Human Genetics and Genomic Medicine
Gholson Lyon
with Max Doerfel, Jason O'Rawe, Yiyang Wu, Han Fang, Prashant Kota, Laura Jimenez

The Lyon laboratory focuses on analyzing human genetic variation and understanding how genetic mutations contribute to severe idiopathic neuropsychiatric disorders. We do this by studying large pedigrees living in the same geographic location, where one can study the expressivity and segregation of variants in a similar environmental background and with fewer population stratification concerns. Toward this end, we collect pedigrees in Utah and elsewhere, and then utilize exome and whole genome sequencing to find mutations that segregate with syndromes in the pedigrees. We focus on the discovery of families with rare diseases and/or increased prevalence for syndromes such as intellectual disability, autism and schizophrenia. Once we identify mutations that likely contribute to a disease, we undertake detailed functional studies of these mutations and the biological processes affected. Several projects are still at an early stage, but some of the projects that are sufficiently far enough along to discuss publicly are listed below.

Ogden Syndrome and the Amino-terminal acetylation of proteins
with Max Doerfel, Yiyang Wu, Ronen Marmorstein (Philadelphia, PA), Thomas Arnesen (Norway), Nathalie Reuter (Norway), Petra van Damme (Belgium)

We have previously identified a lethal X-linked disorder of infancy comprising a distinct combination of distinctive craniofacial features producing an aged appearance, growth failure, hypotonia, global developmental delays, cryptorchidism, and acquired cardiac arrhythmias. The first family was identified in Ogden, Utah, with five affected boys in two generations of family members. A mutation was identified as a c.109T>C (p.Ser37Pro) variant in *NAA10*, a gene encoding the catalytic subunit of the major human N-terminal acetyltransferase (NatA). This same mutation was identified in a second unrelated family, with three affected boys in two generations. This X-linked Malformation and Infantile Lethality Syndrome has been named Ogden Syndrome, in honor of the hometown where the first family resides. This is the first human disease involving a defect in the N-terminal acetylation of proteins, a common (yet vastly understudied) modifications of eukaryotic proteins carried out by N-terminal acetyltransferases (NATs). There is significantly impaired biochemical activity of the mutant hNaa10p, suggesting that a reduction in acetylation of some unidentified proteins by hNaa10p might lead to this disease. There is currently very limited knowledge on the functional importance of Nt-acetylation at the protein level and at the organismal level.

In order to understand the detrimental impact of the Naa10 S37P mutation, we performed structural, molecular and cellular investigations. The recently determined *S. pombe* NatA complex crystal structure was used as a template to generate a model of human NatA, revealing a highly conserved complex. The model allowed for comparison of Naa10 wt and Naa10 S37P within the NatA complex and suggested a decreased flexibility for Naa10 S37P in regions involved in catalysis and at the interface with the auxiliary subunit Naa15. The hydrogen bonding network between Naa10 and Naa15 was also rearranged. *In vitro* enzyme kinetics of Naa10 S37P demonstrated a reduced catalytic capacity, probably due to impaired peptide substrate binding. In agreement with the structural model, Naa10 S37P displayed a reduced capacity to form a stable NatA complex. N-terminal acetylome analyses of patient B-cells and fibroblasts provided a survey of Nt-acetylation in human non-cancer cells. In line with previous NatA knockdown data, patient derived S37P B-cells and fibroblasts have reduced Nt-acetylation for a subset of NatA-type substrates compared to cells from healthy family members, demonstrating *in vivo* perturbation of Naa10 (NatA) mediated Nt-acetylation in Ogden syndrome males. THOC7, one of the affected proteins, was shown to depend on Nt-acetylation for its

stability. Ogden syndrome fibroblasts further displayed abnormal cell proliferation and migration capacity. Therefore, the Ogden syndrome mutant Naa10 is impaired in NatA complex formation and catalytic capacity, and patient cells display reduced *in vivo* Nt-acetylation and cellular phenotypes potentially linked to the defects observed in the males suffering from this disease.

Low concordance of multiple variant-calling pipelines, practical implications for exome and genome sequencing
with Jason O'Rawe, Kai Wang (Los Angeles, California)

To facilitate the clinical implementation of genomic medicine by next-generation sequencing, it will be critically important to obtain accurate and consistent variant calls on personal genomes. Multiple software tools for variant calling are available, but it is unclear how comparable these tools are or what their relative merits in real-world scenarios might be. We sequenced 15 exomes from four families using commercial kits (Illumina HiSeq 2000 platform and Agilent SureSelect version 2 capture kit), with approximately 120X mean coverage. We analyzed the raw data using near-default parameters with five different alignment and variant-calling pipelines (SOAP, BWA-GATK, BWA-SNVer, GNUMAP, and BWA-SAMtools). We additionally sequenced a single whole genome using the sequencing and analysis pipeline from Complete Genomics (CG), with 95% of the exome region being covered by 20 or more reads per base. Finally, we validated 919 single-nucleotide variations (SNVs) and 841 insertions and deletions (indels), including similar fractions of GATK-only, SOAP-only, and shared calls, on the MiSeq platform by amplicon sequencing with approximately 5000X mean coverage.SNV concordance between five Illumina pipelines across all 15 exomes was 57.4%, while 0.5 to 5.1% of variants were called as unique to each pipeline. Indel concordance was only 26.8% between three indel-calling pipelines, even after left-normalizing and intervalizing genomic coordinates by 20 base pairs. There were 11% of CG variants falling within targeted regions in exome sequencing that were not called by any of the Illumina-based exome analysis pipelines. Based on targeted amplicon sequencing on the MiSeq platform, 97.1%, 60.2%, and 99.1% of the GATK-only, SOAP-only and shared SNVs could be validated, but only 54.0%, 44.6%, and 78.1% of the GATK-only, SOAP-only and shared indels could be validated. Additionally, our analysis of two families (one with four individuals and the other with seven), demonstrated additional accuracy gained in variant discovery by having access to genetic data from a multi-generational family. Our results suggest that more caution should be exercised in genomic medicine settings when analyzing individual genomes, including interpreting positive and negative findings with scrutiny, especially for indels. We advocate for renewed collection and sequencing of multi-generational families to increase the overall accuracy of whole genomes.

Advancing precision medicine through clinical grade whole genome sequencing, return of results, and neuromodulation.
with Jason O'Rawe, Han Fang, Martin Reese (California), Gerry Higgins (Washington, D.C.), Karen Eilbeck (Utah), Reid Robison (Utah)

For widespread precision medicine to become a reality, many things must be optimized, including clinical-grade sample collection, high-quality sequencing data acquisition, digitalized phenotyping, rigorous generation of variant calls, and online sharing of medical history and genomic data with research participants and others. We report the detailed phenotypic characterization, clinical-grade whole genome sequencing (WGS), and two-year outcome of a man with severe obsessive compulsive disorder (OCD) treated with deep brain stimulation (DBS) of the nucleus accumbens / anterior limb of the internal capsule. As part of his integrated medical care, his genome was sequenced and variants detected in the Illumina WGS Clinical

Laboratory Improvement Amendments (CLIA)-certified laboratory. It is increasingly apparent that mental illness results from a constellation of genetic and environmental factors, and, consistent with this, WGS did not reveal any one mutation of large effect, but instead showed that he carries several alleles that have been shown to elevate risk for mental illness. This includes the p.Val66Met variant in brain derived neurotrophic factor (BDNF), the p.Glu429Ala allele in methylenetetrahydrofolate reductase (MTHFR), and the p.Asp7Asn allele in choline O-acetyltransferase (ChAT). We identified thousands of other variants in his genome, including pharmacogenetic variants, and one mutation led to the discovery that he has untreated bilateral cataracts and other visual disturbances. We archived all data in the GVFclin format and also returned many results to this person. Since implantation of the deep brain stimulator, this man has reported steady improvement, highlighted by a decline in his Yale-Brown Obsessive Compulsive Scale (YBOCS) score from ~38 to a score of ~25. A rechargeable Activa RC neurostimulator battery has been of major benefit in terms of facilitating a degree of stability and control over the stimulation. His psychiatric symptoms reliably worsen within hours of the battery becoming depleted, thus providing confirmatory evidence for the efficacy of DBS for OCD in this person. To our knowledge, this was the first study in the clinical neurosciences that integrates detailed neuropsychiatric phenotyping, deep brain stimulation for OCD and clinical-grade WGS, with management and the first return of WGS results to a person with severe mental illness.

The Characterization and Analysis of an Idiopathic Intellectual Disability Syndrome via Whole Genome Sequencing Analysis
with Jason O'Rawe, Yiyang Wu, Alan Rope and Jeffrey Swensen (University of Utah),

We continued this year to study a new idiopathic syndrome with intellectual disability and distinctive facial dysmorphology. The propositi are two affected male brothers, aged 10 and 12 respectively, with severe intellectual disability, autism-like behavior, attention deficit issues, and very distinctive facial features, including broad, upturned nose, sagging cheeks, downward sloping palpebral fissures, relative hypertelorism, high-arched palate, and prominent ears. Their parents are nonconsanguineous and are both healthy, and the family history does not demonstrate any other members with anything resembling this current syndrome. X-chromosome inactivation assays reveal skewing in the mother, suggesting the possibility of an X-linked disorder. High-density genotyping arrays in the mother, father and two sons have not revealed previously known copy number variants (CNVs) that might contribute to the phenotype. Whole genome sequencing has led to the identification of several rare variants that are currently being characterized further, including in other unaffected members of the extended family.

Publications

1. **Lyon GJ**, Segal JP. Practical, ethical and regulatory considerations for the evolving medical and research genomics landscape. Applied & Translational Genomics. 2013;2(0):34-40. doi: http://dx.doi.org/10.1016/j.atg.2013.02.001.
2. O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome medicine. 2013;5(3):28. doi: 10.1186/gm432. PubMed PMID: 23537139; PubMed Central PMCID: PMC3706896.
3. O'Rawe JA, Fang H, Rynearson S, Robison R, Kiruluta ES, Higgins G, et al. Integrating precision medicine in the study and clinical treatment of a severely mentally ill person. PeerJ. 2013;1:e177. doi: 10.7717/peerj.177. PubMed PMID: 24109560; PubMed Central PMCID: PMC3792182.
4. Shi L, Zhang X, Golhar R, Otieno FG, He M, Hou C, et al. Whole-genome sequencing in an autism multiplex family. Molecular autism. 2013;4(1):8. doi: 10.1186/2040-2392-4-8.

PubMed PMID: 23597238; PubMed Central PMCID: PMC3642023.
5. Wang K, Kim C, Bradfield J, Guo Y, Toskala E, Otieno FG, et al. Whole-genome DNA/RNA sequencing identifies truncating mutations in RBCK1 in a novel Mendelian disease with neuromuscular and cardiac involvement. Genome medicine. 2013;5(7):67. doi: 10.1186/gm471. PubMed PMID: 23889995.

In Press:
6. **Lyon GJ**, O'Rawe, J. "Clinical Genetics of Neurodevelopmental Disorders". In: Mitchell K, editor. The Genetics of Neurodevelopmental Disorders. Preprint published on BioRxiv preprint server: doi: 10.1101/000687: Wiley; 2014.
7. **Lyon GJ**, Bird, L., Rope, A. "X-linked Malformations and Infantile Lethality". In: Zimmerman C, editor. Epstein's Inborn Errors of Development: Oxford University Press; 2014. In Press.