Human Genetics and Genomic Medicine

The amino-terminal acetylation of proteins, with Max Doerfel, Yiyang Wu, Jonathan Crain, and collaborators, including Jim Lupski (Baylor), Linyan Meng (Baylor), Joseph Wu (Stanford), Thomas Arnesen (Norway), Goo Taeg Oh (Korea) and Jozef Gecz (Australia).

We previously identified the genetic basis of a X-linked, infantile lethal Mendelian disorder, involving a p.S37P missense mutation in *NAA10*, encoding the catalytic subunit of NatA, involved in amino terminal-acetylation (NTA) of proteins. To date, there are >20 individuals with damaging mutations in *NAA10* or *NAA15,* with the latter gene encoding the dimeric binding partner for Naa10. Both *NAA10* and *NAA15* are relatively intolerant of variation. Phenotypic features of patients carrying disruptive variation in NatA include craniofacial anomalies, hypotonia, global developmental delays, cardiac arrhythmia, and/or cardiomyopathy. Functional analysis of the *NAA10* missense mutations has demonstrated impaired biochemical activity and/or a reduced capacity to form a stable NatA complex, although there is no proteome-wide decrease in NTA, thus implying tissue-specific and/or temporal substrate-specific effects. Attempts to knockdown or knockout *NAA10* in human cells results in cell death, making it difficult to assess proteome-wide NTA in human KO cells. Mice with decreased or absent *Naa10* survive throughout embryogenesis, but display a range of phenotypes that are variably expressive and include increased early neonatal lethality (likely related to congenital heart defects), supernumerary thoracic rib and vertebrae, piebaldism, and/or urogenital and renal abnormalities. Some of these phenotypes overlap with that seen with the missense mutation in humans, but survival of the knockout mice was unexpected, given the mutation intolerance of this gene in humans. Extensive proteomic analysis of mouse embryonic fibroblasts has not shown any proteome-wide decrease in NTA in the KO cells, with one possible explanation being our discovery of a previously unannotated *NAA10*-like gene in mice, which lacks a human ortholog, and which we refer to as *NAA12*. A rabbit polyclonal antibody raised with specificity to Naa12 shows expression of Naa12 in many mouse tissues. We used CRISPR to create indels in this new putative *NAA12*, and we are currently characterizing these knockout mice, along with breeding them to make the double knockout for *NAA10* and *NAA12*. Given the potential redundancy in mice, fibroblasts from one affected human proband control were reprogrammed into induced pluripotent stem-cells and differentiated into cardiomyocytes (CMs). Electrophysiological experiments show rate-dependent prolongation of the effective refractory period in patient CMs, suggesting a cell-autonomous defect in the CMs.

The auxiliary subunit of the NatA complex, NAA15, is the dimeric binding partner for NAA10. Through a genotype-first approach with whole-exome or genome sequencing (WES/WGS) and targeted sequencing analysis, we identified and phenotypically characterized 37 individuals from 32 unrelated families with 25 different *de novo* or inherited, dominantly acting likely gene disrupting (LGD) variants in *NAA15.* Clinical features of affected individuals with LGD variants in *NAA15* include variable levels of intellectual disability, delayed speech and motor milestones, and autism spectrum disorder. Additionally, mild craniofacial dysmorphology, congenital cardiac anomalies and seizures are present in some subjects. RNA analysis in cell lines from two of the patients showed degradation of the transcripts with LGD variants, likely due to nonsense mediated decay. Functional assays in yeast confirmed a deleterious effect for two of the LGD variants in *NAA15*. We propose that defects in NatA-mediated N-terminal acetylation (NTA) lead to variable levels of neurodevelopmental disorders in humans, supporting the importance of the NatA complex in normal human development.

Scikit-ribo: Accurate estimation and robust modeling of translation dynamics at codon resolution, with Han Fang, Yi-Fei Huang, Aditya Radhakrishnan, Adam Siepel, and Michael Schatz.

Ribosome profiling (Riboseq) is a powerful technique for measuring protein translation, however, sampling errors and biological biases are prevalent and poorly understand. Addressing these issues, we developed Scikit-ribo (https://github.com/schatzlab/scikit-ribo), the first open-source software for accurate genome-wide A-site prediction and translation efficiency (TE) estimation from Riboseq and RNAseq data. Scikit-ribo accurately identifies A-site locations and reproduces codon elongation rates using several digestion protocols. We showed that commonly used RPKM-derived TE estimation is prone to biases, especially for low-abundance genes. Scikit-ribo introduces a codon-level generalized linear model with ridge penalty that correctly estimates TE while accommodating variable codon elongation rates and mRNA secondary structure. This corrects the TE errors for over 2000 genes in *S. cerevisiae*, which we validated using mass spectrometry of protein abundances and allows the determination of the Kozak-like sequence directly from Riboseq. We also completed an analysis of coverage requirements needed for robust codon-level analysis, and quantified the artifacts that can occur from cycloheximide treatment.

Outlier gene expression reveals recurrent dysregulation in rare disease pedigrees, with Sara Ballouz, Max Dörfel, Jonathan Crain, Megan Crow, and Jesse Gillis.

In disease expression analysis, looking for shared functional signals from a set of genes which exhibit differential expression is commonplace. We examine the complement as a possibility, that disease genes display "outlier" or unexpected expression relative to broader patterns of functional expression variation. Using six families from the rare *TAF1* syndrome disease cohort, we performed family-specific differential expression analyses and find that functional characterization of top candidates enriches for common pathways unlikely to be specifically linked to disease. However, by filtering away common expression changes using known co-expression, we lose all functional enrichment and are left with a small number of outliers characteristic of each proband. Two of these outlier genes are highly recurrent across pedigrees (FDR <2.63e-05) and are the primary commonality among the cohort as a whole. This suggests that systems analysis may be relevant to rare diseases principally as a means of filtering out biological signals unrelated to disease.

We have also continued to work on X-linked Dystonia-Parkinsonism (XDP), which is an adult-onset neurodegenerative disease endemic to the island of Panay, Philippines, where its prevalence is reported to be 5.74 cases per 100,000 individuals with a mean age at onset of 39.7 years. The documented clinical phenotype most frequently combines features of dystonia and parkinsonism in a characteristic temporal progression, beginning with hyperkinetic symptoms at early stages and progressing to predominantly hypokinetic movements at later stages. Various genetic studies have raised the possibility that a defect in *TAF1* may somehow underlie XDP pathogenesis. *TAF1* encodes TATA-Binding Protein (TBP)-Associated Factor-1 (*TAF1*), a component of the TFIID complex which mediates transcription by RNA polymerase II (RNAPII) and has emerged in recent years as a significant disease target. In addition to the XDP-related sequence variants in this gene, we reported coding variation in *TAF1* to be associated with severe neurodevelopmental defects and intellectual disability. Given the essential function of *TAF1* in transcriptional regulation in all cells, it is not known how these reported XDP-specific sequence variants may cause tissue-specific defects and/or highly specific clinical phenotypes. Given the apparently distinct phenotypic spectra between XDP, which we propose to be associated with a noncoding mechanism that results in partial reduction of *TAF1* in males, and the recently reported *TAF1* neurodevelopmental syndrome, we compared phenotypic data available between individuals in both cohorts. Videotaped clinical exams from five individuals from three of these families were evaluated by the movement disorder neurologists who characterized the clinical phenotypes in our XDP cohort. Two affected individuals (IV-3, age 21 years, and III-2, age 23 years) from the previously described extended family with a large CNV duplication involving *TAF1* had severe neurologic deterioration, including prominent oropharyngeal dysphagia, which is characteristic of XDP. They also displayed cervical and jaw-opening dystonias, which are cardinal features of XDP. The probands in the other families exhibited milder dystonic features, but two of these individuals (brothers from Family 1) were much younger in age (12 and 14 years old). We also explored the expression data from the XDP fibroblast and neural derivative cell models and from RNA isolated directly from blood from six TAF1 neurodevelopmental syndrome families that we previously reported, but detected no statistically significant overlap between DEGs among the different cohorts and cell types.

Expanding collection and sequencing of other rare genetic syndromes, with Yiyang Wu, Han Fang, Reid Robison (Utah), Kai Wang (California), David Goldstein (Columbia), Alan Rope (Oregon), Jim Lupski (Baylor), Jennifer Posey (Baylor), and others.

We continue to meet and collect many families in Utah and elsewhere with very rare, idiopathic genetic syndromes, which we are sequencing and analyzing with many groups. The total number of DNA samples collected to date is approaching 2000, and this includes detailed phenotyping information. We have been making extensive use of Human Phenotype Ontology terms, and the PI was an author on a review concerning the current progress with the development and integration of HPO in various research settings. The PI continues to publish case reports and cohort studies of various rare genetic syndromes.

Collaborating on genetics of Tourette Syndrome, with the Tourette Syndrome Association International Consortium for Genetics.

The PI continues to collaborate on this international effort to understand the genetics of Tourette Syndrome. Psychiatric comorbidity is common in Tourette syndrome (TS); when present, these conditions typically cause more distress and impairment than do tics. High rates of attention-deficit/hyperactivity disorder

(ADHD) and obsessive-compulsive disorder (OCD) are well documented and thought to be core components of the TS phenotype; however, few studies have fully characterized other comorbidities. We therefore continue to characterize the prevalence and impact of psychiatric comorbidity in a large sample of individuals with TS and their family members.

References

1. Hirschtritt ME, Darrow SM, Illmann C, Osiecki L, Grados M, Sandor P, Dion Y, King RA, Pauls D, Budman CL, et al: **Genetic and phenotypic overlap of specific obsessive-compulsive and attention-deficit/hyperactive subtypes with Tourette syndrome.** *Psychol Med* 2018, **48:**279-293.

2. Fang H, Huang YF, Radhakrishnan A, Siepel A, Lyon GJ, Schatz MC: **Scikit-ribo Enables Accurate Estimation and Robust Modeling of Translation Dynamics at Codon Resolution.** *Cell Syst* 2018.

3. Kohler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Ayme S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, et al: **The Human Phenotype Ontology in 2017.** *Nucleic Acids Res* 2017, **45:**D865-D876.

4. Huang AY, Yu D, Davis LK, Sul JH, Tsetsos F, Ramensky V, Zelaya I, Ramos EM, Osiecki L, Chen JA, et al: **Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome.** *Neuron* 2017, **94:**1101-1111 e1107.

5. Greenberg E, Tung ES, Gauvin C, Osiecki L, Yang KG, Curley E, Essa A, Illmann C, Sandor P, Dion Y, et al: **Prevalence and predictors of hair pulling disorder and excoriation disorder in Tourette syndrome.** *Eur Child Adolesc Psychiatry* 2017.

6. Fang H, Wu Y, Yang H, Yoon M, Jimenez-Barron LT, Mittelman D, Robison R, Wang K, Lyon GJ: **Whole genome sequencing of one complex pedigree illustrates challenges with genomic medicine.** *BMC Med Genomics* 2017, **10:**10.

7. Dorfel MJ, Fang H, Crain J, Klingener M, Weiser J, Lyon GJ: **Proteomic and genomic characterization of a yeast model for Ogden syndrome.** *Yeast* 2017, **34:**19-37.

8. Darrow SM, Hirschtritt ME, Davis LK, Illmann C, Osiecki L, Grados M, Sandor P, Dion Y, King R, Pauls D, et al: **Identification of Two Heritable Cross-Disorder Endophenotypes for Tourette Syndrome.** *Am J Psychiatry* 2017, **174:**387-396.

9. Darrow SM, Grados M, Sandor P, Hirschtritt ME, Illmann C, Osiecki L, Dion Y, King R, Pauls D, Budman CL, et al: **Autism Spectrum Symptoms in a Tourette's Disorder Sample.** *J Am Acad Child Adolesc Psychiatry* 2017, **56:**610-617 e611.