Neuron

CellPress

Previews

# Advancing discovery of risk-altering variants for complex diseases by functionally informed fine-mapping

You Chen,[1,2] Andrew G. Clark,[1,3] and Haiyuan Yu[2,3,*]
[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[2]Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY 14853, USA
[3]Department of Computational Biology, Cornell University, Ithaca, NY 14853, USA
*Correspondence: haiyuan.yu@cornell.edu
https://doi.org/10.1016/j.neuron.2022.02.018

**Pinpointing causal variants at risk loci identified by genome-wide association studies (GWAS) has been a great challenge. In this issue of *Neuron*, Zhang et al. present a fine-mapping approach, RefMap, integrating functional genomics with GWAS summary statistics to prioritize causal variants for amyotrophic lateral sclerosis.**

Despite the great success in robust mapping of numerous genetic loci contributing to complex diseases, genome-wide association studies (GWAS) have been less successful in determining the underlying causal variants at these risk loci, motivating a continued effort to develop fine-mapping methods (heuristic approaches, penalized regression models, and Bayesian methods) (Schaid et al., 2018). The power and resolution of these methods is limited by the correlated presence of nearby variants (linkage disequilibrium, LD), but can be improved by integrating functional annotations enriched for disease-related phenotypes. There are often tens to hundreds of variants in high LD with the reported associated single nucleotide polymorphisms (SNPs), any of which can be potentially causal (Hormozdiari et al., 2014). An *ad hoc* review of genomic annotations is often applied to SNPs selected by fine-mapping methods, which could be cumbersome and biased. An alternative approach is to integrate functional genomics as prior information for Bayesian fine-mapping methods (Weissbrod et al., 2020). Given that over 90% of GWAS-identified variants lie in non-coding regions, and that they are enriched in regulatory elements that can be cell type specific (Maurano et al., 2012), Zhang et al. (2022) developed a Bayesian network, RefMap, which utilizes epigenetic profiling of induced pluripotent stem cell (iPSC)-derived motor neurons (MNs), the key cell type for the pathogenesis of amyotrophic lateral sclerosis (ALS), to determine the prior probability of disease association. With integration of functional genomics and GWAS summary statistics (Figure 1), RefMap identified 690 ALS-associated genes, which represents a 5-fold increase in recovered SNP heritability compared to conventional methods. Functional characterization of top candidate genes indicates that the cascade of ALS pathogenesis is initiated by distal axonal dysfunction, upstream of TDP-43 mislocalization, which contradicts the previous view that axonal dysfunction is secondary to TDP-43 pathology (Briese et al., 2020).

To simplify fine-mapping, the genome was first partitioned into 1-Mb blocks, and each block was further divided into 1-kb subregions with varying numbers of SNPs. Each region was assigned a disease-association score based on the SNPs within it and their Z-scores estimated from a published ALS GWAS (van Rheenen et al., 2016). Next, the disease-association score for each region was weighted by an epigenetic score, which was calculated from a linear combination of the overlapping ratios of that region with the peaks of ATAC-seq and histone ChIP-seq (H3K27ac, H3K4me1, and H3K4me3). The presence of accessible chromatin (measured by ATAC-seq) and the enrichment of histone modifications (assayed by ChIP-seq) are commonly used to define active regulatory elements. The 1-kb regions with final weighted scores (designated as Q-scores) over the cutoff of 0.95 were defined as significant regions. The authors further mapped these ALS-associated regions with their target genes expressed in MNs. A region was mapped to a particular gene if one of two criteria was met: (1) the region overlapped the gene or the ±10-kb area flanking that gene; or (2) the region overlapped a loop anchor harboring the transcription start site (TSS) of that gene based on Hi-C data of iPSC-derived MNs.

The resulting 690 ALS-associated genes were found to be enriched with clinically reportable ALS genes (ClinVar) and an independently curated list of ALS genes including previous GWAS loci. For those missing from the RefMap gene list, the authors hypothesized that there might be a functional overlap via protein-protein interactions (PPIs). They mapped RefMap genes onto a high-confidence human PPI network and demonstrated a shorter average path distance between novel RefMap genes and known ALS genes than expected by chance. In contrast, genes identified by other fine-mapping tools (MAGMA, Pascal, and PAINTOR) failed to overlap significantly with previously reported ALS genes.

Furthermore, variation in RefMap genes accounted for 36% of the SNP heritability estimated by LD score regression, a 5-fold gain over previous reports. Applying RefMap to another ALS GWAS dataset (van Rheenen et al., 2021) replicated the

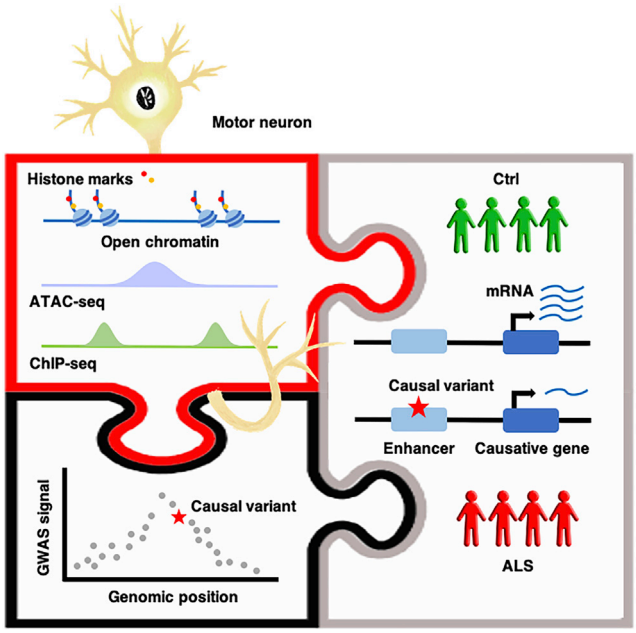


**Figure 1. Cell-type-specific regulatory profiling is the key missing puzzle piece for gene discovery of complex diseases**

To resolve the mismatch between the high heritability of ALS and the relatively low number of known causative genes, Zhang et al. (2022) developed a new fine-mapping model, RefMap, that integrates GWAS summary statistics with epigenetic profiling data (ATAC-seq and ChIP-seq) from a disease-relevant cell type (motor neurons). The variants in regulatory elements result in downregulation of RefMap genes and eventually contribute to the pathogenesis of ALS.

great majority of candidate genes. These genes were also found to be significantly enriched for being haploinsufficient (haploinsufficiency scores) and less tolerant to loss-of-function mutations (LoFtool and o/e scores) and other types of mutations (RVIS scores), compared to all protein-coding genes. PPI network analysis identified two modules (M421 and M604) enriched with RefMap genes, both of which were associated with the distal axon and neuromuscular junction, consistent with the notion of ALS as a distal axonopathy. In iPSC-derived MNs from neurologically normal individuals, RefMap genes had higher expression levels compared to the overall transcriptome.

To further validate the functional significance of these RefMap genes, the authors investigated transcriptome changes in cases versus controls, assuming that ALS-associated variants in regulatory regions may contribute to the pathogenesis by altering expression level of the target genes. Re-analysis of two published RNA-seq datasets (postmortem brain tissues; iPSC-derived MNs) confirmed that RefMap genes in ALS patients were significantly downregulated compared to those in controls. The longitudinal transcriptomic profiling of the SOD1-G93A-ALS mouse model also identified a distinct group of RefMap homologs that were progressively downregulated through consecutive disease stages and enriched for distal axon function, overlapping with the functions ascribed to M421 and M604 modules.

Given the reduced expression of RefMap genes in ALS patients, the authors hypothesized that rare mutations that can alter the expression of these genes would also increase the risk of developing ALS. Burden testing of rare missense variants in an exome sequencing dataset (3,864 ALS patients and 7,839 controls) demonstrated a lower median p value for ALS-associated genetic burden in RefMap genes than expected by chance. The analysis of an independent whole-genome sequencing cohort (5,594 ALS patients and 2,238 controls) revealed that patients with a higher burden of rare loss-of-function mutations within RefMap genes had earlier ages of disease onset.

Genetic convergence between common and rare variant analysis highlighted three promising causal genes (*BNC2*, *KANK1*, and *ADAMTSL1*) and one protective gene (*VAV2*), none of which had been linked to ALS previously. Expression levels of all these genes in iPSC-derived MNs from ALS patients were associated with the ages of onset. Knockdown of *BNC2* and *KANK1* by CRISPR resulted in reduced neuronal viability, shorter neurites, and reduced branch length in cultured neurons differentiated from SH-SY5Y cells. To further elucidate the order of key events in the pathogenic cascade, the authors engineered loss of *KANK1* function in iPSC-derived MNs from an aged healthy control. Increased apoptosis and nuclear fragmentation were observed in *KANK1*-edited MNs, indicating neuronal toxicity caused by loss of *KANK1* function. At day 40 post-differentiation, *KANK1*-edited MNs demonstrated electrophysiological dysfunction, reflected in a moderate increase in input resistance and reduced resting membrane potential and whole-cell capacitance. At day 45, *KANK1*-edited MNs exhibited dramatic loss of nuclear TDP-43 and aggregation of cytoplasmic TDP-43-positive proteins, a pathological hallmark of ALS. Overall, their data placed axonal dysfunction upstream of TDP-43 mislocalization in the cascade of pathogenesis, suggesting that *KANK1* upregulation might be a therapeutic target for ALS patients harboring *KANK1* mutations.

This work adds to the rapidly growing toolbox for GWAS fine-mapping, which prioritizes variants for functional follow-up and yields insights into pathogenesis of complex diseases. More work will be needed to explore the generalizability and applicability of RefMap using molecular profiling of disease-relevant cell types for gene discovery of other complex diseases. Leveraging diverse types of functional annotations may be one approach to increase the power and resolution of fine-mapping methods like RefMap. Altogether, this work illustrates the power of incorporating cell-type-specific molecular profiling into statistical methods that prioritize causal variants at GWAS risk loci and sheds light on the pathogenic mechanisms underlying complex diseases.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (UM1 HG009393 and R01 DK115398 to H.Y.).

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

Briese, M., Saal-Bauernschubert, L., Lüningschrör, P., Moradi, M., Dombert, B., Surrey, V., Appenzeller, S., Deng, C., Jablonka, S., and Sendtner, M. (2020). Loss of Tdp-43 disrupts the axonal transcriptome of motoneurons accompanied by impaired axonal translation and mitochondria function. Acta Neuropathol. Commun. *8*, 116. https://doi.org/10.1186/s40478-020-00987-6.

Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. Genetics *198*, 497–508. https://doi.org/10.1534/genetics.114.167908.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195. https://doi.org/10.1126/science.1222794.

Schaid, D.J., Chen, W., and Larson, N.B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat. Rev. Genet. *19*, 491–504. https://doi.org/10.1038/s41576-018-0016-z.

van Rheenen, W., Shatunov, A., Dekker, A.M., McLaughlin, R.L., Diekstra, F.P., Pulit, S.L., van der Spek, R.A., Võsa, U., de Jong, S., Robinson, M.R., et al.; PARALS Registry; SLALOM Group; SLAP Registry; FALS Sequencing Consortium; SLAGEN Consortium; NNIPPS Study Group (2016). Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat. Genet. *48*, 1043–1048. https://doi.org/10.1038/ng.3622.

van Rheenen, W., van der Spek, R.A.A., Bakker, M.K., van Vugt, J.J.F.A., Hop, P.J., Zwamborn, R.A.J., de Klein, N., Westra, H.J., Bakker, O.B., Deelen, P., et al.; SLALOM Consortium; PARALS Consortium; SLAGEN Consortium; SLAP Consortium (2021). Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. Nat. Genet. *53*, 1636–1648. https://doi.org/10.1038/s41588-021-00973-1.

Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., Schoech, A.P., van de Geijn, B., Reshef, Y., Márquez-Luna, C., et al. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nat. Genet. *52*, 1355–1363. https://doi.org/10.1038/s41588-020-00735-5.

Zhang, S., Cooper-Knock, J., Weimer, A.K., Shi, M., Moll, T., Marshall, J.N.G., Harvey, C., Nezhad, H.G., Franklin, J., Souza, C.D.S., et al.; Project MinE ALS Sequencing Consortium (2022). Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis. Neuron *110*, 992–1008.

# Serotonergic fast lane from taste detection to preparatory digestive actions

**Tongtong Wang,[1] Allan-Hermann Pool,[2] and Yuki Oka[1,*]**
[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA
[2]Department of Neuroscience, Department of Anesthesiology and Pain Management, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, TX, USA
*Correspondence: yoka@caltech.edu
https://doi.org/10.1016/j.neuron.2022.02.023

**Nutrient detection through the taste system triggers various physiological changes in the body. In this issue of *Neuron*, Yao and Scott (2022) identify two distinct classes of serotonergic neurons in *Drosophila* that transform sweet and bitter taste signals into endocrine and digestive responses.**

Just a bite of the food triggers a cascade of responses from salivary secretion and intestinal movement to gastric acid secretion, getting our body ready for the feast. How does the brain swiftly orchestrate these physiological responses upon sensory detection? Currently, the circuit-level mechanisms that achieve these functions remain mostly obscure. Yao and Scott now uncover the role of the serotonergic system in taste-driven priming of the endocrine and digestive systems (Yao and Scott, 2022).

Traditional view of nutrient homeostasis is a feedback model in which nutrient depletion drives ingestion and nutrient absorption quenches appetite. Besides this basic mechanism, several recent studies have shed light on startling feedforward mechanisms that peripheral nutrient detection rapidly modulates the activity of brain appetite circuits (Augustine et al., 2020). These feedforward signals enable animals to optimize internal milieu from as early as foraging and gustatory evaluation of food. To date, many studies have investigated how the brain activity is regulated by peripheral sensory signals. But much less attention has been paid to how the brain controls various physiological responses in a feedforward fashion.

Taste is the primary sensory modality for assessing nutrient quality. The taste system has similar coding logic between vertebrates and invertebrates. Specialized primary gustatory neurons are tuned to detect one of the basic taste qualities. Pathways for the two most salient tastes, sweet and bitter, are segregated from sensory input to higher relays in the brain areas. Past studies suggested that taste detection itself drives rapid physiological changes. For example, sweet taste signals trigger insulin release preceding any blood glucose change, preparing animals for a surge of food ingestion and facilitating