Human Genetics and Genomic Medicine

Gholson Lyon's lab focuses on analyzing human genetic variation and its role in severe neuropsychiatric disorders and rare diseases, including intellectual disability, autism, and schizophrenia. By recruiting large groups of related individuals living in the same geographic location (e.g., Utah), Lyon's lab can study the breadth and depth of genetic variants in a similar environmental background. Using the exome—the parts of the genome that code for protein—and whole-genome sequencing, the lab looks for mutations that segregate with syndromes in the various populations, and the lab undertakes comprehensive functional studies of many of the newly identified mutations. There are ~12 billion nucleotides in every cell of the human body, and there are ~25-100 trillion cells in each human body. Given somatic mosaicism, epigenetic changes and environmental differences, no two human beings are the same, particularly as there are only ~7 billion people on the planet. One of the next great challenges for studying human genetics will be to acknowledge and embrace complexity. Every human *is* unique, and the study of human disease phenotypes (and phenotypes in general) will be greatly enriched by moving from a deterministic to a more stochastic/probabilistic model. The dichotomous distinction between 'simple' and 'complex' diseases is completely artificial, and must consider a spectrum of diseases that are variably manifesting in each person. Comprehensive ancestry tracking and detailed family history data, when combined with whole genome sequencing or at least cascade-carrier screening, might eventually facilitate a degree of genetic prediction for some diseases in the context of their familial and ancestral etiologies. However, it is important to remain humble, as our current state of knowledge is not yet sufficient, and in principle, any number of nucleotides in the genome, if mutated or modified in a certain way and at a certain time and place, might influence some phenotype during embryogenesis or postnatal life. The Lyon lab continued to advance our research agenda with several publications in 2015 [1-8].

The amino-terminal acetylation of proteins, with Max Doerfel, Yiyang Wu, Jonathan Crain, and collaborators.

More than 85 % of human proteins are acetylated at their N-terminal amino group, hence, N-terminal acetylation (NTA) is one of the most abundant modifications of eukaryotic proteins. Despite its discovery more than 30 years ago, very little is known about the cellular effects/functions of this modification. In humans, 6 distinct N-terminal amino-acetyltransferases (NATs) catalyze the transfer of an acetyl group from acetyl-CoA to the Na-terminal amino group of their specific target proteins. The major human acetyltransferase, NatA, consists of an auxiliary subunit, Naa15, and a catalytic active subunit, Naa10. We have previously described two families with a lethal X-linked disorder of infancy called Ogden syndrome. This disorder comprises a distinct combination of an aged appearance, craniofacial anomalies, hypotonia, global developmental delays, cryptorchidism and cardiac arrhythmias. Using X chromosome exon sequencing, we identified a c.109T>C (p.Ser37Pro) variant in Naa10 as contributing to this disease. Biochemical analysis and immunoprecipitation assays in combination with LCMS demonstrated a reduced catalytic capacity and revealed an impaired binding of the S37P mutant towards specific interaction partners, including Naa15 and Naa50. Analysis of the N-terminal acetylome of patient cells revealed a decreased acetylation of a subset of NatA substrates, indicating that a reduced binding capability and an affected enzymatic activity of the Naa10 S37P mutation is a prominent feature in Ogden Syndrome. Characterization of *NAA10*/*NAA15* knockout yeast strains revealed various phenotypes, including growth defects at elevated temperatures and altered sensitivity towards cytotoxic stresses. These effects could be rescued by overexpressing human wild type Naa15/Naa10 from plasmids; however, overexpressing mutant Naa15/Naa10 S37P only partially rescue these effects. Interestingly, introduction of both human Naa15/Naa10 wt and S37P mutant into the endogenous locus of the corresponding yeast genes failed to reverse the effects. We also continued our efforts with establishing induced pluripotent stem cells (iPSCs) from skin fibroblasts from one of the boys with Ogden Syndrome, and we have also established knockout mice for NAA10 and knockin mice containing the mutation of interest in NAA10. Ongoing work will focus on characterizing the cells and mice.

TAF1 Syndrome: Characterization and Analysis of an Idiopathic Intellectual Disability Syndrome, with Jason O'Rawe, Yiyang Wu, Han Fang, Laura Jimenez Barron, Ed Yang (Boston), Alan Rope (Oregon) and Jeffrey Swensen (Arizona), Reid Robison (Utah), Kai Wang (California) and other collaborators.

This past year, we expanded and published our study on a new X-linked genetic syndrome associated with mutations in TAF1 and manifesting with global developmental delay, intellectual disability (ID), characteristic facial dysmorphology, generalized hypotonia, and variable neurologic features, all in male individuals. Simultaneous studies using diverse strategies led to the identification of nine families with overlapping clinical presentations and affected by de novo or maternally inherited single-nucleotide changes.

Two additional families harboring large duplications involving TAF1 were also found to share phenotypic overlap with the probands harboring single-nucleotide changes, but they also demonstrated a severe neurodegeneration phenotype. Functional analysis with RNA-seq for one of the families suggested that the phenotype is associated with downregulation of a set of genes notably enriched with genes regulated by E-box proteins. In addition, knockdown and mutant studies of this gene in zebrafish have shown a quantifiable, albeit small, effect on a neuronal phenotype. Our results suggest that mutations in TAF1 play a critical role in the development of this X-linked ID syndrome.

Development of comprehensive whole genome sequencing analysis pipelines, with Han Fang, Jason O'Rawe, Laura Jimenez Barron, Yiyang Wu, Michael Schatz, Giuseppe Narzisi, Kai Wang (California), Max He (Wisconsin).

We continued developing various bioinformatics approaches for the analysis of exome and whole genome sequencing data. For example, in one project, we showed that the accuracy of detection of small insertions and deletions (indels) is greater when using whole genome sequencing versus exon capture and sequencing. We also calculated that 60X WGS depth of coverage from the Illumina HiSeq platform is needed to recover 95% of indels detected by Scalpel. While this is higher than current sequencing practice, we proposed that the deeper coverage may save total project costs because of the greater accuracy and sensitivity. Finally, we investigated sources of INDEL errors (e.g., capture deficiency, PCR amplification, homopolymers). We reported over the past 12 months the results of several other ongoing bioinformatics projects as well, as shown in the below publications. For example, we developed SeqHBase, a big data-based toolset for analysing family-based sequencing data to detect *de novo*, inherited homozygous or compound heterozygous mutations that may be disease contributory. We demonstrated SeqHBase's high efficiency and scalability, which is necessary as WGS and WES are rapidly becoming standard methods to study the genetics of familial disorders. We also published an opinion piece regarding the current state of uncertainty quantification in DNA sequencing applications, and we proposed methods that can be used for accounting and propagating these errors and their uncertainties through subsequent calculations.

Genome-wide variant analysis of simplex autism families with an integrative clinical-bioinformatics pipeline, with Laura T. Jiménez-Barrón, Jason A. O'Rawe, Yiyang Wu, Margaret Yoon, Han Fang, and Ivan Iossifov.

Autism spectrum disorders (ASDs) are a group of developmental disabilities that affect social interaction and communication and are characterized by repetitive behaviors. There is now a large body of evidence that suggests a complex role of genetics in ASDs, in which many different loci are involved. Although many current population-scale genomic studies have been demonstrably fruitful, these studies generally focus on analyzing a limited part of the genome or use a limited set of bioinformatics tools. These limitations preclude the analysis of genome-wide perturbations that may contribute to the development and severity of ASD-related phenotypes. To overcome these limitations, we developed and utilized an integrative clinical and bioinformatics pipeline for generating a more complete and reliable set of genomic variants for downstream analyses. Our study focused on the analysis of three simplex autism families consisting of one affected child, unaffected parents, and one unaffected sibling. All members were clinically evaluated and widely phenotyped. Genotyping arrays and whole-genome sequencing were performed on each member, and the resulting sequencing data were analyzed using a variety of available bioinformatics tools. We searched for rare variants of putative functional impact that were found to be segregating according to de novo, autosomal recessive, X-linked, mitochondrial, and compound heterozygote transmission models. The resulting candidate variants included three small heterozygous copy-number variations (CNVs), a rare heterozygous de novo nonsense mutation in MYBBP1A located within exon 1, and a novel de novo missense variant in LAMB3. Our work demonstrated how more comprehensive analyses that include rich clinical data and whole-genome sequencing data can generate reliable results for use in downstream investigations.

KBG syndrome involving a single base insertion in *ANKRD11*, with Janet Malcolmson, Robert Kleyner, David Tegay, Kenneth Ward (Utah), Justine Coppinger (Utah), Annette Maughan (Utah), Glenn Maughan (Utah), Lesa Nelson (Utah), Kai Wang (California), Reid Robison (Utah).

KBG syndrome is a rare autosomal dominant genetic condition characterized by neurological involvement, macrodontia and distinct facial, hand and skeletal features. Over 70 cases have been reported; however it is likely that KBG syndrome is underdiagnosed due to lack of comprehensive characterization of the heterogeneous phenotypic features. We describe the clinical manifestations in a male referred at 11 years of

age, who exhibited symptoms including epilepsy, developmental delay, distinct facial features and hand anomalies, without positive genetic diagnosis. Subsequent exome sequencing identified a novel *de novo* heterozygous single base pair insertion (c.6015dupA) in *ANKRD11* which was validated by Sanger sequencing. We predict that this insertion leads to a premature stop codon and loss of function in *ANKRD11,* thereby implicating it as contributing to the proband's symptoms and yielding a molecular diagnosis of KBG syndrome for the case.

SCN8A Mutation in Child Presenting with Seizures and Developmental Delays, with Janet Malcolmson, Robert Kleyner, David Tegay, Whit Adams (Utah), Kenneth Ward (Utah), Justine Coppinger (Utah), Lesa Nelson (Utah), Kai Wang (California), Reid Robison (Utah).

The *SCN8A* gene encodes the Nav1.6 neuronal voltage-gated sodium channel alpha subunit. Mutations in this gene have been associated with early infantile epileptic encephalopathy type 13. With the use of whole exome sequencing, a missense mutation was identified in a 4-year-old female who initially exhibited symptoms at the age of 5-months, after she received routine vaccinations. Determining the molecular etiology of this proband's epileptic encephalopathy has improved her management and treatment.

Expanding collection and sequencing of other rare genetic syndromes, with Jason O'Rawe, Yiyang Wu, Han Fang, Reid Robison (Utah), Kai Wang (California) Alan Rope (Oregon), and others.

We continue to meet and collect many families in Utah and elsewhere with very rare, idiopathic genetic syndromes. The total number of DNA samples collected to date is approaching 2000, and this includes detailed phenotyping information. Some of these samples have undergone exome or whole genome sequencing, and we are currently analyzing 18 whole genomes and 50 exomes generated as part of this project. This includes the ongoing analysis of nine whole genomes from a pedigree with Prader–Willi Syndrome (PWS), Hereditary Hemochromatosis, Familial Dysautonomia (FD), and Tourette Syndrome.

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. Dorfel MJ, Lyon GJ: **The biological functions of Naa10 - From amino-terminal acetylation to human disease.** *Gene* 2015, **567:**103-131.
2. Guo Y, Ding X, Shen Y, Lyon GJ, Wang K: **SeqMule: automated pipeline for analysis of human exome/genome sequencing data.** *Sci Rep* 2015, **5:**14283.
3. He M, Person TN, Hebbring SJ, Heinzen E, Ye Z, Schrodi SJ, McPherson EW, Lin SM, Peissig PL, Brilliant MH, et al: **SeqHBase: a big data toolset for family based sequencing data analysis.** *J Med Genet* 2015, **52:**282-288.
4. Hirschtritt ME, Lee PC, Pauls DL, Dion Y, Grados MA, Illmann C, King RA, Sandor P, McMahon WM, Lyon GJ, et al: **Lifetime prevalence, age of risk, and genetic relationships of comorbid psychiatric disorders in Tourette syndrome.** *JAMA Psychiatry* 2015, **72:**325-333.
5. Jiménez-Barrón LT, O'Rawe JA, Wu Y, Yoon M, Fang H, Iossifov I, Lyon GJ: **Genome-wide variant analysis of simplex autism families with an integrative clinical-bioinformatics pipeline.** *Cold Spring Harb Mol Case Stud* 2015, **1:**a000422.
6. Myklebust LM, Van Damme P, Stove SI, Dorfel MJ, Abboud A, Kalvik TV, Grauffel C, Jonckheere V, Wu Y, Swensen J, et al: **Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects.** *Hum Mol Genet* 2015, **24:**1956-1976.
7. O'Rawe JA, Ferson S, Lyon GJ: **Accounting for uncertainty in DNA sequencing data.** *Trends Genet* 2015, **31:**61-66.
8. O'Rawe JA, Wu Y, Dorfel MJ, Rope AF, Au PY, Parboosingh JS, Moon S, Kousi M, Kosma K, Smith CS, et al: **TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations.** *Am J Hum Genet* 2015, **97:**922-932.