographically Isolated Russian-speaking Population: Implications for Genetic and Clinical Studies of Developmental Language Disorder. *Applied Psycholinguistics*, 34(5), 971-1003

Phenotype does not breed true



Rakhlin, N., Kornilov, S.A., Palejev, D., Koposov, R.A, Chang, J.T., & Grigorenko, E.L. (2013). The Language Phenotype of a Small Geographically Isolated Russian-speaking Population: Implications for Genetic and Clinical Studies of Developmental Language Disorder. *Applied Psycholinguistics*, 34(5), 971-1003

All combinations of deficits are found



Rakhlin, N., Kornilov, S.A., Palejev, D., Koposov, R.A, Chang, J.T., & Grigorenko, E.L. (2013). The Language Phenotype of a Small Geographically Isolated Russian-speaking Population: Implications for Genetic and Clinical Studies of Developmental Language Disorder. *Applied Psycholinguistics*, 34(5), 971-1003

Aims of the study

- Perform a genome-wide association study of DLD in AZ
  - Microarray SNP panel (Illumina's 370k-Duo)
- Using a multivariate set of naturalistic, ecologically-valid phenotypes (elicited speech samples)
- Explore the role of other types of genetic variation (CNVs, ROHs, rare variants) in DLD in AZ
  - Microarray SNP panel (Illumina's 370k-Duo)
  - Whole exome sequencing of severely affected probands (75 bp paired-ended, HiSeq 2500)
- Illuminate the role of potential novel DLD candidate genes in the etiology of the disorder using neurophysiological endophenotypes
  - Network connectomics (graph theory measures)

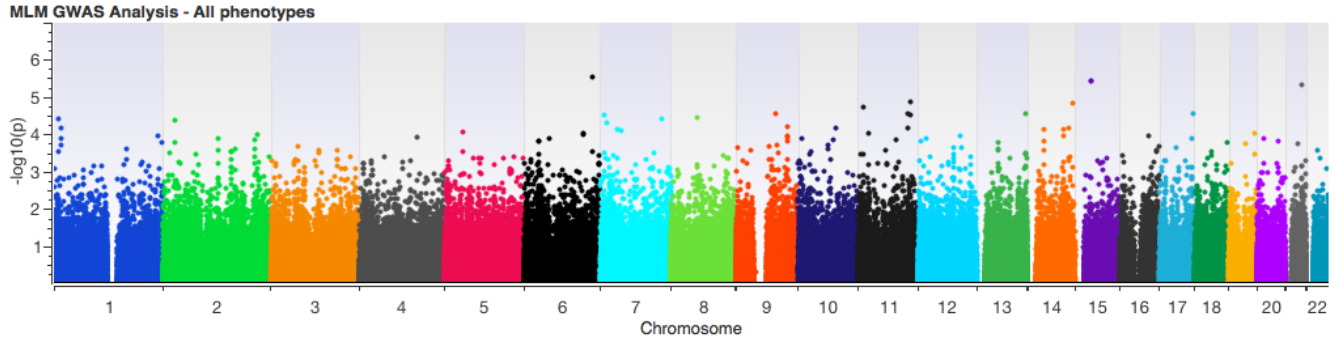
Kornilov, S.A., Rakhlin, N., Koposov, R., Lee, M., Yrigollen, C., Caglayan, A., Magnuson, J.S., Mane, S., Chang, J., & Grigorenko, E.L. (Under revision). A genome-wide association and whole-exome sequencing study of Developmental Language Disorder in a geographically isolated population.

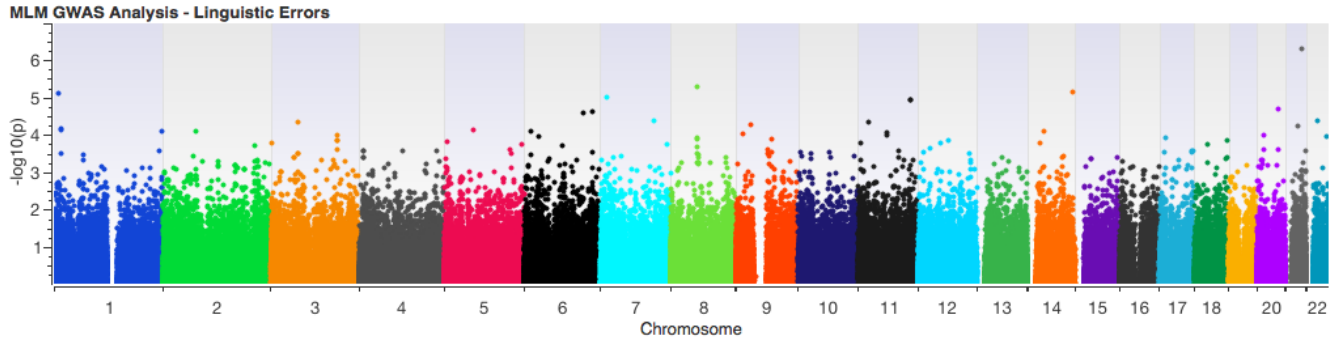
Methods

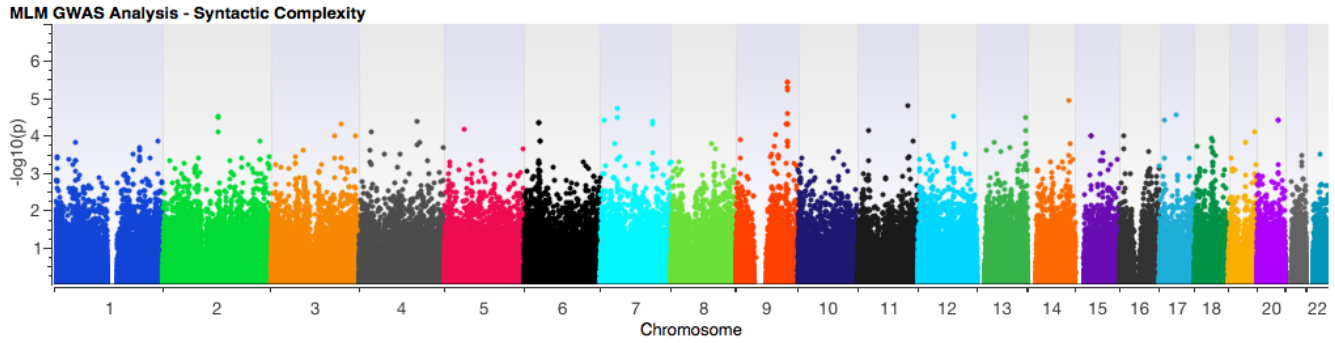
- Sample
  - 359 AZ residents (124 children and 235 adults)
- Phenotype
  - Linguistic Errors (Phonetic/Prosodic Characteristics, Wellformedness, Semantic/Pragmatic Errors)
  - Syntactic Complexity (MLUw, Complex Structures)
- Specimens
  - DNA extracted from peripheral blood, saliva, or buccal cells
- Genotyping
  - Illumina's Human CNV 370k-Duo BeadChip (370k SNPs)
  - QCed down to 223,580
  - Most processing and analyses performed in SVS (GoldenHelix, Inc)
- Association testing
  - SNP - EMMAX and GEMMA, controlling for age, gender, rank-normalized phenotypes
  - Gene-based testing – Hybrid Set-Based Test (HYST)
  - CNVs – FBAT CNV, CNVRuler/Conan/regression
  - ROHs – logistic and linear regression

Association findings: SNP

- A multitude of intriguing individual SNP findings, yet none survived corrections for multiple testing

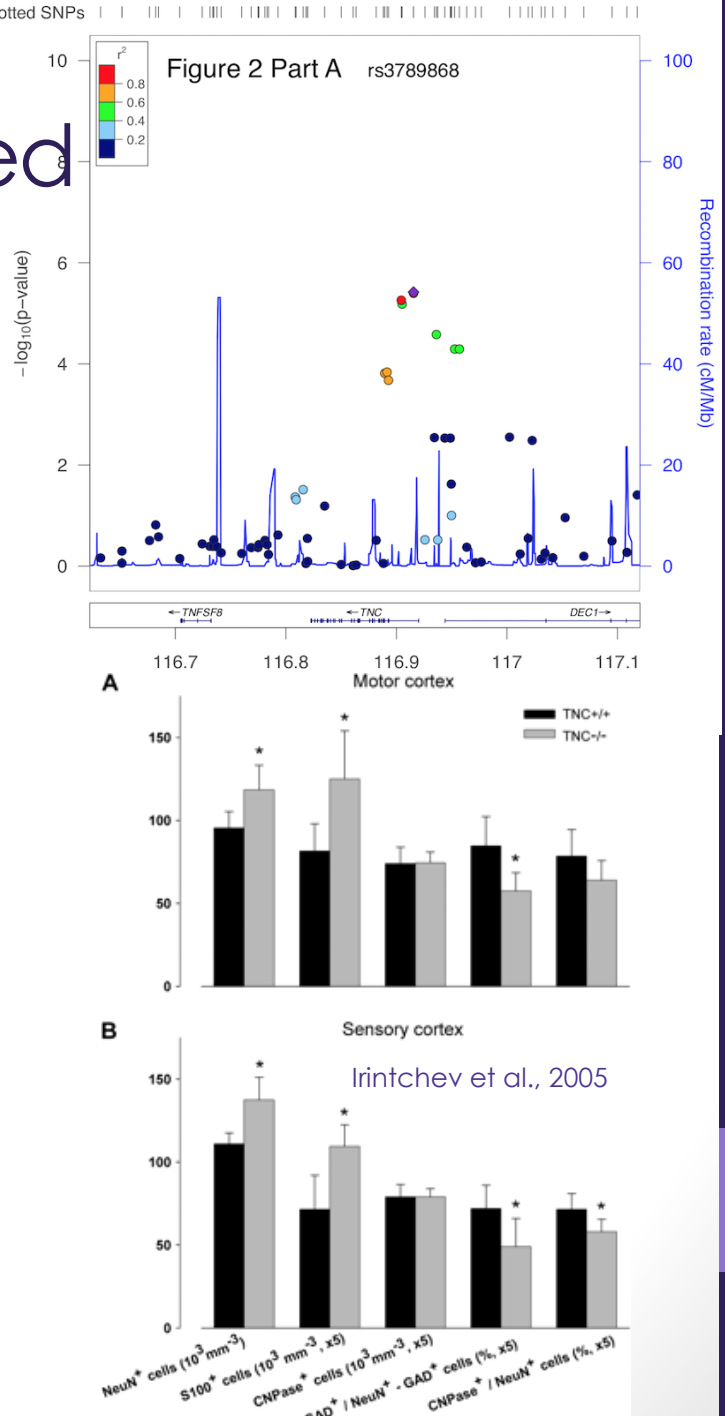






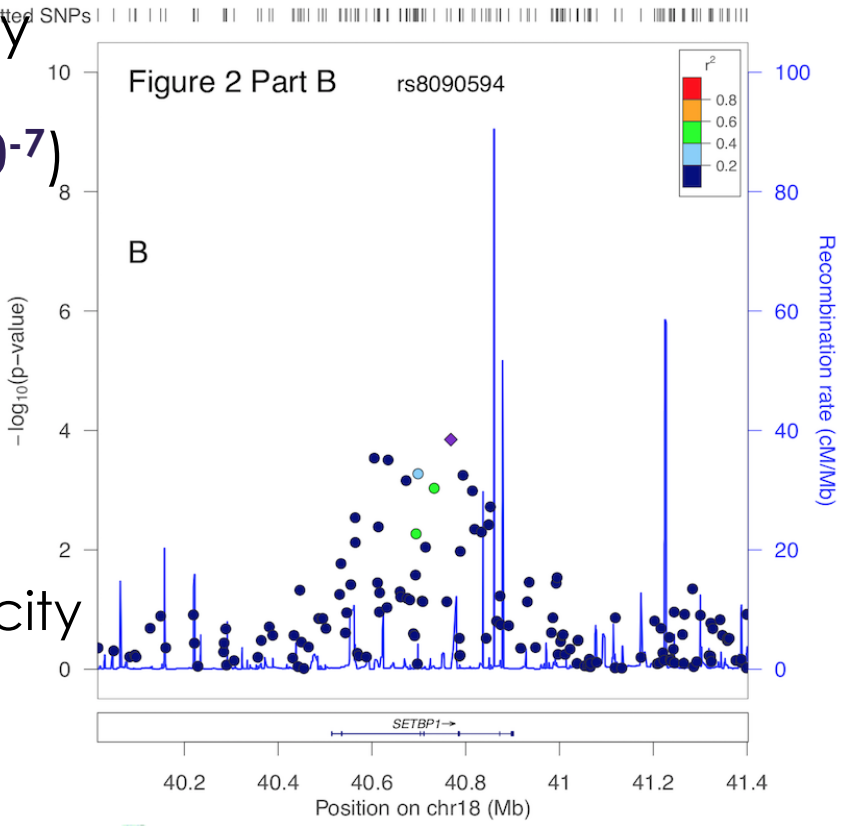
Associations: Gene-based

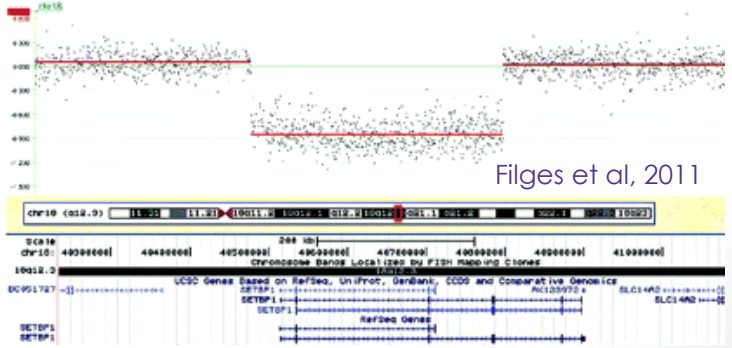
- Multiple SNPs in the *TNC* (9q33) gene were in the top set of hits for Syntactic Complexity
  - TNC codes for extracellular matrix protein tenascin – involved in neural development
  - Involved in autosomal dominant deafness
  - Plays a role in cochlear development
  - TNC-deficient mice show structural and functional cortical abnormalities – e.g., higher neuronal density, abnormal dendrite morphogenesis (Gurevicius et al., 2009; Irintchev et al., 2005)



Association findings: *SETBP1*

- *SETBP1* (18q21) is significantly associated with Syntactic Complexity (at $p = 5.47 \times 10^{-7}$)
- Little known about the function, but
  - Involved in DNA replication, apoptosis, and transcription regulation
  - Inhibits proteins involved in regulation of synaptic plasticity (Colbran, 2004)
  - Mutations cause Schinzel-Giedion syndrome (MIM#269150)
  - Haploinsufficiency associated with severe DLD (Bouquillon et al., 2011; Filges et al., 2011; Marseglia et al., 2012)





Whole Exome Sequencing

- 12 severely affected DLD probands
- Illumina's HiSeq 2500 platform (Illumina, Inc)
- NimbleGen EZ Exome SeqCap v2 (Roche Nimblegen), 75 bp paired-end
- Aligned to hg19 using Novoalign
- Variant calling performed using GATK (DePristo et al., 2011)
- Prioritization and filtration done using SVS (GoldenHelix, Inc) and eXtasy (variant prioritization by genomic data fusion - pathogenicity, haploinsufficiency prediction, similarity to other candidate genes; Sifrim et al., 2013)

Kornilov, S.A., Rakhlin, N., Koposov, R., Lee, M., Yrigollen, C., Caglayan, A., Magnuson, J.S., Mane, S., Chang, J., & Grigorenko, E.L. (Under revision). A genome-wide association and whole-exome sequencing study of Developmental Language Disorder in a geographically isolated population.

Exome sequencing findings

- 14 coding sequence variants (each is carried by at least 4 out of 12 probands)
- Genes that regulate neural development: frameshift mutations in *NT5DC2*, missense SNVs in *NECAB1, ILK*
- Missense variant in *CDH2* – 7/12 in AZ but only 2% in the 1000 Genomes dataset
- Stop gain variant in *TCP10L2*
- Missense SNV in *TRIP6* (7q22.1) and a frameshift deletion in *ENTHD1* (22q13) – regulate post-natal neural stem cell maintenance and synaptic vesicle endocytosis at nerve terminals
- Possible commonality of the pathway with SETBP1 findings (common regulator)

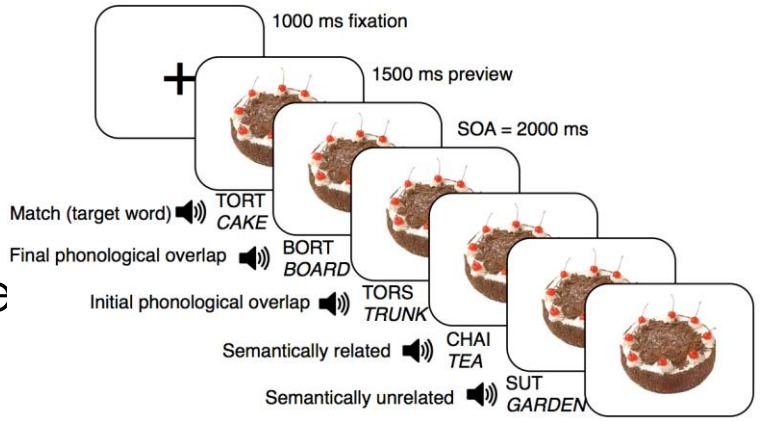
Kornilov, S.A., Rakhlin, N., Koposov, R., Lee, M., Yrigollen, C., Caglayan, A., Magnuson, J.S., Mane, S., Chang, J., & Grigorenko, E.L. (Under revision). A genome-wide association and whole-exome sequencing study of Developmental Language Disorder in a geographically isolated population.

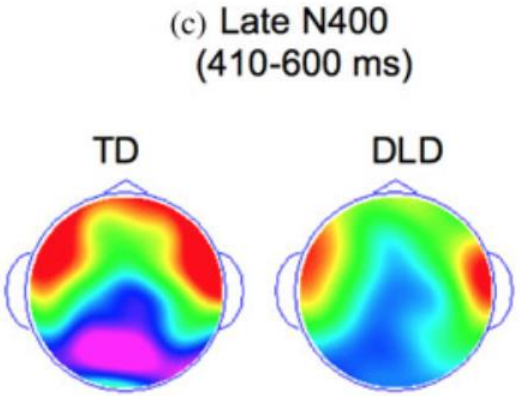
SETBP1 and neural development

- Neurodevelopmental disorders are associated with dysregulation of cortical networks – e.g., ASD, ADHD, ID (Kleinhans et al., 2008; Hoekzema et al., 2014)
- Connectivity patterns are under strong genetic control (23-89%; Shutte et al., 2013; Smit et al., 2008)
- We examined the role of SETBP1 in DLD in AZ by
  - Performing a targeted association study
  - In the mixed sample of affected and unaffected AZ children (n=39)
  - Using EEG/ERP indices of neural activation during word processing
  - **Network connectomics**: using higher-order properties of the functional intracortical networks (e.g., cohesion) as endophenotypes

SETBP1 and neural development: Methods

- Five SNPs, including rs8085464
  - (*p* = .0004 in the GWAS)
- EEG recorded from 64 electrode in the picture expectation paradigm (Kornilov et al., 2015)
- Estimated functional lagged coherence between 18 language ROIs in the intracortical space using eLORETA
- Used graph theory to evaluate properties of networks (cohesion, average path length, transitivity)
- In six EEG bands
- Corrected for multiple comparisons (Bonferroni)





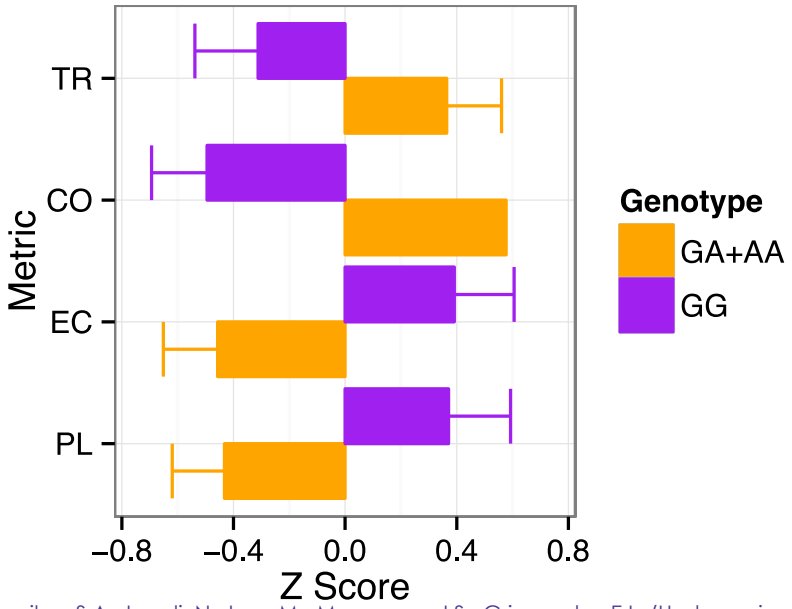
Kornilov et al., 2015

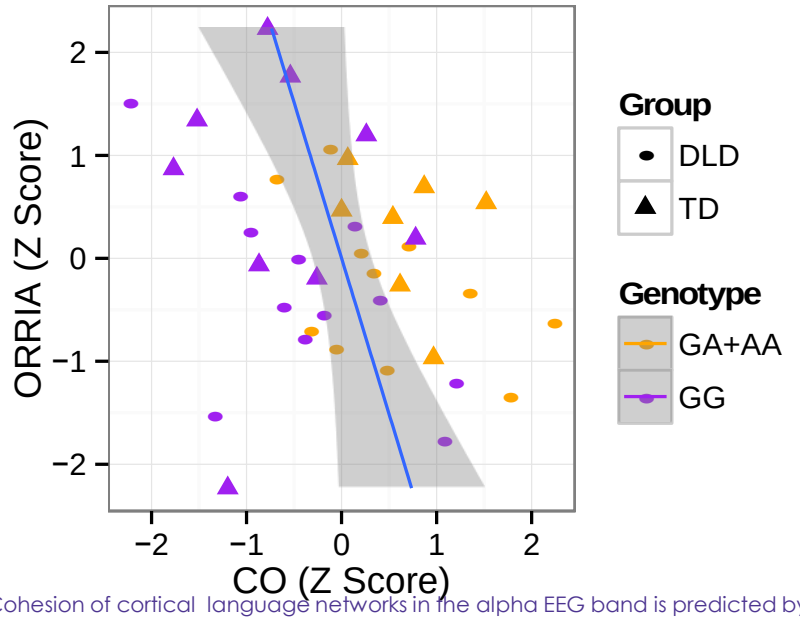
Kornilov, S.A., Magnuson, J.S., Rakhlin, N., Landi, N., & Grigorenko, E.L. (2015). Lexical processing deficits in children with developmental language disorder: An event-related potentials study. Development and Psychopathology, 27, 459-476. doi:10.1017/S0954579415000097

Kornilov, S.A., Landi, N., Lee, M., Magnuson, J.S., Grigorenko, E.L. (Under review). Cohesion of cortical language networks in the alpha EEG band is predicted by a common polymorphism in the *SETBP1* gene.

SETBP1 and neural development

- rs8085464 explained 26% of variance in language network cohesion (corrected p = .0145) in the alpha EEG band
- Each copy of the minor allele increased coherence by ¾ SD units
- Highly conserved (GERP = 340)

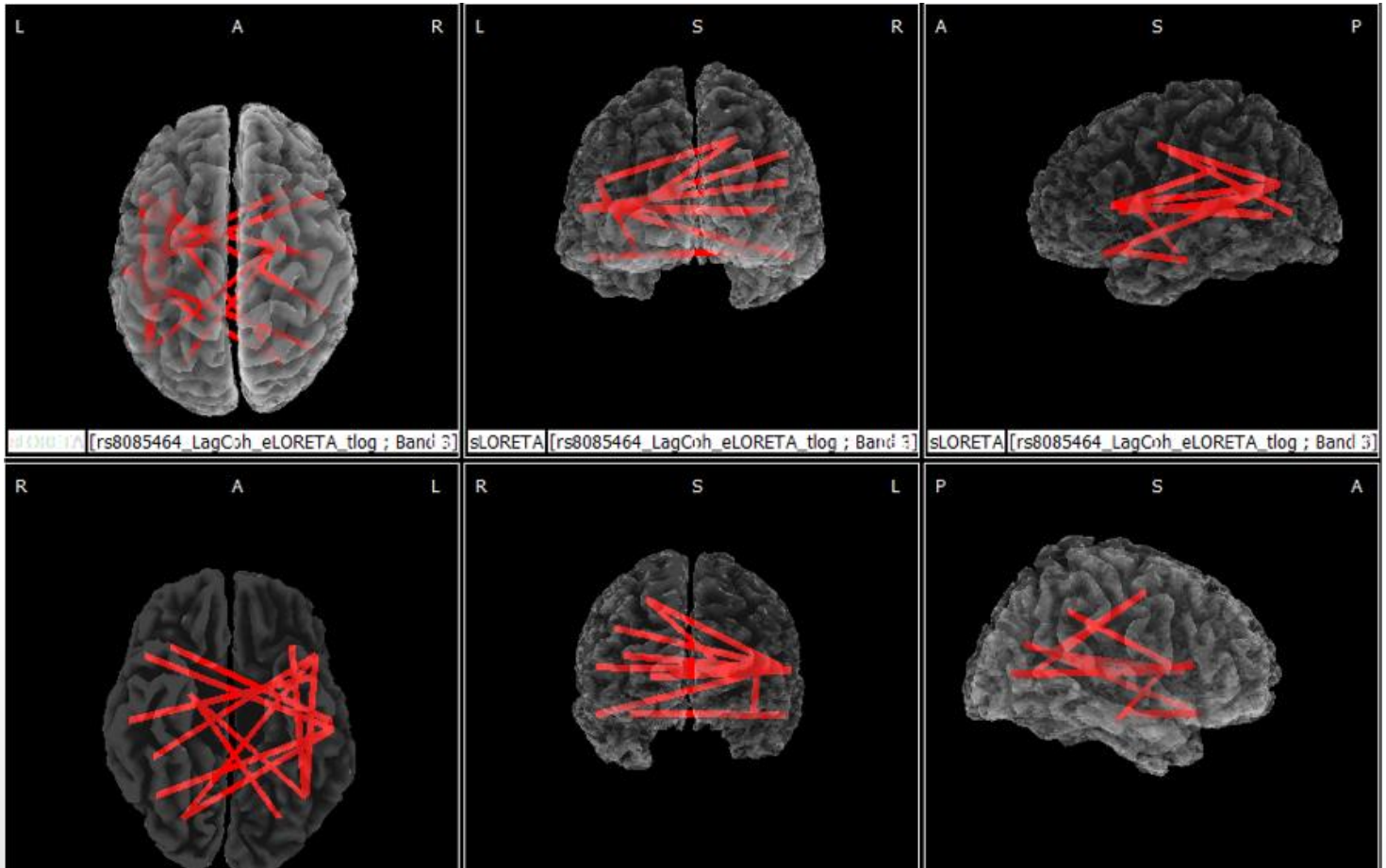




Kornilov, S.A., Landi, N., Lee, M., Magnuson, J.S., Grigorenko, E.L. (Under review). Cohesion of cortical language networks in the alpha EEG band is predicted by a common polymorphism in the *SETBP1* gene.

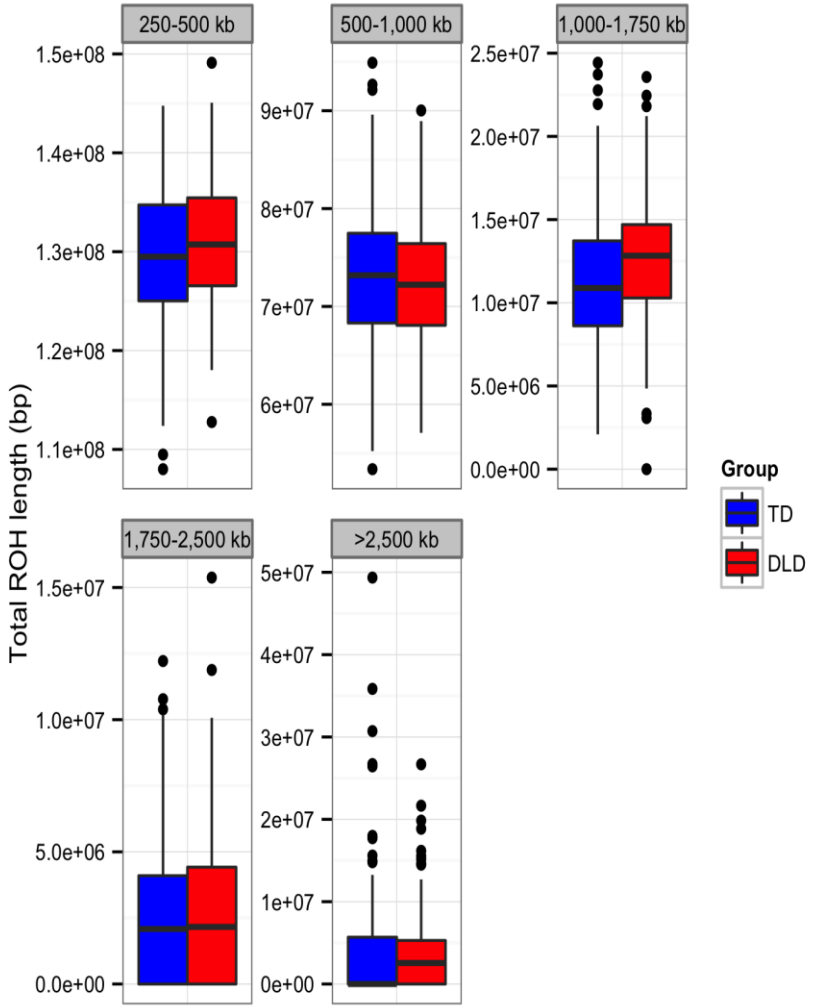
SETBP1 and neural development

- Increased cohesion stems from overconnectivity
  - Eleven pairwise coherences were stronger in minor allele carriers
  - Interhemispheric rather than intrahemispheric
    - Left auditory cortex and pars triangularis



Other findings

- Although (F)(P)BAT-CNV analyses suggested the presence of several highly-significant CNVRs associated with DLD in AZ, qPCR did not validate these
- Cumulative ROH burden was associated with DLD status in AZ ($p < .001$), yet, no single ROH region reached genome-wide significance after corrections for multiple testing



Kornilov, S.A., Rakhlin, N., Koposov, R., Lee, M., Yrigollen, C., Caglayan, A., Magnuson, J.S., Mane, S., Chang, J., & Grigorenko, E.L. (Under revision). A genome-wide association and whole-exome sequencing study of Developmental Language Disorder in a geographically isolated population.

Conclusion

- DLD is a multivariate, multi-factorial complex disorder
- Even in the presence of reduced background heterogeneity, DLD is likely regulated by multiple genetic pathways/genes and even types of variants
- A GWAS study of DLD in an isolated AZ population revealed *SETBP1* as a novel candidate DLD gene
- *SETBP1* seems to regulate the development of cortical networks involved in language processing
- Pending replication in independent samples and functional studies

Kornilov, S.A., Rakhlin, N., Koposov, R., Lee, M., Yrigollen, C., Caglayan, A., Magnuson, J.S., Mane, S., Chang, J., & Grigorenko, E.L. (Under revision). A genome-wide association and whole-exome sequencing study of Developmental Language Disorder in a geographically isolated population.

Thank you!


Elena Grigorenko


Sergey Kornilov


Maria Lee


Joseph Chang


Lesley Hart


Roman Koposov


Natalia Rakhlin


James Magnuson


Nicole Landi

- **EGLab** (Yale University, Child Study Center, School of Medicine) http://www.yale.edu/eglab
- Funding agencies
  - National Institutes of Health Grant R01 DC007665 (E.L.G., Principal Investigator)
  - National Science Foundation Grant CAREER 0748684 (J.S.M., Principal Investigator)
  - National Science Foundation IGERT Training Grant 114399 (J.S.M., Principal Investigator)
  - Government of the Russian Federation (grant No 14.Z50.31.0027; E.L.G., Principal Investigator)
- GoldenHelix, Inc
- Northern State Medical Academy, Arkhangelsk, Russia
- We thank the children who participated in the study and their families for their cooperation and the local medical officials of the AZ community for their help with data collection.

Kornilov, S.A., Rakhlin, N., Koposov, R., Lee, M., Yrigollen, C., Caglayan, A., Magnuson, J.S., Mane, S., Chang, J., & Grigorenko, E.L. (Under revision). A genome-wide association and whole-exome sequencing study of Developmental Language Disorder in a geographically isolated population.